POST-TRANSLATIONAL MODIFICATIONS AND CONFORMATIONAL TRANSITIONS OF THE INTRINSICALLY DISORDERED ONCOPROTEINS HIGH-MOBILITY GROUP A

Settore scientifico-disciplinare BIO/10

DOTTORANDA
ELISA MAURIZIO

RESPONSABILE DOTTORATO DI RICERCA
Prof. GIANNINO DEL SAL

RELATORE
Prof. GUIDALBERTO MANFIOLETTI
Università degli Studi di Trieste

SUPERVISORE/TUTORE
Prof. GUIDALBERTO MANFIOLETTI
Università degli Studi di Trieste

CORRELATORE
Dott. RICCARDO SGARRA, PhD
Università degli Studi di Trieste

ANNO ACCADEMICO 2009-2010
Abstract

Intrinsically disordered proteins (IDPs) are flexible molecules, able to adapt to the surfaces of different molecular partners by means of specific, but easily reversible interactions. IDPs carry out pivotal biological functions participating in almost all cell signaling and regulatory pathways. Importantly, IDPs activities are finely modulated by the addition/removal of numerous post-translational modifications (PTMs) which are important conformational modulators. Prototypes of IDPs are High Mobility Group A (HMGA) proteins, which are expressed at high levels and play essential functions both in embryonic and cancer cells. HMGA protein family (HMGA1a, HMGA1b, and HMGA2) belong to the non-histone HMG chromatin protein super-family and are multifunctional architectural transcription factors. HMGA conformational adaptability and intricate pattern of dynamic and constitutive PTMs are thought to be responsible for this multifunctionality.

We performed a liquid chromatography-mass spectrometry (LC-MS) screening in twenty different cell lines in order to evaluate HMGA proteins PTM pattern and we evidenced relevant intra-family differences. Moreover, we focused on the poorly characterized HMGA2 and we mapped HMGA2 phosphorylation sites by mass sequencing demonstrating that, similarly to HMGA1, it is phosphorylated on the acidic C-terminal tail by CK2. Importantly, this modification turned out to affect HMGA2 DNA binding.

Since truncated HMGA proteins are more oncogenic than full-length ones and since HMGA are in vivo heavily modified on their C-terminal domain, we dissected the role of this domain and its phosphorylations from a structural point of view. We probed HMGA IDPs compactness and accessibility taking advantage of an innovative approach combining limited proteolysis and MS-based techniques. By limited proteolysis, ESI (electrospray ionization)-MS, and IMS (ion mobility separation)-MS we demonstrated that HMGA can assume a compact conformation and that their compactness degree is dependent upon the presence of the acidic C-terminal domain and its constitutive phosphorylations. Moreover, LC-MS analyses after enzymatic assays showed that HMGA forms with a deletion of acidic C-terminal tail are more susceptible to PTMs, thus supporting the idea that the acidic tail is involved in masking the accessibility of modifying enzymes to their own consensus sites.

We evidenced macroscopic differences regarding PTMs affecting the three HMGA family members and provided the first data about in vivo HMGA2 PTMs and their effect on DNA binding. Our structural investigations revealed a structure/PTMs relationship dictated by the presence of the C-terminal domain. This evidence, together with the already known in vivo functional outcome of HMGA C-terminal truncation, suggests a structure/function link between HMGA tails, their PTMs, and their oncogenic properties, paving the way for the development of interfering therapeutic strategies based on targeting HMGA proteins.
# Table of contents

List of abbreviations ............................................................................................................. 1

Chapter 1. Introduction ............................................................................................................ 3
  1.1 Intrinsically disordered proteins .................................................................................. 3
  1.2 Sequence signatures of intrinsic disorder .................................................................. 4
  1.3 Experimental characterization of disordered proteins ................................................. 5
  1.4 Prediction methods ..................................................................................................... 7
  1.5 IDPs and evolution ....................................................................................................... 8
  1.6 General structural characteristics of disordered proteins ........................................... 9
  1.7 Mechanisms of coupled folding and binding .............................................................. 11
  1.8 IDPs are hubs in protein interaction networks ........................................................... 13
  1.9 Tight regulation of IDPs: from transcript synthesis to protein degradation ............... 15
  1.10 The biological “cost” of disordered proteins: the D^2 concept ................................ 15
  1.11 Intrinsically disordered HMG proteins are key factors in cancer development ........ 16
    1.11.1 HMG proteins in malignant tumors ................................................................. 16
    1.11.2 HMG proteins in benign tumors ...................................................................... 17
    1.11.3 HMG proteins act also as tumor suppressors ..................................................... 18
  1.12 HMG protein family ................................................................................................ 18
  1.13 Transcriptional and post-transcriptional control of HMG expression ....................... 20
  1.14 HMG molecular network ......................................................................................... 22
  1.15 Mechanisms of action of HMG proteins ................................................................... 24
  1.16 Fine tuning of HMG binding by post-translational modifications ............................ 27
    1.16.1 Phosphorylation ................................................................................................. 27
    1.16.2 Acetylation ........................................................................................................ 31
    1.16.3 Methylation ....................................................................................................... 33
    1.16.4 Other minor modifications: poly(ADP-rybosyl)ation, SUMOylation, and formylation. 33
  1.17 HMGPTMs in different stages of neoplastic progression ........................................... 34
  1.18 HMG C-terminal tail and neoplastic transformation ................................................... 35
  1.19 Aims of the study ..................................................................................................... 37

Chapter 2. Materials and Methods ......................................................................................... 38
  2.1 Cell cultures ............................................................................................................... 38
  2.2 Recombinant HMG proteins production ..................................................................... 38
    2.2.1 Competent cells preparation .............................................................................. 39
    2.2.2. Bacterial transformation ................................................................................... 39
    2.2.3 Protein expression ............................................................................................. 39
  2.3 HMG proteins extraction ............................................................................................ 39
  2.4 GST-fused PRMTs production and extraction ............................................................ 40
  2.5 Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis (SDS-PAGE) ........ 41
  2.6 Liquid Chromatography-Mass Spectrometry (LC-MS) screening ........................... 41
  2.7 RP-HPLC purification of HMG proteins .................................................................... 42
  2.8 Protein quantification ............................................................................................... 42
    2.8.1 Waddell method ................................................................................................ 42
2.9 In vitro HMGA modification assays .........................................................44
  2.9.1 In vitro phosphorylation of HMGA proteins by Casein kinase 2 (CK2) ...44
  2.9.2 In vitro comparative phosphorylation of FL and CT HMGA proteins by Cyclin dependent kinase 1 (CDK1) ........................................44
  2.9.3 In vitro comparative methylation of FL and CT HMGA proteins by Protein arginine N-methyltransferases (PRMT1 and PRMT6) ........44
2.10 LC-MS analyses of HMGA ........................................................................45
2.11 HMGA2 protein phosphate group mapping .................................................45
  2.11.1 Enzymatic digestions ........................................................................45
  2.11.2 Phospho-peptides enrichment .............................................................45
  2.11.3 LC-MS/MS analyses (LC-Tandem Mass Spectrometry analyses) ..........46
2.12 Electrophoretic Mobility Shift Assays (EMSA) .............................................47
  2.12.1 DNA probes .....................................................................................47
  2.12.2 [γ-32P] labeling of DNA probe ..........................................................47
  2.12.3 Protein/DNA binding .......................................................................48
  2.12.4 Electrophoretic analysis .....................................................................48
2.13 Electrospray Ionization – Mass Spectrometry (ESI-MS) analyses ...............48
2.14 Limited proteolysis analyses .....................................................................49
2.15 Ion Mobility Separation – Mass Spectrometry (IMS-MS) analyses ............50
2.16 Immunoprecipitation (IP) of XHMGA2β-a from PCA extracts ..................50
2.17 Circular dichroism (CD) measurements ....................................................50

Chapter 3. Results and Discussion ....................................................................52
  3.1 Liquid Chromatography-Mass Spectrometry (LC-MS) screening ...............52
  3.2 Identification of HMGA and HMGN post-translational modifications ..........55
  3.3 Differences among the post-translational modifications of HMGA and HMGN proteins in different cell types ........................................65
  3.4 HMGA2 post-translational modifications in an animal model .....................67
  3.5 HMGA2 is multi-phosphorylated by CK2 on the acidic tail .......................69
  3.6 HMGA2 phosphate group mapping ..........................................................71
    3.6.1 Time Course phosphorylation assay ................................................71
    3.6.2 Mass sequencing analyses ................................................................73
  3.7 Phosphorylation of the HMGA2 acidic C-terminal tail affects its DNA binding properties .................................................................84
  3.8 Could HMGA highly phosphorylated acidic tail have a conformational role? 86
  3.9 The C-terminal tail of HMGA proteins shields the protein-protein interaction domain ..............................................................88
  3.10 Conformational role of HMGA acidic tail ...............................................93
  3.11 C-terminal truncation of HMGA proteins makes them more susceptible to PTMs 101

Chapter 4. Conclusions ....................................................................................105

Collaborations .................................................................................................110
References .......................................................................................................111
Attachments ....................................................................................................119
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>acetylation</td>
</tr>
<tr>
<td>Abl</td>
<td>abelson murine leukemia viral oncogene homolog</td>
</tr>
<tr>
<td>AP-1/c-Jun</td>
<td>activator protein 1/proto-oncogene c-Jun</td>
</tr>
<tr>
<td>APE1/REF-1</td>
<td>DNA-(apurinic or apyrimidinic site) lyase</td>
</tr>
<tr>
<td>ATF-2</td>
<td>activating transcription factor 2</td>
</tr>
<tr>
<td>ATM</td>
<td>ataxia-telangiectasia mutated</td>
</tr>
<tr>
<td>Bax</td>
<td>Bcl-2-associated X protein</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Breast cancer type 1 susceptibility protein</td>
</tr>
<tr>
<td>CBP</td>
<td>CREB-binding protein</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>CDK1</td>
<td>cycline dependent kinase 1</td>
</tr>
<tr>
<td>c-Fos</td>
<td>proto-oncogene c-Fos</td>
</tr>
<tr>
<td>CID</td>
<td>collision induced dissociation</td>
</tr>
<tr>
<td>CK2</td>
<td>casein kinase 2</td>
</tr>
<tr>
<td>c-Myc</td>
<td>myc proto-oncogene protein</td>
</tr>
<tr>
<td>CREB</td>
<td>cyclic AMP-responsive element-binding protein</td>
</tr>
<tr>
<td>Crk</td>
<td>proto-oncogene c-Crk</td>
</tr>
<tr>
<td>CSD</td>
<td>charge state distribution</td>
</tr>
<tr>
<td>CT</td>
<td>C-terminal truncated</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>DFF45/40</td>
<td>DNA-fragmentation factor 45/40-kDa subunit</td>
</tr>
<tr>
<td>E2F1</td>
<td>transcription factor E2F1</td>
</tr>
<tr>
<td>EGF</td>
<td>epidermal growth factor</td>
</tr>
<tr>
<td>eIF4E</td>
<td>eukaryotic translation initiation factor 4E</td>
</tr>
<tr>
<td>EMSA</td>
<td>electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>ESI-MS</td>
<td>electrospray ionization-mass spectrometry</td>
</tr>
<tr>
<td>FGF</td>
<td>fibroblast growth factor</td>
</tr>
<tr>
<td>FL</td>
<td>full-length</td>
</tr>
<tr>
<td>FRAP</td>
<td>fluorescence recovery after photobleaching</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td>HDAC1</td>
<td>histone deacetylase1</td>
</tr>
<tr>
<td>HIPK2</td>
<td>homeodomain-interacting protein kinase</td>
</tr>
<tr>
<td>HMG (A-B-N)</td>
<td>high mobility group (A/B/N)</td>
</tr>
<tr>
<td>ID</td>
<td>intrinsically disordered</td>
</tr>
<tr>
<td>IDPs</td>
<td>intrinsically disordered proteins</td>
</tr>
<tr>
<td>IL-1β</td>
<td>interleukin-1β</td>
</tr>
<tr>
<td>IL-2</td>
<td>interleukin-2</td>
</tr>
<tr>
<td>IL-4</td>
<td>interleukin-4</td>
</tr>
<tr>
<td>IMS-MS</td>
<td>ion mobility separation-mass spectrometry</td>
</tr>
<tr>
<td>INF-β</td>
<td>interferon-β</td>
</tr>
<tr>
<td>IP</td>
<td>immunoprecipitation</td>
</tr>
<tr>
<td>IR</td>
<td>insulin receptor</td>
</tr>
<tr>
<td>IRFs</td>
<td>interferon regulatory transcription factors</td>
</tr>
<tr>
<td>(p)KID</td>
<td>(phosphorylated) kinase-inducible transcriptional-activation domain</td>
</tr>
<tr>
<td>KIX</td>
<td>KID binding domain</td>
</tr>
<tr>
<td>Ku70/80</td>
<td>ATP-dependent DNA helicase II 70 KDa and 80 kDa subunits</td>
</tr>
<tr>
<td>LC-MS</td>
<td>liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography-tandem mass spectrometry</td>
</tr>
<tr>
<td>LC-UV</td>
<td>liquid chromatography-ultraviolet absorbance detection</td>
</tr>
<tr>
<td>LIM domain</td>
<td>Lin11, Isl-1 &amp; Mec-3 domain (two contiguous zinc finger domains)</td>
</tr>
<tr>
<td>LPP</td>
<td>lipoma preferred partner</td>
</tr>
<tr>
<td>M</td>
<td>methylation</td>
</tr>
<tr>
<td>MAPK</td>
<td>mitogen-activated protein kinase</td>
</tr>
<tr>
<td>Mdm2</td>
<td>E3 ubiquitin-protein ligase Mdm2/murine double minute-2</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>METAP2</td>
<td>methionine aminopeptidase 2</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MYCN</td>
<td>V-myc myelocytomatosis viral related oncogene, neuroblastoma derived</td>
</tr>
<tr>
<td>Nek2</td>
<td>serine/threonine-protein kinase Nek2/never in mitosis A-related kinase 2</td>
</tr>
<tr>
<td>NF-kB</td>
<td>nuclear factor of kappa light polypeptide gene enhancer in B-cells</td>
</tr>
<tr>
<td>Nkx2.5</td>
<td>homeobox protein NK-2 homolog E</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>ORC</td>
<td>origin of replication complex</td>
</tr>
<tr>
<td>P</td>
<td>phosphorylation</td>
</tr>
<tr>
<td>P120E4F</td>
<td>transcription factor E4F1</td>
</tr>
<tr>
<td>p21</td>
<td>cyclin-dependent kinase inhibitor 1 or CDK-interacting protein 1</td>
</tr>
<tr>
<td>p300</td>
<td>histone acetyltransferase p300</td>
</tr>
<tr>
<td>p53</td>
<td>cellular tumor antigen p53</td>
</tr>
<tr>
<td>PCA</td>
<td>perchloric acid</td>
</tr>
<tr>
<td>PCAF/GCN5</td>
<td>p300/CREB-associated factor/GCN5</td>
</tr>
<tr>
<td>PDGF</td>
<td>platelet-derived growth factor</td>
</tr>
<tr>
<td>PKC</td>
<td>protein kinase C</td>
</tr>
<tr>
<td>PML</td>
<td>promyelocytic leukemia protein</td>
</tr>
<tr>
<td>Pol II</td>
<td>DNA polymerase II</td>
</tr>
<tr>
<td>PONDR</td>
<td>predictor of natural disordered regions</td>
</tr>
<tr>
<td>pRb</td>
<td>retinoblastoma protein</td>
</tr>
<tr>
<td>PRMT(1/3/6)</td>
<td>protein arginine methyl-transferase(1/3/6)</td>
</tr>
<tr>
<td>PTEN</td>
<td>Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN/phosphatase and tensin homolog</td>
</tr>
<tr>
<td>PTMs</td>
<td>post-translational modifications</td>
</tr>
<tr>
<td>Ras</td>
<td>rat sarcoma GTPase</td>
</tr>
<tr>
<td>RBBP</td>
<td>retinoblastoma-binding protein</td>
</tr>
<tr>
<td>RC3</td>
<td>neurogranin</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>reverse phase-high pressure liquid chromatography</td>
</tr>
<tr>
<td>RUVBLs</td>
<td>rB-like proteins</td>
</tr>
<tr>
<td>SAHF</td>
<td>senescence-associated heterochromatic foci</td>
</tr>
<tr>
<td>SARs/MARs</td>
<td>scaffold/matrix associated regions</td>
</tr>
<tr>
<td>SAXS</td>
<td>small-angle X-ray scattering</td>
</tr>
<tr>
<td>SDCCAG10</td>
<td>serologically defined colon cancer antigen 10</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulphate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SH3</td>
<td>SRC Homology 3 Domain</td>
</tr>
<tr>
<td>SM22α</td>
<td>smooth muscle protein 22-alpha/transgelin</td>
</tr>
<tr>
<td>Smad</td>
<td>mothers against decapentaplegic homolog</td>
</tr>
<tr>
<td>Snail1</td>
<td>zinc finger protein SNAI1</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>SP1</td>
<td>specificity protein 1 transcription factor</td>
</tr>
<tr>
<td>SRF</td>
<td>serum response factor</td>
</tr>
<tr>
<td>STAT3</td>
<td>signal transducer and activator of transcription3</td>
</tr>
<tr>
<td>Swi/Snf</td>
<td>SWItch/Sucrose NonFermentable chromatin-remodeling complex</td>
</tr>
<tr>
<td>TF</td>
<td>transcription factor</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>TGF-α</td>
<td>transforming growth factor-α</td>
</tr>
<tr>
<td>TGF-β</td>
<td>transforming growth factor-β</td>
</tr>
<tr>
<td>TIC</td>
<td>total ion count chromatogram</td>
</tr>
<tr>
<td>TPA</td>
<td>12-O-tetradecanoylphorbol-13-acetate</td>
</tr>
<tr>
<td>U1snRNP 70KDa</td>
<td>U1 small nuclear ribonucleoprotein 70 kDa</td>
</tr>
<tr>
<td>Ubc9</td>
<td>SUMO-conjugating enzyme UBC9/ubiquitin carrier protein 9</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>Wat</td>
<td>wingless/integrated signaling molecules</td>
</tr>
<tr>
<td>4E-BP1</td>
<td>eukaryotic translation initiation factor 4E-binding protein 1</td>
</tr>
</tbody>
</table>
1. Introduction

Interactions among proteins are crucial events underlying all biological processes. This is the reason why cells constantly control protein contacts using specific regulative mechanisms. In particular, the addition of reversible covalent post-translational modifications (PTMs) to peculiar amino acids embedded in binding regions is often the faster way to directly modulate protein conformations, interactions, and therefore function.

The architecture of protein-protein interaction networks is settled by the presence of highly interconnected hubs usually belonging to the intrinsically disordered proteins category and performing dynamically regulated interactions with a huge number of different molecular partners. These proteins possess a high level of intrinsic dynamics and flexibility because of the lack of any fixed conformational structure. Prototypes of this category are HMGA architectural transcription factors that have been defined as central hubs of nuclear function (Reeves, 2001); in fact, they participate in many different cellular processes being able to adapt to a large number of molecular partners. HMGA are finely modulated by numerous and different PTMs and, as well as histones, are among the most modified proteins in the nucleus.

HMGA carry out pivotal biological roles and they are referred to as oncofetal proteins since they are highly expressed and play essential functions both during embryonic development and neoplastic transformation. Moreover, given their well assessed causal role in cancer, HMGA are considered novel suitable targets for cancer chemotherapy.

1.1 Intrinsically disordered proteins

Proteins not adopting well-defined, stable three-dimensional structures, either entirely or in parts, under physiological conditions when alone in solution, are defined as intrinsically or natively disordered/denatured/unfolded/unstructured proteins (IDPs or IUPs). Such different words reflect differences in the underlying biophysical traits of these regions (Schlessinger et al., 2009; Dyson and Wright, 2005; Uversky, 2002; Uversky et al., 2000).

In the last decade, systematic studies on IDPs demolished one of the cornerstones in protein biology, chemistry and physics: the “structure-function paradigm” (Wright and Dyson, 1999). This concept claims that the specific function of a protein is determined by its unique and rigid three-dimensional structure. But, for IDPs this paradigm doesn’t work, in fact they occupy a unique
structural and functional niche of protein kingdom in which function is directly linked to structural disorder. In fact, disordered segments fold on binding to different biological targets or constitute flexible linkers that have a role in the assembly of regulative macromolecular complexes.

### 1.2 Sequence signatures of intrinsic disorder

IDPs have unique sequence features. In fact, a strong amino acid compositional bias that leads to a low mean hydrophobicity and a high net charge is a signature of intrinsic disorder in proteins (Figure 1.1) (Uversky, 2002; Uversky et al., 2000).

![Figure 1.1](image)

**Figure 1.1** IDPs proteins are characterized by low overall hydrophobicity and high net charge. Mean net charge versus mean hydrophobicity plot for a set of 275 folded (blue squares) and 91 IDPs (red circles). The green line represents the border between extended IDPs and ordered proteins (Dunker et al., 2008, adapted from Uversky et al., 2000).

In fact, IDPs amino acid alphabet is depleted of hydrophobic and bulky amino acids (V, L, I, M, F, W and Y) which are often referred to as “order promoting” residues, and enriched in “disorder promoting” small, polar, and charged residues (Q, S, P, E, K and, on occasion, G and A) compared with the average of folded proteins. The lack of hydrophobic residues prevents the formation of a hydrophobic core necessary for a stable three-dimensional fold and the large number of uncompensated side chain charges contributes to destabilize any compact state by electrostatic repulsion. Moreover, polar and charged residues, which favorable interact with water, confer to unstructured regions a large solvent-accessible area (Uversky et al., 2000; Gsponer and Babu, 2009; Dyson and Wright, 2005 and references therein).

Other sequence-related biases in disordered regions include the depletion in cysteine, that often occurs in enzyme active sites or that forms disulfide bonds in globular proteins, and the high content in proline, known to disfavor a rigid secondary structure (Tompa, 2002).
The amino acid composition of disordered regions was also found to correlate with their length. For example, short disordered stretches are mainly negatively charged whereas long unstructured regions are either positively or negatively charged, but on average, nearly neutral (Radivojac et al., 2004). Two types of patterns are highly abundant in disordered regions: a proline-rich pattern and a (positively or negatively) charged pattern (Lise and Jones, 2005). Interestingly, many proline-rich motifs in unstructured regions are important for protein-protein interactions. For instance, (i) proline-rich stretches are abundant in molecular recognition features (MoRFs - short disordered motifs of about 10-70 aa that exhibit molecular recognition and binding functions) of IDPs (for example α-synuclein, caseins and tau protein) (Mohan et al., 2006; Syme et al., 2002), and (ii) the P-X-X-P motif in SH3-binding domains is important for mediating protein-protein interactions (for example the binding of Abl kinase with Crk SH3 domain) (Fuxreiter et al., 2007; Zheng et al., 2010).

Moreover, IDPs tend to have a low sequence complexity, although it is not a general rule. Low complexity regions are regions with a biased composition having a strong overrepresentation of a few residues (homopolymeric runs, short-period repeats, etc.) (Romero et al., 2001; Wootton, 1994)

All the above analyzed biochemical features go beyond the classical oversimplified view of IDPs as fully unstructured, featureless proteins underlining that “order” at the level of amino acid composition and primary sequence reconciles with IDPs functional perfection (Tompa, 2002).

1.3 Experimental characterization of disordered proteins

In contrast to ordered protein regions, intrinsically disordered (ID) regions are not characterized by the atom equilibrium positions and dihedral angle equilibrium values around which amino acid residues spends most of the time. In fact, IDPs and ID regions exist as dynamic ensembles of rapidly inter-converting conformers in which atom positions and backbone Ramachandran angles vary significantly over time with no specific equilibrium values (Radivojac et al., 2007). Therefore, the experimental assignment of disordered regions is problematic.

One way to overcome this problem is by measuring biophysical characteristics that are associated with the lack of ordered three-dimensional structure (Schlessinger et al., 2009).

The most direct way of identifying disordered regions in proteins is to look for residues with missing backbone coordinates in three-dimensional structures of proteins determined by X-ray crystallography. In fact, disordered amino acid stretches fail to crystallize into fixed structures and, therefore, scatter X-rays incoherently and are invisible in the resulting electron density maps (Romero et al., 1997).
To study proteins that are disordered along their entire length, novel nuclear magnetic resonance (NMR) techniques turned out to be very versatile and powerful tools in elucidating structure and internal molecular dynamics in the disordered states. Solution NMR can monitor unfolded polypeptide conformations and their transitions at atomic resolution. Differently from traditional NMR limited and often qualitative descriptions, new quantitative parameters reporting on the averages of local conformations and long-range interactions of denatured proteins can now be measured. This information has been used to obtain plausible all-atom models of the unfolded state at increasing accuracy (Meier et al., 2008).

The other technique that is used most often is far-UV circular dichroism (CD) spectroscopy, which detects the amount (or lack) of secondary structure (alpha-helix, beta sheet and random coil structures). The ellipticity spectrum of IDPs has a large negative peak at around 200 nm and a value close to zero at 220 nm, distinct enough from that of ordered conformations to allow identification of partially or fully unstructured proteins. Moreover, residual structure on the order of 10-20% can be determined (Kelly et al., 2005).

Further spectroscopic techniques supplement these data and can be also used to monitor folding transitions associated with binding. For example Fourier transform infrared (IR) spectroscopy also provides information on the secondary structure, near-UV CD, UV spectroscopy and fluorimetry report on the environment of aromatic residues, and Raman optical activity measurements characterize protein secondary structure and provide a snapshot of conformations allowing distinction between different types of disorder (Tompa, 2002; Syme et al., 2002).

Unfolded conformations can also be detected and characterized by hydrodynamic techniques. Gel filtration (size-exclusion chromatography), small-angle X-ray scattering (SAXS), sedimentation analysis and dynamic light scattering provide information on hydrodynamic parameters. Other techniques can add an extra dimension to these studies. Differential scanning calorimetry unveils the absence of cooperative folding transitions characteristic of a well-defined tertiary fold. Proteolytic assays easily identify ID regions in proteins for their higher sensitivity to proteolytic cleavage with respect to globular domains (Fontana et al., 2004). Moreover, also heat-stability assays can be used to study IDPs. In fact, these proteins do not lose solubility at elevated temperatures because they do not have hydrophobic core.

A further method is sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), which is routinely used to assess molecular weight (MW) of proteins. Because of their unusual amino acid composition, IDPs bind less SDS than other proteins and their apparent MW is often 1.2–1.8 times higher than the real one calculated from sequence data or measured by mass spectrometry.
(A list of techniques useful to detect and characterize IDPs can be found in Tompa, 2002 and Radivojac et al., 2007).

Recently, a newly developed degradation assay using 20S proteasome turned out to be a suitable system for operational definition of IDPs. In fact, 20S proteasomes can digest IDP sequences under conditions in which native, and even molten globule states (see par. 1.6), are resistant and, furthermore, can give information on IDPs state since protein-protein interactions have been demonstrated to protect IDPs against 20S proteasomal action (Tsvetkov et al., 2008).

Interestingly, advances in the field of mass spectrometry in the last years have suggested powerful new tools for protein conformational investigations. These approaches are all based on the principle that most proteins retain their native conformation while moving from the solution to the gas phase during electrospray ionization (Yan et al., 2004, Kaltashov and Abzalimov, 2008).

Since each of the above mentioned experimental techniques captures different aspects of protein disorder there is no single gold standard for order/disorder assignment, instead, several experimental methods have to be used in concert (Schlessinger et al., 2009).

1.4 Prediction methods

The use of bioinformatics methods of prediction and classification of proteins has greatly enhanced research on intrinsically unstructured proteins.

Although the wide range of biophysical techniques that have been developed for studying IDPs only bioinformatic techniques can tackle this issue in general at the present. One of the main reasons for the predictability of unstructured regions is their amino acid compositional bias.

Several tools (by now more than 25) have been developed to predict protein disordered segments (Ferron et al., 2006; Uversky et al., 2009). Generally these methods rely on sequence features such as amino acid composition, predicted secondary structure, sequence complexity, flexibility, hydropathy and sequence variability and are trained on available datasets of experimentally characterized disordered proteins. Some of the most used programs are PONDR (Li et al., 1999), FoldIndex (Prilusky et al., 2005), GlobPlot (Linding et al., 2003), DISOPRED (Ward et al., 2004) and DisEMBL (Linding et al., 2003bis).

PONDR (Predictor Of Natural Disordered Regions – www.pondr.com) algorithm was the first predictor to be developed and it is now available in various versions, each having its own specificities. Figure 1.2 shows two PONDR predictions of p53 protein together with a scheme of the experimentally characterized ordered/disordered regions (Oldfield et al., 2008).
It is important to underline that different predictors rely on different physicochemical parameters and can be considered complementary. However, it has been estimated that the per residue prediction accuracy of the most elaborated programs reaches 85%.

![Figure 1.2 PONDR prediction and experimentally characterized regions of disorder and order in p53.](image)

Figure 1.2 PONDR prediction and experimentally characterized regions of disorder and order in p53. p53 experimentally characterized regions of disorder (red) and order (blue) are indicated by the horizontal bar. The plot gives predictions of disorder (scores > 0.5) and order (scores < 0.5) for two PONDR predictor versions (adapted from Oldfield et al., 2008).

In parallel with prediction algorithms, also databases of disordered protein regions have been generated. Among these, DisProt is the largest publicly available dataset that provides information about experimentally assessed proteins that lack fixed three-dimensional structure, either entirely or in part, in their native states (Sickmeier et al., 2007). The current release, DisProt 5.6, consists of 638 proteins and 1368 protein regions (www.disprot.org).

1.5 IDPs and evolution

The development of computational approaches to predict unstructured segments from protein sequence coupled with the availability of genomic data for various prokaryotic and eukaryotic species has led to an extensive use of these programs in order to quantify the abundance of IDPs in the different kingdoms of life. Analysis of sequence data from 29 genomes indicates that IDPs are highly prevalent, and that the proportion of proteins that contain ID segments increases with the increasing complexity of an organism: eukaryotes » archaea » eubacteria (Dunker et al., 2000; Dyson and Wright, 2005).
As regards mammals, about 75% of their signaling proteins and about 50% of their total proteins are predicted to contain long disordered regions (> 30 residues), whereas about 25% of their proteins are predicted to be fully disordered (Dunker et al., 2008bis).

Evolutionary investigations indicate that IDPs evolve at a significant faster rate than ordered proteins because they have fewer evolutionary constraints. These faster rates include changes that result in amino acid substitutions, insertion, deletions, and especially repeat expansions of genetically unstable tandemly repeated short segments (Huntley and Golding, 2000; Brown et al., 2002; Tompa, 2003; Lin et al., 2007). Notably, several studies indicate that the functions of ID regions are maintained even in the face of this rapid evolution allowing the extension and functional expansion of IDPs class (Brown et al., 2010; Tompa, 2003).

Moreover, alternative splicing, that provides an important mechanism for enhancing protein diversity in multicellular eukaryotes, is more frequently mapped in regions encoding for disordered rather than ordered protein sequences (Romero et al., 2006).

### 1.6 General structural characteristics of disordered proteins

In terms of their structure, globular proteins can be defined as being in one of four states of folding: native (ordered), molten globule, pre-molten globule or unfolded. This concept was defined *The Protein Quartet model* of protein functioning (Figure 1.3). A native state is the highest level of three-dimensional organization of a globular protein and it is determined by the complex set of non-covalent interactions defining protein tertiary structure. The molten globule was originally defined as a compact state of a protein with total native-like secondary structure but disordered tertiary structure, whereas the pre-molten globule state has only ~50% native secondary structure and loses completely the globular structure. In the less compact random coil the amino acid chain has random distribution of dihedral angles (Uversky, 2002). In fact, a polymer molecule is randomly coiled when internal rotation can take place at about every single bond of the molecule with the same freedom with which it would take place in a molecule of low molecular weight containing the same kind of bonds (Tanford, 1968). However, a globular protein rarely, if ever, behaves as a true random coil, especially in non-denaturing conditions.

Protein Quartet conformational classification of “normal” globular protein folding states can be also used to explain the different conformations adopted by all proteins in their functional state (Uversky, 2002).
Moreover, proteins have been defined to fall onto a *structural continuum*, from highly extended heterogeneous unstructured states, to compact but disordered molten globules, to multi-domain or modular proteins that might have flexible or disordered linker regions, and, finally, to tightly folded and ordered single domains (Dyson and Wright, 2005). A graphical representation for this concept is reported in the examples in Figure 1.4.

**Figure 1.3. The Protein Quartet model.** Protein Quartet model of protein functioning. In accordance with this model, function arises from four specific conformations of the polypeptide chain (ordered forms, molten globules, premolten globules, and random coils) and transitions between any of the states (Uversky, 2002).

**Figure 1.4. The continuum of protein structure.** Examples on the continuum of protein structure: an unstructured conformational ensemble, which is represented by the interaction domain of activator for thyroid hormone and retinoid receptors (ACTR); a molten globule-like domain such as the nuclear-receptor co-activator-binding domain (NCBD) of CBP; linked folded domains such as the first three zinc fingers of transcription factor-III A (TFIIIA); and free eukaryotic translation-initiation factor (eIF)4E, which is mostly folded with only local disorder (adapted from Dyson and Wright, 2005).

In this complex view, IDPs represent a protein family showing a complete (or almost complete) loss of any ordered structure having (i) larger hydrodynamic dimensions compared to typical native globular proteins with corresponding molecular mass, (ii) low content of secondary structure, and (iii) high intra-molecular flexibility. However, IDPs do not represent a uniform family, but rather two structurally different groups: intrinsic coils (or extended IDPs), with no ordered secondary structure, and pre-molten globules (or collapsed IDPs), more compact, exhibiting some amount of residual secondary structure. Since amino acid sequences of intrinsic coils and pre-molten globule
are similar, some other sequence features (e.g. propensity to form secondary structure) have to be taken into account (Uversky, 2002).

Recent findings underlined a decisive role played by charge interactions on conformational properties of IDPs. In fact, in proteins with high net charge, charge repulsion can lead to a pronounced expansion, whereas in proteins with a similar number of positive and negative charges, charge attraction can drive the collapse of the chain showing a compaction of the molecule (Muller-Spath et al., 2010).

An intriguing feature of IDPs is their capability to undergo a disorder-to-order transition during or prior functioning. The degree of these structural rearrangements varies over a very wide range, from coil to pre-molten globule transitions to formation of rigid ordered structures. Therefore, IDPs functional transitions follow the Protein Quartet model, with biological activity arising from unique conformations of the polypeptide chain and transitions between any of them (Uversky, 2002).

### 1.7 Mechanisms of coupled folding and binding

The phenomenon describing conformational functional transitions of IDPs is named *coupled folding and binding*. During this process an IDP folds into a more ordered structure concomitant with binding to its target.

There is an entropic cost associated with the disorder-to-order transition, but the key thermodynamic driving force for the binding reaction is generally a favorable enthalpic contribution, which gives an example of enthalpy-entropy compensation (Dyson and Wright, 2005). Coupled folding and binding frequently gives rise to a protein complex with high specificity and relatively low affinity, which is appropriate for signal transduction proteins that must not only associate specifically to initiate the signaling process, but must also be capable of rapid dissociation when signaling is complete.

One of the better characterized examples of coupled folding and binding is that undergone by the phosphorylated kinase-inducible transcriptional-activation domain (pKID) of the transcription factor CREB. The KID polypeptide is intrinsically disordered when free in solution, both as isolated peptide and in full-length CREB, but it folds to form a pair of orthogonal helices on binding to its target domain (KIX) in the transcriptional co-activator CBP (Radhakrishnan, 1997 and 1998). The intrinsic disorder of KID can be reliably predicted from its amino acid sequence, as can be predicted an inherent helical propensity in the region that undergoes the coupled folding and binding transition (Figure 1.5).
Chapter 1

Introduction

Figure 1.5 Illustration of disorder-to-order transition upon binding. A. Representation of coupled folding and binding of pKID with CBP KIX domain (adapted from Zhou 2001). B. Amino acid sequence of KID. Green: small residues, uncharged hydrophilic residues, and Pro. Yellow: hydrophobic residues. Red: acidic residues. Blue: basic residues. Helices that were predicted using The PredictProtein server, but which are not observed in the free peptide, are shown in grey above the sequence. The locations of the helices that are observed in pKID on its binding to KIX are shown in pink below the sequence (Dyson and Wright, 2005).

Conformational transitions in IDPs can involve just a few residues, as in KID case, or bigger entire protein domains. For example, the 116-residue N-terminal domain of DNA-fragmentation factor 45-kDa subunit (DFF45) is unstructured in solution, but folds into an ordered globular structure on forming a heterodimeric complex with DFF40 (Zhou et al., 2001).

Often different structures can be adopted with different partners. p53 represents the best example of a protein that can use different disordered regions to bind to different partners and which also has several individual disordered regions each binding to multiple partners. One of these regions is shown in Figure 1.6 together with a representation of the possible conformational transitions that undergoes (Oldfield et al., 2008).

Figure 1.6 p53 C-terminal region fold differently upon binding to different molecular partners. Sequence and structure comparison for four interacting complexes involving a binding region located on the C-terminus of p53 (aa 374-388). Primary, secondary, and quaternary structure of p53 complexes with S100β, sirtuin, CBP and cyclin A2 are represented. In the four complexes p53 binding region displays all three major secondary structure types: it becomes a helix when binding to S100β, a sheet when binding to sirtuin, and a coil with two distinct backbone trajectories when binding to CBP and cyclin A2 (modified from Oldfield et al., 2008).
Moreover, a recent paper showed how post-translational modifications (PTMs) are also implied in the coupled folding and binding mechanism. In fact, the major regulatory function of S65 phosphorylation in the intrinsically disordered protein 4E-BP1 is to modulate the propensity of its binding motif to fold into a conformation that fits the binding surface on its partner eIF4E (Tait, 2010). The binding modes attainable by disordered proteins are highly diverse, creating a multitude of unusual complexes. Although the majority of studied intrinsic disorder-based complexes are ordered or static entities originating due to global or local disorder-to-order transitions, a new striking development is the discovery of dynamic complexes in which IDPs continuously sample an ensemble of rapidly interconverting conformations mostly devoid of structure even in their bound state (Uversky, 2011).

1.8 IDPs are hubs in protein interaction networks
IDPs have been shown to hold key positions in mediating signaling and coordinating regulatory events in the recently mapped protein-protein interaction networks of several different eukaryotes. In fact, IDPs are able to perform dynamically regulated interactions with a huge number of different molecular partners, both in time and in space and for this reason they are considered protein network hubs (Dunker et al., 2005). A single IDP can fulfill many different unrelated functions and can be often reused in multiple pathways to produce different physiological outcomes. This phenomenon is called moonlighting and confers a huge complexity to interactomes (Tompa et al., 2005).

IDPs peculiar biochemical features and conformational adaptability confer to them important properties that are essential to exploit the function of molecular hubs in protein interaction networks. These properties are:

- **enrichment in post-translational modifications (PTMs) sites.** In flexible IDPs peptide consensus sites are easily reached by the catalytic sites of modifying enzymes to introduce or remove a PTM. For example, as regard phosphorylation, recent computational studies of kinase-substrate network in yeast have shown that IDPs are enriched for sites that can be modified with respect to structured proteins. These data suggest an important role of this modification in fine-tuning IDPs function and half-life (Gsponer et al., 2008; Collins et al., 2008);

- **use of PTMs.** IDPs flexibility facilitates the binding of different enzymes such as kinases, phosphatases, acetyltransferases, deacetylases, methyltransferases, ubiquitin ligases and others. Single modification or combination of diverse PTMs create “marks” that are
recognized by different binding effectors which specifically mediate the downstream outcome (Gsponer and Babu, 2009);

- **protein population shift and conformational selection due to PTMs.** PTMs serve as important modulators of the conformational energy landscape of IDPs, which in turn regulates their binding preferences (Figure 1.7) (Ma and Nussinov, 2009);

- **scaffolding and recruitment of different binding partners.** IDPs can act as scaffold proteins that serve as backbone for the regulated assembly of different signaling partners and anchor the macromolecular complexes to specific sub-cellular localizations (Gsponer and Babu, 2009). In fact, IDPs possess a high capture radius (*fly-casting model* - Levy et al., 2004) and a large interaction surface area that allow high intermolecular association rates and simultaneous binding to several protein, DNA, or RNA molecules.

![Figure 1.7 The energy landscape of IDPs conformations, the effects of post-translational modifications and their relationship to function. A. The x-axis depicts the conformational ensemble. Conformations that are geometrically similar lie close to each other. The y-axis depicts the population size. B and C. The dynamic conformational selection of IDPs (IUPs) through PTMs and molecular interactions. Two PTMs are shown: phosphorylation (P) and acetylation (K). Both result in conformational selection and population shift in the ensemble of structures. Many structural clusters coexist for a seemingly unstructured protein. PTMs create allosteric perturbation sites, propagating through the structures like waves. The observable outcome is a shift in the distribution of the population, biasing the ensemble towards conformers whose structures are favored to bind specific partners. A specific conformation is selected by a binding partner with best complementarity to the IDP binding site (Ma and Nussinov, 2009).](image-url)
Chapter 1

Introduction

1.9 Tight regulation of IDPs: from transcript synthesis to protein degradation
Because of their unusual structural and important functional properties, the presence of IDPs in a cell may need to be carefully monitored. In fact, altered abundance of IDPs is associated with several disease conditions (par. 1.10).

Integrated analysis of multiple large-scale datasets highlighted that mRNAs encoding highly IDPs are generally less abundant than transcripts encoding structured proteins and this difference resides in a lower half-life, and in particular in decay rates. As regard to this, polyA tail length and interaction with specific RNA binding proteins seem to be important factors in modulating the stability of transcripts encoding IDPs (Gsponer et al., 2008; Ma et al., 2009).

Moreover, IDPs are also finely regulated at the protein level. In fact, IDPs have a lower rate of protein synthesis and a lower protein half-life with respect to structured proteins. The availability of many IDPs is fine-tuned by PTMs and interactions with other factors that regulate their accessibility to the degradation machinery (Gsponer et al., 2008).

Therefore, both transcriptional and post-translational mechanisms are actively involved in controlling IDPs presence in a cell and can promote increased abundance and/or longer half-life upon the presence of specific stimuli.

1.10 The biological “cost” of disordered proteins: the D² concept
Since IDPs are among major cellular regulators, recognizers and signal transducers, when misregulated they became key players in the development of pathological conditions.

Many proteins involved in human diseases such as cancer, Parkinson’s disease and other synucleinopathies, Alzheimer’s, prion diseases, diabetes, and cardiovascular diseases are either completely disordered or contain long disordered regions. These diseases originate from molecular misidentification, misregulation, and missignaling due to increased abundance or conformational alterations of causative IDPs.

Such evidence gives rise to the “disorder in disorders or D² concept” (Uversky et al., 2008).

Moreover, since IDPs display peculiar interacting properties and are key players in cellular interaction networks, they became very attractive targets for the development of drugs aimed at interfering with their protein-protein and protein-nucleic acid interactions (Dunker et al., 2008bis).

Among the most known IDPs involved in diseases there are, for example, α-synuclein, tau protein, p53, BRCA1 and HMGA.
1.11 Intrinsically disordered HMGA proteins are key factors in cancer development

The High Mobility Group A (HMGA) proteins that are considered to be prototypical IDPs (Dunker et al., 2005) are non-histone chromatin factors known to play crucial roles in cancer onset and progression.

HMGA proteins (HMGA1 and HMGA2) are defined oncofetal factors because they are highly expressed in undifferentiated rapidly proliferating cells during embryonic development and neoplastic transformation whereas they are completely absent or barely detectable in normal differentiated cells.

Elevated levels of HMGA proteins are diagnostic markers of both neoplastic transformation and increased metastatic potential of a large number of different types of human cancers.

Generally, HMGA overexpression is a feature of malignant tumors and is causally related to neoplastic cell transformation, whereas rearrangements of HMGA genes are a feature of most benign human mesenchymal tumors (Fusco and Fedele, 2007).

1.11.1 HMGA proteins in malignant tumors

HMGA have been related to cancer since their first identification in rat thyroid cells transformed by retroviruses. In fact, HMGA level of expression was found much higher in transformed cells compared to normal cells in which they resulted practically absent (Giancotti et al., 1985), moreover further studies validated the association of these proteins with highly malignant phenotypes (Giancotti et al., 1987). Subsequently, several evidences indicated that HMGA proteins have oncogenic activities, being causally involved in neoplastic transformation (Berlingieri et al., 1995; Scala et al., 2000; Berlingieri et al., 2002).

Studies on rat prostatic cancer cells suggested a role for HMGA proteins in the progression of malignant tumors. Comparing the Dunning R-3327-H benign rat prostatic tumor cell line with the highly aggressive MatLyLu cell line, HMGA turned out to be up-regulated in the latter and HMGA overexpression correlated with metastatic ability rather than with growth rate (Bussemakers et al., 1991). Moreover, HMGA1 expression has been shown to increase in high grade prostate tumors (Tamimi et al., 1993).

Subsequently, many cancer types have been analyzed by immunohistochemistry for HMGA expression and it has been demonstrated that many human neoplasias, including thyroid, prostatic, cervical, colorectal, pancreatic, gastric, breast, lung and ovarian carcinoma and head and neck tumors, show a strong increase of HMGA1 proteins. HMGA2 overexpression has been reported in pancreatic and lung carcinomas, squamous carcinomas of the oral cavity, in bladder cancer, and colorectal cancers. Importantly, HMGA proteins expression in human tumors is associated with a
high malignant phenotype and is a poor prognostic index as their overexpression often correlates with the presence of metastasis and with a reduced survival (Fusco and Fedele, 2007; Fedele and Fusco 2010 and references therein; Wang, et al., 2011; Yang et al., 2011).

In addition to solid tumors, HMGA are implicated also in the development of haematological neoplasias. In fact, HMGA1 are overexpressed in several leukaemic cell lines and most human leukaemias and, interestingly, they have been found to play crucial roles in megakaryocyte differentiation (Pierantoni et al., 2003). Moreover, HMGA2 rearrangements following diverse chromosomal translocations have been observed in several myeloid malignancies and turned out to be linked with poor prognosis (Odero et al., 2005). Further in vitro and in vivo studies confirmed these data and demonstrated that HMGA are involved as causal factors in inducing haematological neoplasias: HMGA overexpression induces tumorigenicity, whereas their decreased expression abrogates transformation (Wood et al., 2000), and mouse models overexpressing both Hmga1 or Hmga2 develop lymphomas (Baldassarre et al., 2001; Xu et al., 2004).

1.1.1.2 HMGA proteins in benign tumors
HMGA2 gene is one of the most rearranged genes in benign human tumors mainly of mesenchymal origin, including lipomas, uterine leiomyomas, pulmonary chondroid hamartomas, fibroadenomas of the breast, endometrial polyps, pleomorphic adenomas of the salivary glands and vulvar aggressive angiomyxoma. Moreover, also HMGA1 gene has been found rearranged in several of these tumors (Fusco and Fedele, 2007 and references therein).

The rearranged HMGA2 gene encodes for a truncated or a chimeric HMGA2 protein lacking its C-terminal portion. Particularly, the preferred chromosomal rearrangement in lipomas gives rise to a translocation where HMGA2 is fused to the LIM domain of LPP (lipoma-preferred partner) gene (Petit et al., 1996). Among HMGA2 fusion partners there are also some tumor-suppressor genes that lose their function upon translocation, however several evidences have suggested that HMGA2 truncation rather than disruption or misregulation of the fusion partner is the crucial event in the process of tumorigenesis. Moreover, also mouse models confirm the role of truncated HMGA2 in promoting lipomatosis, lipomas, and other benign mesenchymal tumors including fibroadenomas of the breast and salivary glands adenomas (Fusco and Fedele, 2007 and references therein).

Rearrangements and overexpression of the HMGA2 gene have also been described in human pituitary adenomas (Finelli et al., 2002; Pierantoni et al., 2005). Hmga2 and Hmga1 overexpression in mice also causes pituitary tumorigenesis, but genes are not rearranged (Fusco and Fedele, 2007).
1.11.3 HMGA proteins act also as tumor suppressors

Surprisingly, HMGA have been found to display also anti-oncogenic properties. In fact, both heterozygous and homozygous *Hmgal*-null mice develop age-dependent splenomegaly associated with myelo-lymphoproliferative disorders, resembling various human B-cell lymphomas. The onset of a malignant phenotype in these mice might be due to the loss of regulation of T-cell specific cytokines, resulting in a B-cell expansion. These data unveil an unsuspected haploinsufficient tumor-suppressor role for HMGA proteins (Fedele et al., 2006).

1.12 HMGA protein family

HMGA proteins belong to the High Mobility Group (HMG) super-family of proteins together with HMGB and HMGN. HMG proteins are abundant heterogeneous nonhistone components of chromatin that act as “architectural transcription factors” modulating nucleosome and chromatin structure and orchestrating the efficient participation of other proteins in vital nuclear activities such as transcription, replication and DNA repair.

HMGA, HMGB and HMGN families are distinguished from each other by their unique DNA-binding motifs (*AT-hook, HMG-Box and Nucleosomal-binding domain*, respectively), by their preferred binding substrates, by the changes they induce in their substrates and by the different subset of cellular processes they influence (Reeves, 2010).

HMGA protein family consists of three main members: HMGA1a, HMGA1b, which are alternatively spliced products of the same gene (*HMGAI*), and HMGA2, which is the product of a different but highly related gene (*HMGAI2*) that has a high amino acid sequence homology (about 55%) with HMGA1 proteins (Sgarra et al., 2004). HMGA proteins occur widespread in nature and homologues for the human HMGA proteins have been found in yeast, insects, plants and birds, as well as in all mammalian species examined (Cleynen and Van de Ven, 2008).

Human *HMGAI* and *HMGAI2* genes are located on chromosome 6 (6p21) and 12 (12q13-15), respectively. These genes are very different in length; in fact *HMGAI* spans 10.1 kb whereas *HMGAI2* spans more than 220 kb with a third intron of about 140 kb. Notably, this intron is the central molecular target of almost all HMGA2 chromosomal aberrations resulting in truncation or generation of fusion genes that are found in many tumors (Fusco and Fedele, 2007).

HMGA are small proteins of about one hundred amino acid residues with a modular sequence organization (Figure 1.8). In fact, all HMGA proteins are characterized by the following domains:

- three highly positively charged DNA binding domains (DBDs) called *AT-hooks*, since they bind the minor groove of AT-rich DNA stretches;
- a protein-protein interaction domain, which partially overlaps with the second DBD;
• a highly acidic C-terminal tail that is constitutively phosphorylated with no clear functional assignment (Sgarra et al., 2008).

Figure 1.8 Schematic representation of HMGA proteins. Amino acid sequence and functional domains of HMGA proteins are reported: AT-hook (DNA binding domain); P/P interaction (Protein/Protein interaction domain); Acidic tail (C-terminal domain). HMGA1a sequence comprises all 106 amino acid residues indicated in the HMGA1 sequence, whereas in HMGA1b isoform the 11 amino acid stretch indicated as “splicing region” is absent. HMGA2 sequence consists of 108 amino acid residues.

HMGA1 proteins differ for 11 amino acid residues between the first and the second DBD that are absent in HMGA1b isoform. Whereas, HMGA2 differs from HMGA1a and HMGA1b mainly for 25 N-terminal amino acid residues and a short peptide of 12 residues between the third AT-hook and the acidic C-terminal tail called spacer region that does not have any equivalent in HMGA1 proteins.

As shown in Figure 1.8 the three AT-hooks are differentially spaced along the protein molecules resulting in an interactive modular system constituted by a set of three proteins able to establish interactions with differently spaced AT-rich DNA regions (Sgarra et al., 2004). All the three AT hooks contribute to DNA/chromatin-binding in vivo. However, the first and the second AT hooks are the main mediators of DNA binding, whereas the third AT hook plays only a secondary, but nevertheless cooperative part. For proper binding to chromatin in vivo, two functional AT hooks are necessary and sufficient and this is in accordance with several in vitro investigations (Harrer et al., 2004).

In addition to recognize the structure of the minor groove of AT-rich DNA as a target of binding, HMGA proteins also recognize and bind to non-B-form DNAs with unusual structural features such as four-way and three-way junctions, bent and supercoiled DNA, base-unpaired regions of AT-rich DNA and distorted or flexible regions of DNA on isolated nucleosome core particles (Reeves, 2010).

Moreover, recent evidences show that HMGA AT-hooks are involved also in RNA binding, both in a sequence-specific manner during splicing events (Manabe et al., 2003) and in a structure-specific
manner, with the recognition of G-rich RNA oligonucleotides that potentially form G-quadruplex and of snRNA loops secondary structure (Norseen et al., 2008; Eliebrecht et al., 2010 and 2011).

HMGA have a high content of K, R, E, D and P amino acid residues typical of IDPs, which are clustered at the level of their DBDs (R, K and P) and of their acidic tail (E and D) (Sgarra et al., 2010). Both CD and NMR studies (Lehn et al., 1988; Evans et al., 1995) indicated that HMGA proteins lack structure, exhibiting a random coil-like conformation over its entire length. The atypical electrophoretic mobility of this molecule (Lund et al., 1983) also suggested a high content of extended structure.

The only structural information regarding these IDPs concerns the region comprising the second and the third DBD, which was shown to undergo a disorder-to-order transition upon binding to DNA. In fact this region assumes a planar, crescent-shaped configuration when specifically bound to the minor groove of short AT-rich stretches of DNA (Reeves and Nissen, 1990; Evans et al., 1995; Huth et al., 1997).

The lack of any conformational information on HMGA bound to protein partners does not exclude folding upon binding mechanisms typical of IDPs.

1.13 Transcriptional and post-transcriptional control of HMGA expression

HMGA gene expression can be regulated in response to a large number of cellular stimuli and transcriptional HMGA deregulation is probably the major mechanism involved in the aberrant expression of HMGA in tumors.

HMGA expression can be induced by regulative cascades activated by many factors known to be intimately associated with the promotion of tumor progression and metastasis (Cleynen and Van de Ven, 2008; Reeves, 2001), among which:

- cell growth factors (EGF, PDGF, FGF, TGF-α, and TGF-β);
- tumor-promoting agents (TPA and phorbol esters);
- harmful environmental factors (endotoxin, hypoxia, morphine, and viral infections)
- differentiation factors (retinoic acid, butyrate, and calcium ionophores);
- cytokines (IL-1β and INF-β);
- oncogenes (c-Myc and Ras).

Moreover, among transcription factors, AP-1 (Ogram and Reeves, 1995; Cleynen et al., 2007), c-Myc (Wood et al., 2000bis), MYCN (Giannini et al., 2005), and SP1 (Cleynen et al., 2007) have been described to directly target HMGA1 promoter.

Several oncogenic pathways inducing HMGA genes have been described:
- **TGF-β/Smad** signaling induces *HMGA2* transcription in mammary epithelial cells to elicit epithelial-mesenchymal transition (Thuault et al., 2006);
- **Ras/MAPK** pathway induces *HMGA2* to efficiently repress glucocorticoid receptor/dexamethasone-stimulated transcription of α-subunit of epithelial Na⁺ channel gene and STAT3-mediated transactivation in salivary rat epithelial cells (Zentner et al., 2001). In addition, the activation of Ras/MAPK signaling induces the expression of *HMGA1* in cancer cells with a mechanism that has been suggested to require a complex cooperation between SP1 family members and AP-1 factors (Cleynen et al., 2007). Finally, *HMGA2*, but not *HMGA1*, is induced by this pathway in human pancreatic cancer cells where HMGA2 is involved in the maintenance of Ras-induced epithelial-mesenchymal transition by stimulating Snail expression (Watanabe et al., 2009);
- **Wnt/β-catenin** pathway induces *HMGA1* expression through c-Myc activation resulting in the maintenance of proliferation of gastric cancer cells (Akaboshi et al., 2009).

At a post-transcriptional level HMGA expression is regulated by miRNAs. The tumor suppressor miRNA let-7 represses *HMGA2* by binding to multiple target sites on 3’UTR of *HMGA2* transcript thus negatively regulating protein synthesis (Lee and Dutta, 2007; Mayr et al., 2007). Moreover, *HMGA2* was found to be a target for miRNA-98 in head and neck squamous cell carcinomas (Hebert et al., 2007) and a putative target of miR-196a-2, a miRNA regulated by HMGA1 (De Martino et al., 2009).

Importantly, the loss of 3’UTR that occurs during chromosomal aberrations in mesenchymal tumors could activate *HMGA2* oncogene expression that is normally repressed by the action of miRNAs (Figure 1.9) (Fusco and Fedele, 2007).

![Figure 1.9 Loss of microRNA-mediated repression of HMGA2 in mesenchymal tumours. A. The let-7 microRNA binds to the 3’ region of the HMGA2 gene and thus downregulates HMGA2 expression by inhibiting translation and/or causing the degradation of the mRNA. B. Following truncation or fusion with ectopic sequences, HMGA2 loses the sites previously recognized by let-7, resulting in increased HMGA2 mRNA levels. Here, HMGA2 is in blue and its fusion partner is in red (Fusco and Fedele, 2007).](image-url)
Recently, also HMGA1 turned out to be post-transcriptionally regulated by miRNA, in fact miR-16 interacts with 3’UTR of *HMGA1* transcript downregulating its expression (Kaddar et al., 2009).

### 1.14 HMGA molecular network

HMGA ID proteins are highly connected *hubs* in protein-protein molecular networks and HMGA up-to-date interactome counts more than one hundred different factors. Most of these factors have been identified by proteomic approaches, but for some of them also functional interaction data are available (Sgarra et al., 2010).

Following the *guilt-by-association principle* claiming that “clues to the function of a protein can be obtained by seeing whether it interacts with another protein of known function” (Oliver, 2000), HMGA proteins turned out to be involved in many different biological processes that can be associated to the five networks listed below (Sgarra et al., 2010).

- **Transcriptional regulation.** About forty different transcription factors (TFs) interact with HMGA proteins and for the vast majority of these interactions a function of HMGA proteins in modulating TF activity has been described. Importantly, HMGA interacts with (i) p53 interfering with its oligomerization thus deregulating p53-dependent p21, Bax and Mdm2 gene expression and impairing the correct execution of the apoptotic process (Frasca et al., 2006; Pierantoni et al., 2006); (ii) pRb displacing HDAC1 from pRB/E2F1 complex and thereby alleviating the transcriptional repressive effect of this enzyme and favoring cell proliferation (Fedele et al., 2006); (iii) Smads modulating their binding activity in the transcriptional activation of Nkx2.5 gene during cardiogenesis (Monzen et al., 2008). Moreover HMGA and Smads co-regulate the transcriptional network necessary for the TGF-β-dependent epithelial-mesenchymal transition (Thuault et al., 2006); (iv) nuclear receptors enhancing nuclear hormone-mediated transcription (Nagpal et al., 1999; Massaad-Massade et al., 2002 and 2004).

- **RNA processing.** HMGA can bind both to RNA sequences and RNA binding proteins. In particular, HMGA1 mediates aberrant splicing of Presenilin-2 pre-mRNA via direct protein-protein interaction with U1 snRNP 70K protein (Norseen et al., 2008; Manabe et al., 2003 and 2007).

- **DNA repair and chromatin remodeling and dynamics.** HMGA proteins bind several factors known to be involved in regulating chromatin accessibility, DNA replication, transcription and DNA repair. In fact, HMGA have been shown to interact with (i) core histones and nucleosome core particles probably being involved in the modulation of the chromatin fiber structure and plasticity. In fact, it has been hypothesized that HMGA can
participate at nucleosome disassembly in order to facilitate the binding of TF to DNA (Sgarra et al., 2010, Reeves and Wolffe, 1996, Reeves and Nissen, 1993); (ii) origin of replication complex (ORC) subunits recruiting them to heterocromatin domains in order to facilitate DNA replication (Norseen et al., 2008; Thomae et al., 2008); (iii) several factors that are members of co-activators/co-repressor complexes modulating transcription (such as RBBPs, RUVBLs, CBP and PRMT6) (Sgarra et al., 2005 and 2006); (iv) proteins involved in several DNA repair mechanisms such as Ku70/Ku80, APE1/REF-1 and ATM (Sgarra et al., 2005 and 2008; Pentimalli et al., 2008). The function of the direct interaction among HMGA and DNA repair factors has not been clearly elucidated yet; however several evidences unveil a transcriptional role for HMGA in different DNA repair processes. HMGA proteins seem to influence cellular DNA damage response both in a negative or positive manner depending on the different DNA-repair pathways and biological contexts considered (Sgarra et al., 2010; Bullerdiek and Rommel, 2010; Li et al., 2009; Summer et al., 2009; Reeves and Adair, 2005).

- **Modifying enzymes.** Fifteen different modifying enzymes have been found to act on HMGA proteins, and a stable interaction has been demonstrated for eight of them (PRMT6 - arginine N-methyltransferase; HIPK2, Nek2 and ATM - kinases; p300 - histone acetyltransferase, Ubc9 - SUMO-conjugating enzyme, SDCCAG10 - peptidyl-prolyl cis-trans isomerase, and METAP2 - methionine aminopeptidase). These enzymes can directly modulate HMGA interactions with DNA or other molecular partners by adding PTMs or they can be recruited by HMGA proteins at specific sites where they can exploit their modifying functions towards other proteins (Sgarra et al., 2010).

- **Others.** This category comprises all those proteins for which it is difficult to assign a function for the interaction or those which could have too many functions. Among these multifunctional proteins nucleophosmin and nucleolin are of particular interest, however no functional data are available yet (Sgarra et al., 2005 and 2008).

Figure 1.10 summarizes some of the most relevant functional outcomes of the protein-protein interaction network of HMGA proteins. The modulation of HMGA expression through the activation of different pathways, such as Ras/MAPK and TGF-β or through the action of miRNAs, affects the activity of cellular factors playing key roles in several processes (apoptosis, cell proliferation, etc.).
Figure 1.10 Influence of HMGA proteins molecular network on key cellular processes. HMGA proteins, by protein-protein interactions, modulate the activity of key factors affecting several cellular processes involved in embryonic development and neoplastic transformation (Sgarra et al., 2010).

1.15 Mechanisms of action of HMGA proteins

The role of HMGA proteins in cell transformation is essentially based on their ability to down- or up-regulate expression of genes that have a critical role in control of cell proliferation and invasion. HMGA architectural transcription factors do not have transcriptional activity per se and three main mechanisms of action have been described:

- **general action on chromatin structure.** HMGA proteins have been shown to be important elements associated with DNA matrix- and scaffold-associated regions (SARs or MARs). SARs are specific segments of genomic DNA that have a high affinity for the nuclear matrix and that are enriched in AT sequences (> 70%). These regions anchor chromatin to the nuclear scaffold and organize topologically independent DNA domains which have functional roles both in DNA replication and transcription (Galande, 2002). The binding of
HMGA proteins to these regions de-repress transcription by displacement of histone H1 from DNA. In fact, HMGA can compete with histone H1 for the same AT-rich regions on SARs and can interfere with the highly cooperative binding of H1 to DNA (Zhao et al., 1993) (Figure 1.11 A).

However, despite HMGA well assessed role of transcriptional anti-repressor molecules, surprisingly they have been shown to be preferentially enriched in transcriptionally inactive, condensed chromatin and in chromosomes during mitosis (Harrer et al., 2004). Moreover, HMGA proteins accumulate in condensed chromatin of senescent cells constituting essential structural elements of senescence-associated heterochromatic foci (SAHFs) and actively cooperate with tumor suppressors for SHAFs formation and for repression of genes associated with cell proliferation (Narita et al., 2006). These evidences suggest an anti-oncogenic role for HMGA proteins. Therefore, HMGA are multifunctional architectural factors regulating both positively or negatively transcription and chromatin compactness in dependence of the molecular context they are embedded in.

- **specific action at the level of target promoter/enhancer regions.** HMGA proteins have a crucial role in assembling and modulating multiprotein stereospecific complexes, also called enhanceosomes, at the level of promoter/enhancer regions of their target genes. HMGA molecules can directly bind to DNA modifying its conformation and consequently facilitating the binding of a group of other TFs or they can interact with both DNA and TFs generating a macromolecular complex bound to DNA (Figure 1.11 B) (Sgarra et al., 2004). HMGA proteins have been shown to participate in this way in the regulation of many genes, for example IFN-β (Thanos and Maniatis, 1992), IL-2 receptor α (Reeves et al., 2000; Reeves and Beckerbauer, 2001), IL-4 (Klein-Hessling et al., 1996) and IR (Brunetti et al., 2001; Foti et al., 2003).

HMGA1a role in INF-β gene expression is the best characterized and most representative mechanism of action for these proteins. When HMGA1a is not bound at the level of INF-β promoter, transcriptional activators NF-kB, ATF-2/c-Jun and IRFs recognize their binding sites with low affinity due to the unfavorable intrinsic DNA curvature of the promoter. Importantly, the binding of HMGA1a molecules at the level of two specific AT-regions (PRDII/NRDI and PRDV) unbends the DNA, thus favoring the binding of transcriptional activators on their own consensus sequences. This allosteric effect on DNA results in a significant enhancement of activators’ promoter binding affinity in the absence of protein-protein interactions with HMGA. Macromolecular assembly is then completed by protein-
protein interactions between all the components, thus leading to a remarkably stable nucleoprotein structure, the so called enhanceosome. This new structure is responsible for the modification and remodeling of a nucleosome that masks the TATA-box and, consequently, transcription can start. Importantly, INF-β enhanceosome stability is finely modulated by HMGA PTMs (par. 1.16.2) (Thanos and Maniatis, 1992; Yie et al., 1999; Munshi et al., 2001).

- **modulation of transcription factors affinity towards DNA.** HMGA can also influence gene transcription in a DNA independent manner, through direct protein-protein interactions with TFs by inducing changes in their DNA binding affinities (Figure 1.11 C) (Sgarra et al., 2004).

For example, HMGA1a binds to serum response factor SRF enhancing its transcriptional activity towards c-Fos and SM22α genes (Chin et al., 1998) whereas HMGA2 associates with E1A-regulated transcriptional repressor p120E4F interfering with its binding to the cyclin A promoter thus activating cyclin A transcription and favoring progression of cell cycle (Tessari et al., 2003). These examples underline again a role for HMGA in promoting cell cycle progression and cell proliferation.

All HMGA mechanisms of action are finely tuned by PTMs that modulate their multi-interacting properties with both DNA and proteins (Figure 1.11) (Sgarra et al., 2004).
1.16 Fine tuning of HMGA binding by post-translational modifications

As previously shown, HMGA proteins actively participate in cell transformation by affecting multiple and different proliferative and regulative key nodes both regulating their gene expression and protein functions.

HMGA multifunctionality is probably due to their natively disordered status, their modular organization in functional domains and their fine tuning by PTMs addition. In fact, HMGA are among the most extensively modified proteins in the nucleus suggesting the existence of a HMGA PTMs code analogous to the histone code (Zhang and Wang, 2008 and 2010).

In particular, HMGA1 protein modifications have been largely investigated for over twenty years and it turned out that about one third of HMGA1a and HMGA1b amino acid residues are potentially affected by PTMs (summarized in table 1 and schematized in Figure 1.12). Some of these modifications are considered dynamic since they have been demonstrated to regulate specific processes, therefore involving only few HMGA molecules inside the nucleus, otherwise others are considered constitutive since they are assumed to occur on HMGA proteins independently from the cellular context.

HMGA are regulated by a heterogeneous series of PTMs, such as phosphorylation, acetylation, methylation, ADP-rybosilation, formylation and SUMOylation.

1.16.1 Phosphorylation

Since their discovery all HMGA proteins have been found to be among the most heavily phosphorylated proteins in the nucleus and for this reason they have been recognized as phosphoproteins (Lund et al., 1985; Elton and Reeves, 1986).

Further researches revealed that the acidic C-terminal tail of HMGA1 and HMGA2 proteins is constitutively phosphorylated in vivo by Casein Kinase 2 (CK2) at the level of multiple S/T residues (Palvimo et al., 1989; Goodwin, 1998). CK2 is a ubiquitous serine/threonine-selective protein kinase. Looking at the extreme variety of its targets and partners it is suggested that all or nearly all cellular functions are more or less directly subjected to modulation by CK2. In particular, CK2 appears to play a central role in the regulation of gene expression and protein synthesis/degradation as a mediator of stress stimuli and as a powerful survival agent whose general strategy is to counteract programmed cell death by impinging at different levels on the complex apoptotic machinery. Moreover, CK2 is abnormally high, both in terms of protein and activity, in a wide variety of tumors (Pinna and Allende, 2009).

As regard HMGA, CK2 phosphorylation sites have been mapped both in vitro and in vivo only for HMGA1 proteins. CK2 catalyzes the phosphorylation of S98, S101 and S102 in HMGA1a and the
corresponding sites in HMGA1b (i.e., S87, S90 and S91) (numbers refer to proteins without initial methionine that is always missing in HMGA; hereafter numbers will refer only to the HMGA1a isoform for simplicity) (Palvimo and Linnala-Kankkunen, 1989; Ferranti et al., 1992; Jiang and Wang, 2006). Due to the high content of negatively charged amino acid residues, the C-terminal domain is generally believed to participate in protein-protein interactions rather than binding to DNA, however, an effect of this phosphorylation on HMGA1-DNA interaction was noted. In fact, IL-4 stimulation in B lymphocytes has been shown to increase HMGA1 C-terminal phosphorylation that has been demonstrated to reduce HMGA1 binding affinity towards human germ line ε promoter \textit{in vitro} (Wang et al., 1995). These data only suggest a mechanism in which phosphorylation triggered by IL-4 or other cytokines could regulate the function of HMGA1 on gene transcription, however the functional role of CK2 constitutive phosphorylations occuring on HMGA acidic tails remains still unclear.

Moreover, no functional information is available also for the other member of HMGA family, HMGA2.

Beside constitutive C-terminal phosphorylation, numerous other signal-responding and dynamic phosphorylations have been described.

Importantly, early studies showed that HMGA1 phosphorylations of T52 and T77 are cell cycle-related; in fact these PTMs take place at G2/M-phase and cause a strong decrease of HMGA1 DNA-binding affinity. The enzyme responsible for these modifications is the \textbf{Cycline Dependent Kinase 1} (CDK1, previously known as p34\textsuperscript{cdc2}), a serine/threonine kinase that is activated by its binding with its cyclin partners and that acts as key player in cell cycle regulation (Nissen et al., 1991; Reeves et al., 1991).

---

**Figure 1.12 Scheme of post-translational modifications mapped on HMGA1 protein sequence.** All PTMs that have been mapped on HMGA1 proteins (HMGA1a and HMGA1b) are indicated on HMGA1 amino acid sequence. Modifying enzymes assignment is also reported, when available.
Table 1. Post translational modifications of HMGA1 proteins. HMGA1a and HMGA1b PTMs so far identified are listed in correspondence of the amino acid residue where they have been mapped. Modifying enzymes and functional significance are also indicated, if available. A: acetylation. M: methylation and P: phosphorylation (Maurizio et al., 2009 – Supplementary table 1 and references therein).
Phosphorylation by CDK1 occurs also on HMGA2 protein impairing the interaction of its second AT-hook with the PRDII element of INF-β gene promoter; however the exact CDK1 phosphorylation sites have never been mapped for HMGA2 (Schwanbeck et al., 2000).

Moreover, the **Homeodomain-Interacting Protein Kinase 2 (HIPK2)**, a serine/threonine kinase that have been shown to exert a potent inhibitory effect on cell growth at the G2/M phase of the cell cycle, can both bind and phosphorylate HMGA1 proteins (Pierantoni et al., 2001). Recent mapping data revealed that HIPK2 and CDK1 phosphorylate HMGA1a at the same amino acid residues (i.e. S35, T52 and T77) *in vitro*, although the two kinases exhibit different site preferences. The preference for HIPK2 phosphorylation followed the order T77>T52>S35, whereas the sequence for CDK1 phosphorylation is T52>T77>S35 (Zhang and Wang, 2007). These phosphorylations have been shown to reduce the binding affinity of HMGA1a to human germ line ε promoter; however the decrease in DNA binding affinity induced by HIPK2 phosphorylation was not as pronounced as that introduced by CDK1 phosphorylation. This evidence is in accordance with the notion that the second AT-hook in HMGA1a is more important for DNA binding than the third (Harrer et al., 2004).

Remarkably, HIPK2 is a proapoptotic activator of p53 and its cellular localization is dependent on HMGA1 expression levels. In fact, HMGA1 overexpression inhibits p53 activity by localizing HIPK2 in the cytoplasm, while HIPK2 overexpression reestablishes its nuclear localization and promotes p53-mediated apoptosis (Pierantoni et al., 2007). The major phosphorylation site induced by HIPK2 (T77) is located within HMGA1a protein-protein interaction domain that is involved in p53 binding (Pierantoni et al., 2006). Thus, it has been hypothesized that the HIPK2-induced phosphorylation of HMGA1a can impair the interaction between p53 and HMGA1a and eventually promote p53-mediated apoptosis. On the other hand, HMGA1a overexpression in cancer cells can promote HIPK2 localization to the cytoplasm, which can result in decreased phosphorylation of HMGA1a, enhanced HMGA interaction with p53, and inhibited apoptotic function of p53.

Interestingly, recent studies support an involvement of HMGA1 phosphorylation in DNA damage response. In fact, HMGA1 proteins have been identified as novel downstream targets of the **Ataxia-Telangiactesia Mutated (ATM)** kinase pathway in response to DNA damage (Pentimalli et al., 2008). ATM is a key upstream activator of cellular DNA damage response and specifically phosphorylates serine or threonine residues preceding a glutamine (also known as the SQ motif) (O’Neill et al., 2000). Indeed, HMGA1 have been demonstrated to be phosphorylated by ATM both *in vitro* and *in vivo* at the level of the SQ motif embedded in the acidic tail (Pentimalli et al., 2008). HMGA1 proteins can be phosphorylated also by **Protein Kinase C (PKC)**, a serine/threonine kinase that is activated by signals such as increases in the concentration of diacylglycerol or Ca^{2+}
and that plays important roles in several signal transduction cascades. In fact, HMGA phosphorylation by PKC is the result of the treatment of human mammary epithelial cells with phorbol esters that activate the Ca\(^{2+}\)/phospholipid pathway (Banks et al., 2000). The major sites of PKC phosphorylation on HMGA1a are T20, S43 and S63, however also T74, T75, T76, and T77 can be phosphorylated at a low level. A significant reduction of DNA-binding affinity has been found for HMGA1a protein modified by PKC. In fact PKC phosphorylation attenuates HMGA1a binding to the promoter regions of PKC\(\gamma\) and neurogranin/RC3 genes (Xiao et al., 2000).

Furthermore, Nek2, a serine/threonine kinase activated by the MAPK pathway in mouse pachytene spermatocytes, directly interacts with HMGA2 \textit{in vitro} and in mouse spermatocytes. The interaction does not depend on the activity of Nek2 and seems constitutive. On progression from pachytene to metaphase, Nek2 is activated and HMGA2 is phosphorylated in an MAPK-dependent manner. Nek2 have been shown to modify HMGA2 \textit{in vitro}, but phosphorylation sites are unknown. This phosphorylation decreases the affinity of HMGA2 for DNA and might favor its release from the chromatin. Indeed, most HMGA2 associates with chromatin in mouse pachytene spermatocytes, whereas it is excluded from the chromatin upon the G2/M progression. Because \textit{hmga2\(-/-\)} mice are sterile and show a dramatic impairment of spermatogenesis, it is possible that the functional interaction between HMGA2 and Nek2 plays a crucial role in the correct process of chromatin condensation in meiosis (Di Agostino et al., 2004).

Additionally, several other \textit{in vivo} studies suggested contrasting roles for HMGA phosphorylation with respect to the highly dynamic HMGA-chromatin interaction. In fact, HMGA1a typical hyperphosphorylation-dephosphorylation steps observed during apoptosis progression suggest that an early increase in HMGA phosphorylation can be involved in the displacement from chromatin to allow DNA digestion, whereas subsequent HMGA dephosphorylation can be implicated in the formation of highly condensed chromatin in apoptotic bodies (Diana et al., 2001). Contrastingly, FRAP experiments revealed that HMGA increased phosphorylation correlates with an elevated residence time of HMGA1a in heterochromatin and chromosomes with respect to euchromatic regions (Harrer et al., 2004).

Despite these discordances, HMGA proteins and their phosphorylations appear to be directly involved and play crucial roles in modulating local and global changes of chromatin structure.

\subsection*{1.16.2 Acetylation}

Other than the acetylation at N-termini that affects almost 85\% of eukaryotic proteins (Polevoda and Sherman, 2000), HMGA1 proteins are post-translational acetylated on several lysine residues \textit{in vivo} (Ferranti et al., 1992; Banks et al., 2000; Edberg et al., 2005; Zhang et al., 2007). The most
notable example of the biological significance of HMGA1 acetylation lies in its dynamic control of virus-induced expression of INF-β gene (Munshi et al., 1998; Munshi et al., 2001). The accurate execution of INF-β transcriptional switch (par. 1.15) depends on the ordered and highly controlled acetylation of HMGA1 by two histone acetyltransferases: CREB-binding protein (CBP) and p300/CBP-associated factor/GCN5 (PCAF/GCN5), which acetylate HMGA at distinct lysine residues, inducing opposite effects on enhanceosome stability. In particular, the acetylation of HMGA1 by PCAF/GCN5 at K70 potentiates transcription by stabilizing the enhanceosome, recruiting further activator complexes, and preventing acetylation by CBP, whereas the acetylation of HMGA1 by CBP (in complex with Pol II) at K64 destabilizes the enhanceosome leading to the transcriptional turn-off of the human INF-β gene (Figure 1.13) (Munshi et al., 2001; Struhl, 2001).

A recent study revealed that PCAF and p300 can also acetylate K14, K66, and K73 in HMGA1 beside K64 and K70, both in vitro and in vivo (Zhang et al., 2007).

**Figure 1.13 The IFN-β enhanceosome switch.**

A. In response to viral infection, the enhanceosome-consisting of activators (unlabeled colored ovals) and HMGA1 (HMG1(Y)) (green circle)-is assembled between nucleosomes (grey). The enhanceosome recruits the PCAF/GCN5 complex, which acetylates both HMGA1 at K71 (numbers here refer to HMGA1 sequence with initial methionine - K70 in the text) and histone proteins in the nucleosome (arrows).

B. The stabilized enhanceosome sequentially recruits the CBP-Pol II enzyme complex and then the Swi/Snf nucleosome remodeling complex. C Swi/Snf disrupts the positioned nucleosomes flanking the enhanceosome, leading to binding of the TFIID transcription factor and transcription of the gene. Acetylation at K71 inhibits CBP acetylation at K65 (K64 in the text).

D. CBP acetylates HMGA1 at K65, leading to disruption of the enhanceosome and the cessation of transcription (Struhl, 2001).
1.16.3 Methylation
HMGA1 proteins can be methylated at several residues located principally within their AT-hooks, indicating a potential role in modulating HMGA DNA binding activities. However, clear functional assignments for these PTMs have never been uncovered yet.

Arginine mono-methylation of HMGA1a at R25 is a distinct feature of several tumor cells and the methylation level increases when cells are induced to undergo apoptosis (Diana et al., 2001; Sgarra et al., 2003 and 2003bis). Moreover, R25 was found also both mono- and di-methylated in PC-3 human prostate cancer cells where both isoforms of arginine di-methylation, i.e., symmetric and asymmetric di-methylations, were detected (Zou and Wang, 2005).

R25 methylation is the most abundant methylation of HMGA1a protein in tumor cells, but the enzyme responsible for this modification hasn’t been clearly assigned yet. In vitro studies demonstrated that R25 can be modified by the protein arginine methyl-transferase 1 (PRMT1). Since PRMT1 is the most abundant PRMT in the nucleus it has been suggested that it can be responsible for HMGA R25 methylation also in vivo (Zou et al., 2007). However, PRMT1 produces only asymmetric di-methylarginine in proteins, whereas both the symmetric and asymmetric forms of di-methylations were observed in vivo at R25 (Zou and Wang, 2005). Thus, other PRMTs catalyzing the symmetric di-methylation of HMGA1a have to be identified.

Moreover, HMGA can be methylated also by PRMT6 mainly at R57 and R59, and at a low level at R83 and R85 both in vitro and in vivo (Miranda et al., 2005; Sgarra et al., 2006) and by PRMT3 at R23 in vitro (Zou et al., 2007).

Importantly, HMGA1 binding to AT-rich double stranded DNA substantially decreases the methylation efficiency of all three PRMTs toward HMGA1 proteins, at least in vitro (Zou et al., 2007).

Beside arginine methylation, HMGA1 proteins can be methylated also on lysine residues. A study examining HMGA1 PTMs in cancerous and normal human breast tissues unveiled that K30 and K54 can be mono-methylated in both metastatic and primary tumors (Zou and Wang, 2007).

1.16.4 Other minor modifications: poly(ADP-rybosyl)ation, SUMOylation, and formylation.
HMGA proteins can be poly(ADP-rybosyl)ated in mouse Lewis lung carcinoma cells in presence of Ca$^{2+}$ ions and this modification has been suggested to be part of the general poly(ADP-ribosyl)ation process that accompanies DNA damage in apoptosis (Giancotti et al., 1996).

HMGA2 has been shown to undergo SUMOylation, which can lead to the degradation of promyelocytic leukemia (PML) protein via the ubiquitin-proteasomal pathway. Importantly, arsenic
trioxide treatment stimulated HMGA2 SUMOylation, leading to the formation of HMGA2 nuclear foci surrounding PML nuclear bodies and the stimulation of PML degradation (Cao et al., 2008).

Finally, a study aimed at identifying sites of formylation in histones and other nuclear proteins from a variety of cell types and from human and mouse tissues identified formylation sites in all proteins of HMG super-family. As regard HMGA, only HMGA1 turned out to be formylated on K30 (Wisniewski et al., 2008).

1.17 HMGA PTMs in different stages of neoplastic progression

Only few works addressed the study of HMGA PTMs in cells at different stages of neoplastic transformation. These studies aimed at correlating HMGA PTMs and tumor aggressiveness and were all performed in breast cancer cells or tissues.

HMGA PTMs were analyzed in two human breast epithelial cell lines, MCF-7 and MCF-7/PKC-α representing different stages of neoplastic progression. MCF-7 cell line exhibits many characteristics of normal breast epithelial cells, whereas the MCF-7/PKC-α cell line, a derivative of MCF-7 expressing a transgene coding for PKC-α enzyme, is both malignant and highly metastatic.

HMGA1b protein was found more highly modified than HMGA1a in both cellular systems. Importantly, HMGA1b in MCF-7/PKC-α possesses a unique constellations of phosphorylations, methylations, and acetylations not found on HMGA1a in both cellular systems. Strikingly, HMGA1b from MCF-7/PKC-α turned out to have a higher nucleosome core particle binding affinity than the less phosphorylated HMGA1b from MCF-7 (Banks et al., 2000).

Another report on three MCF-7 cell lines with different metastatic potential focused on HMGA1a and showed that this protein displays increased levels of acetylation and R and K methylation in more metastatic cells and that K45 modification (none, acetylation or dimethylation) varies depending on the metastatic potential of the cells (Edberg et al., 2004).

Moreover, analysis of HMGA1 PTMs in human breast tumor specimens with different carcinoma progression stages and in the paired adjacent normal breast tissues confirmed that a more complex spectrum of HMGA1 PTMs (in particular methylation and phosphorylation) is correlated with a more aggressive malignancy (Zou and Wang, 2007).

All these studies investigated the different nature and sites of HMGA1 PTMs, but the relative abundances were not evaluated. Moreover, the functional role for these modifications has never uncovered and information on the possible linkage between HMGA PTMs and cell neoplastic potential is still fragmentary and needs further investigation.

In addition, no data are available for HMGA2 protein despite the growing number of evidences underlining its fundamental role in tumor development. Probably this resides on the fact that
HMGA2, differently from HMGA1, is not found in all human tumor specimens and therefore it can be more difficult to detect.

1.18 HMGA C-terminal tail and neoplastic transformation

The functional roles of the highly phosphorylated HMGA1 and HMGA2 C-terminal domains have never been discovered, however several pieces of evidence have stressed their importance. The loss of the C-terminal domain has been demonstrated to have an impact on cellular transformation, suggesting an inhibitory role in the regulation of HMGA proliferative and transforming ability. Importantly, HMGA truncation confers an increased growth rate to different cell types and HMGA deleted forms display a higher oncogenic activity with respect to full-length proteins (Pierantoni et al., 2003bis; Fedele et al., 1998; Li et al., 2007). This suggests that the lack of the C-terminal tail could enhance HMGA oncogenic activity through changes leading to alterations of their interactions with DNA, protein partners, and modifying enzymes. As regard this observation, it has been shown that: (i) in electrophoretic mobility shift assays (EMSAs) truncated HMGA2 behaves differently to full-length protein (Noro et al., 2003) and (ii) in farwestern analyses acidic tail truncation enhances HMGA binding affinity towards several protein partners (Sgarra et al., 2005).

In addition, the C-terminal tail of HMGA2 has been demonstrated to have a transactivating function in the context of Snail1 gene regulatory sequences (Thuault et al., 2008) and cells expressing C-terminal tail truncated HMGA2 have a different transcriptional profile with respect to those expressing full-length protein (Henriksen et al., 2010).

Moreover, very recently different biological effects of C-terminal truncated HMGA1 with respect to full-length forms have been observed in transgenic mice (Fedele et al., 2011).

All these data (reassumed in table 2) underline a functional and crucial modulatory role of acidic C-terminal domain towards HMGA protein activities. However, molecular mechanisms of action need to be elucidated.
<table>
<thead>
<tr>
<th>HMGA</th>
<th>Biological system</th>
<th>Functional outcome (CT vs FL)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMGA1</td>
<td>3T3-L1 (murine preadipocytic cells)</td>
<td>increased proliferation rate higher E2F1 DNA binding activity cell cycle de-regulation</td>
<td>Pierantoni et al., 2003bis</td>
</tr>
<tr>
<td>HMGA2</td>
<td>NIH3T3 (murine fibroblasts)</td>
<td>increased proliferation rate malignant transformation: soft agar growth and tumor formation in nude mice</td>
<td>Fedele et al., 1998</td>
</tr>
<tr>
<td>HMGA1b</td>
<td>Rat1a (rat fibroblasts)</td>
<td>higher clonogenicity higher colony average size</td>
<td>Li et al., 2007</td>
</tr>
<tr>
<td>HMGA2</td>
<td>none – EMSA experiment</td>
<td>alteration of HMGA2/DNA complexes mobility by establishment of protein-protein interactions</td>
<td>Noro et al., 2003</td>
</tr>
<tr>
<td>HMGA2</td>
<td>none – farwestern experiment</td>
<td>increased protein binding affinity</td>
<td>Sgarra et al., 2005</td>
</tr>
<tr>
<td>HMGA2</td>
<td>COS1 (monkey kidney SV40 transformed fibroblasts)</td>
<td>failure in transactivating Snail1 promoter</td>
<td>Thuault et al., 2008</td>
</tr>
<tr>
<td>HMGA2</td>
<td>hMSC (human mesenchymal stem cells)</td>
<td>activation/repression of a higher and different number of genes</td>
<td>Henriksen et al., 2010</td>
</tr>
<tr>
<td>HMGA1b</td>
<td>mice carrying CT Hmga1b</td>
<td>higher lipomagenic activity</td>
<td>Fedele et al., 2011</td>
</tr>
</tbody>
</table>

**Table 2. Summary of the biological effects induced by HMGA C-terminal domain deletion.** CT: C-terminal truncated HMGA; FL: full-length HMGA. The functional outcomes (CT vs FL) caused by truncated with respect to full-length HMGA in the indicated biological system are reported.
1.19 Aims of the study

Post-translational modifications (PTMs) of proteins expand nature’s amino acid repertoire (Walsh, 2006). The addition of PTMs, such as phosphorylation, methylation, acetylation, SUMOylation, formylation, to amino acid residues constitutes an extremely important cellular control mechanism. In fact, PTMs addition or removal may alter proteins’ physical and chemical properties thus affecting protein folding, local conformations, localization, stability, interactions and consequently, biological functions. Importantly, the family of intrinsically disordered proteins (IDPs), that coordinate and orchestrate many protein interaction networks, is highly enriched in PTMs sites. HMGA non-histone chromatin factors (HMGA1a, HMGA1b, and HMGA2) belong to this protein class. They have a crucial role in neoplastic transformation and are among the most abundant, highly interconnected, and post-translationally modified proteins in the nucleus. For this reason, we propose to systematically screen PTMs of all three HMGA family members in numerous different cellular contexts (cells of different origin and tumorigenic phenotype) by means of a mass spectrometry-based approach in order to characterize the type, the level and the variability of these modifications. The information achieved with this screening will provide suggestions to further investigate about the role of PTMs and modified protein domains in modulating HMGA activity. Therefore, we propose to address these aspects to elucidate HMGA mechanisms of action in order to provide useful insights for the future development of HMGA interfering strategies.
2. Materials and Methods

2.1 Cell cultures
Twenty human cell lines of different origin and tumorigenic phenotype were chosen for our in vivo screening of HMGA proteins PTMs. They are listed in table 3.1 and categorized following ATCC® description (www.lgcstandards-atcc.org). All these cell types are adherent-growing.

PC-3, DU145, FRO, ARO, NIM 1, and TPC-1 were grown in RPMI 1640 (HyClone). HEK293, C-4I, ME-180, HeLa, Chang, Hep G2, Hep3B, PLC/PRF/5, HBL100, MDA-MB-157, MDA-MB-468, MDA-MB-231, and MCF7 were grown in DMEM (Euroclone).

Growth media contained 10% fetal bovine serum, 2 mM L-glutamine, penicillin (100 U/mL), and streptomycin (100 μg/mL). PZ-HPV-7 cells were grown in Keratinocyte serum free medium (Invitrogen).

Cells were grown at 37 °C in humidified 5% CO₂ incubator and expanded every 3-4 days in a subcultivation ratio depending on confluence level and ATCC recommendations. Cell monolayers were dissociated by trypsin-EDTA (0.05% trypsin, 0.02% EDTA in PBS. PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, and 1.8 mM KH₂PO₄, pH 7.4) at 37°C. Trypsin was neutralized by adding the culture medium; cells were collected by centrifugation (200 g, RT, 5 min), resuspended in fresh medium, and seeded.

Cells were harvested under subconfluence conditions after two washing steps with cold PBS. Cells were collected by centrifugation, frozen, and stored at -80°C.

2.2 Recombinant HMGA proteins production
Recombinant HMGA protein forms produced were the following:

- human HMGA1a full-length, aa 1-106 (henceforward named HMGA1a FL);
- human HMGA2 full-length, aa 1-108 (HMGA2 FL);
- human HMGA1a acidic tail truncated, aa 1-89 (HMGA1a CT);
- human HMGA2 acidic tail truncated, aa 1-93 (HMGA2 CT).

Recombinant HMGA proteins were expressed using the human HMGA cDNAs cloned in the vector pAR3038 (Noro et al., 2003). Proteins were expressed using the Ros2DE3 Escherichia coli strain which contains a plasmid encoding tRNA normally poorly represented in bacteria.
2.2.1 Competent cells preparation
A few drops of Ros2DE3 bacterial suspension (50% glycerol solution, stored at -20°C) were added to 1 mL of SOB medium (20 g/L bacto-tryptone, 5 g/L bacto-yeast extract, and 8.5 mM NaCl) and shaking incubated for 1 h at 37°C until OD$_{600}$ = 0.5. Then, suspension volume was brought to 250 mL with SOB and shaking incubated for 1 h at 37°C until OD$_{600}$ = 0.5. Bacteria were incubated on ice for 10 min, centrifuged to remove medium (3000 g, 4°C, 10 min), resuspended in 80 mL of CCMB buffer (80 mM CaCl$_2$, 20 mM MnCl$_2$, 9 mM MgCl$_2$, and 0.1 M glycerol) and left 20 min on ice. Suspension was then centrifuged (3000 g, 4°C, 10 min) and bacteria resuspended in 20 mL of CCMB buffer. 50 μL aliquots were prepared and stored at -80°C.

2.2.2. Bacterial transformation
Bacterial transformation was performed using heat-shock method. 1 μg of plasmid DNA was added to 50 μL of competent Ros2DE3 cells thawed slowly on ice. Bacteria were keep on ice for 30 min, incubated at 42°C for 90 sec, and put back on ice for 5 min. Then, 200 μL of SOC medium (20 g/L bacto-tryptone, 5 g/L bacto-yeast extract, 8.5 mM NaCl, 20 mM glucose, and 1 mM MgCl$_2$) was added and bacterial suspension shaking incubated for 1 h at 37°C. The resulting culture was spread on a warmed plate containing 25 mL LB-agar medium (10 g/L bacto-tryptone, 5g/L bacto-yeast, 171 mM NaCl, and 14 g/L bacto-agar) supplemented with 50 μg/mL ampicilline and 34 μg/mL chloramphenicol and incubated at 37°C O/N.

2.2.3 Protein expression
A colony from the plate was inoculated into a solution of 25 mL of LB-broth (10 g/L bacto-tryptone, 5g/L bacto-yeast extract, 171 mM NaCl, 50 μg/mL ampicilline, and 34 μg/mL chloramphenicol) and bacteria left grown for 16 h at 37°C. The culture was expanded to 500 mL with LB-broth. Bacteria were grown at 37°C to OD$_{600}$ = 0.6 - 0.8, induced for protein expression with 0.5 mM isopropyl β-D-thiogalactopyranoside for 2 h, and harvested by centrifugation (3000 g, 4°C, 10 min). Bacteria were washed twice with cold PBS, harvested by centrifugation (3000 g, 4°C, 10 min) and stored at -80°C.

2.3 HMG proteins extraction
Extraction with perchloric acid (PCA) provides an easy method for efficient enrichment of both histone H1 and HMG proteins (HMGA, HMGB and HMGN) from a variety of cells and tissues (Zougman and Wisniewski, 2006). Total HMG proteins and histone H1 (from human cells, par. 2.1
Chapter 2  
Materials and Methods

and Xenopus laevis embryos, par 2.16) and recombinant HMGA proteins (from bacteria, par. 2.2) were selectively extracted by means of a 5% PCA (w/v) solution in water (Noro et al., 2003). Bacterial cells were lysed with a volume of PCA equal to pellet volume containing protease inhibitors (Phenylmetylsulfonyl Fluoride saturated solution in isopropanol and Protease Inhibitor Cocktail SIGMA, P8340, both used in a 1/100 ratio with respect to PCA volume) and deacetylase inhibitors (sodium butyrate, few grains of powder). Protease inhibitors were not added for extraction of proteins to be used in proteolytic digestion experiments. Lysates were sonicated for 15 seconds at 20% power (Digital Sonifier 250, Branson) and then centrifuged (2500 g, 4°C, 10 min). The supernatant was recovered and the pellet was subjected to two more extractions. All extraction steps were carried out on ice in order to avoid protein degradation. The extracts were acidified with HCl to ~0.3M and proteins precipitated with 10 volumes of cold acetone at -20°C for 12-18 h. Proteins precipitates were collected by centrifugation (2500 g, 4°C, 10 min), acetone was removed and pellet was air dried. Proteins were resuspended in water, filtered (Durapore PVDF 0.45 µm Ø, Millipore), and checked by SDS-PAGE (par. 2.5).

2.4 GST-fused PRMTs production and extraction

GST-fused human protein arginine methyl-transferases PRMT1 and PRMT6 cloned in the pGEX-6P1 vector (Frankel et al., 2002) were produced in E. coli Ros2DE3 cells as previously described in par. 2.2.2 and 2.2.3. Bacterial cells were resuspended in a volume of cold PBS in a 1/20 ratio with respect to initial culture volume. Cells were subsequently broken by three 10 sec sonication pulses (30% power, Digital Sonifier 250, Branson) on ice. The lysate was centrifuged for 15 min at 10000 g at 4 °C. The supernatant was incubated for 60 min at 4 °C with Glutathione Sepharose® 4B beads (Amersham Biosciences) (300 µL resin suspension for 500 mL initial culture) previously equilibrated with three PBS washes. Resin was then collected by centrifugation (1100 g, 4 °C, 5 min) and washed three times with cold PBS to completely remove unbound proteins. The purified active enzymes were eluted incubating the beads for 15 min at 4 °C with 300 µL of elution solution (30 mM glutathione, 100 mM Tris/HCl, pH 8, 120 mM NaCl) for 15 min at 4 °C. Proteins purity, molecular weight, and concentration were checked by SDS-PAGE (par. 2.5 and 2.8.2).
2.5 Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis (SDS-PAGE)

Proteins were separated using a discontinuous SDS-PAGE gel system.

Stacking gel: polyacrylamide gel (T=5%, C=3.3%) in 0.13 M Tris/HCl, pH 6.8, 0.1% SDS, 1mg/mL ammonium persulfate (APS). Polymerization occurs by addition of TEMED (2μL for 1mL solution).

Running gel: polyacrylamide gel (T=15%, C=3.3%) in 0.38 M Tris/HCl, pH 8.8, 0.1% SDS, 1mg/mL APS. Polymerization occurs by addition of TEMED (4μL for 10mL solution).

Gel thickness: 0.75 mm – length of stacking gel: 0.5 cm and length of running gel: 8 cm – gel width: 8 cm.

Running buffers: 1) 25 mM Tricine, 0.1% SDS adjusted at pH 8.1 with Tris base for HMGA proteins analyses; 2) 25 mM Tris, 192 mM Glycine, 0.1% SDS for GST-fused proteins analyses.

Samples to be analyzed were made 1% SDS (w/v), 15% glycerol (v/v), 5% DTT (w/v) and 0.13 M Tris/HCl, pH 6.8 and boiled at 95°C. Electrophoretic run consisted of two steps: accumulation (50 V for 30 min) and separation (200 V for about 1 h).

Separated proteins were fixed and stained by using a methanol/water/acetic acid solution (in a 5/4/1 volume ratio) containing 0.05% Coomassie Brilliant Blue R 250 (w/v). Dye excess was eliminated by 10% acetic acid (v/v) washes.

2.6 Liquid Chromatography-Mass Spectrometry (LC-MS) screening

For LC-MS analyses PCA extracts were conditioned to a final concentration of 0.1% trifluoroacetic acid (TFA) (v/v).

Reverse Phase-High Pressure Liquid Chromatography (RP-HPLC) of PCA extracts was carried out with a PerkinElmer Life Sciences apparatus (series 200 LC pump and 785A UV-visible detector) using a Vydac Protein C4 column (2.1 x 150 mm, 5 μm). Proteins were eluted using a water - (A)/acetonitrile - (B) gradient reported below and 0.1% TFA (v/v) as a modifier (solvents E Chromasolv for HPLC, Fluka).

The chromatograms were obtained by absorbance detection at 220 nm. At the same time, an aliquot of the eluted solution (1/50) was directly injected into an interfaced mass spectrometer (PE SCIEX, API 1), which gave an equivalent chromatogram by counting total ions (TIC) that reach the mass detector (Wilm et. al, 1996). Search for ion composition was carried out on each TIC peak, and a succession of \( m/z \) values was obtained for every protein form contained in that peak. The identification of all successions of \( m/z \) allowed obtain reconstructed mass spectra (±1 Da) that gave the molecular composition of the analyzed peak.
Spectrometer calibration was obtained directly injecting a polypropylene glycol solution and a control protein solution.

MS parameters were the following: positive ion mode; MS scan range: 850-1350 m/z; step: 0.1 m/z; dwell time: 0.64 msec (one complete scan every 4 sec).

UV chromatogram acquisition and elaboration software: Turbochrom Navigator 4.0 (Perkin-Elmer).

MS acquisition software: RAD 2.5 (PE-SCIEX). MS data analysis software: MAC SPEC 3.22 and BIO MULTIVIEW 1.3.1 (PE-SCIEX).

RP-HPLC gradient optimized for HMGA family proteins separation:

<table>
<thead>
<tr>
<th>Step</th>
<th>Time (min)</th>
<th>Flow (mL/min)</th>
<th>% Solvent A</th>
<th>% Solvent B</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.0</td>
<td>0.2</td>
<td>92</td>
<td>8</td>
<td>isocratic</td>
</tr>
<tr>
<td>2</td>
<td>7.0</td>
<td>0.2</td>
<td>86</td>
<td>14</td>
<td>linear</td>
</tr>
<tr>
<td>3</td>
<td>35.0</td>
<td>0.2</td>
<td>76</td>
<td>24</td>
<td>linear</td>
</tr>
<tr>
<td>4</td>
<td>5.0</td>
<td>0.2</td>
<td>0</td>
<td>100</td>
<td>linear</td>
</tr>
<tr>
<td>5</td>
<td>10.0</td>
<td>0.2</td>
<td>0</td>
<td>100</td>
<td>isocratic</td>
</tr>
<tr>
<td>6</td>
<td>5.0</td>
<td>0.2</td>
<td>92</td>
<td>8</td>
<td>linear</td>
</tr>
<tr>
<td>7</td>
<td>25.0</td>
<td>0.2</td>
<td>92</td>
<td>8</td>
<td>isocratic</td>
</tr>
</tbody>
</table>

2.7 RP-HPLC purification of HMGA proteins

PCA extracts from both human cells and bacteria (par.2.3) were RP-HPLC separated in order to obtain pure HMGA proteins. Endogenous HMGA2 protein from HepG2 cells, recombinant HMGA1a and HMGA2 proteins, and recombinant CK2 phosphorylated HMGA1a and HMGA2 proteins (par 2.9.1) were purified using the chromatographic system and separation gradient described above. A RP Supelco Discovery BIO Wide Pore C5 column (2.1 × 250 mm, 5 μm) was used. Protein fractions were collected and solvent was removed by evaporation with a HETOVAC vacuum dryer (Hetco). Proteins were resuspended in water (LC-MS Chromasolv, Fluka) and quantified (par. 2.8). Proteins purity and molecular masses were checked by both SDS-PAGE and mass spectrometry (par.2.5 and 2.10).

2.8 Protein quantification

2.8.1 Waddell method

HMGA2 protein concentrations were obtained by measuring tryptophan absorbance at 280 nm using an absorption coefficient for tryptophan of 5500 M⁻¹ cm⁻¹.
Purified recombinant HMGA proteins concentration was calculated according to a modified Waddell method. The Waddell coefficient 144 was modified according to HMGA2 calculated concentration and fixed at 154.2. Then, HMGA proteins were quantified measuring the absorbance at 225 nm and 215 nm and using the following equation: protein concentration (μg/mL) = (A_{215} – A_{225}) × 154.2 (Noro et al., 2003).

2.8.2 Densitometry
All proteins used in this work were quantified by densitometry. Optical densities of protein bands resolved in SDS-PAGE and Coomassie stained were determined using Image Scanner and Image Master 2D Software (Amersham Pharmacia Biotech). Unknown protein concentrations were calculated on the basis of a reference protein calibration curve.

2.8.3 Liquid Chromatography-UV detection (LC-UV)
HMGA proteins to be used in comparative enzymatic assays in small equimolar quantities (par. 2.9.2 and 2.9.3) required a more precise quantification that was performed by LC-UV detection. Proteins were analyzed with a 1200 Series Capillary HPLC equipped with a Variable Wavelength Detector G1314B (Agilent Technologies). About 2 μg of protein were made 0.1% TFA, separated on a Zorbax 300SB-C18 column (0.5 × 150 mm, 3.5 μm, Agilent Technologies) using a water 0.05% TFA - (A)/acetonitrile 0.05% TFA - (B) gradient reported below (solvents LC-MS, Chromasolv, Fluka), and detected by absorbance measurement at 220 nm. The software ChemStation Plus (Agilent Technologies) was used for signal acquisition and elaboration. Peak height and area values were used for protein quantification using a calibration curve.

<table>
<thead>
<tr>
<th>Step</th>
<th>Time (min)</th>
<th>Flow (μL/min)</th>
<th>% Solvent A</th>
<th>% Solvent B</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.0</td>
<td>15</td>
<td>98</td>
<td>2</td>
<td>isocratic</td>
</tr>
<tr>
<td>2</td>
<td>26.6</td>
<td>15</td>
<td>78</td>
<td>22</td>
<td>linear</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>15</td>
<td>100</td>
<td>24</td>
<td>linear</td>
</tr>
<tr>
<td>4</td>
<td>2.4</td>
<td>15</td>
<td>100</td>
<td>0</td>
<td>isocratic</td>
</tr>
<tr>
<td>5</td>
<td>1.0</td>
<td>15</td>
<td>2</td>
<td>98</td>
<td>linear</td>
</tr>
<tr>
<td>6</td>
<td>5.0</td>
<td>15</td>
<td>2</td>
<td>98</td>
<td>linear</td>
</tr>
</tbody>
</table>
2.9 In vitro HMGA modification assays

2.9.1 In vitro phosphorylation of HMGA proteins by Casein kinase 2 (CK2)

CK2 phosphorylation was performed by incubating 5 μg of recombinant HMGA proteins with 100-250 Units of CK2 (New England BioLabs) in 50 μl reaction volume (20 mM Tris/HCl pH 7.5, 50 mM KCl, 10 mM MgCl₂, and 200 μM ATP) at 37 °C for 16 h.

In order to obtain HMGA2 with different phosphorylation levels a time course experiment (30 min, 45 min, 1 h, 2 h, 4 h, or 16 h) was performed using 2.5 Units of CK2 for 5 μg of HMGA2. Phosphorylation reactions were stopped by TFA addition (1% final). Phosphorylation patterns were detected by mass spectrometry (par 2.10). CK2 phosphorylated HMGA proteins to be used for phosphate group mapping and ion mobility separation-mass spectrometry analyses were purified by RP-HPLC (par. 2.7).

2.9.2 In vitro comparative phosphorylation of FL and CT HMGA proteins by Cyclin dependent kinase 1 (CDK1)

In vitro phosphorylation of HMGA proteins by CDK1 was performed by incubating equimolar quantities (650 pmoles) of FL and CT recombinant HMGA1a or HMGA2 proteins in the same batch with 20 Units of CDK1 (New England BioLabs) in 30 μL of reaction volume (50 mM Tris-HCl, 10 mM MgCl₂, 2 mM DTT, 1 mM EGTA, 0.01 % Brij 35, pH 7.5, and ATP 200 μM) at 30°C. Phosphorylation reactions were stopped by TFA addition (1% final) at 30 min, 1 h, 2.5 h, 5 h, and 16 h.

Phosphorylation patterns were detected by mass spectrometry (par 2.10).

2.9.3 In vitro comparative methylation of FL and CT HMGA1a protein by Protein arginine N-methyltransferases (PRMT1 and PRMT6)

In vitro methylation reactions of HMGA1a by GST-PRMT1 and GST-PRMT6 were performed both with [³H]-labeled and unlabeled S-adenosyl-L-methionine (SAM). Equimolar quantities (216 pmoles) of FL and CT HMGA1a were incubated in the same batch with GST-PRMT1, GST-PRMT6 or GST (5 μg) in 30 μl of reaction volume (PBS, pH 7.4) in the presence of 4 μl [³H]-SAM (0.55 mCi/mL – Perkin Elmer). Methylation reactions for LC-MS analyses were carried out incubating FL and CT HMGA1a (432 pmoles) in the same batch with GST-PRMT1, GST-PRMT6 or GST (2 μg) in 30 μL of reaction volume (PBS pH 7.4) in the presence of 1 μl SAM 15mM at 30°C for 16 h.

Radioactive methyl group incorporation was visualized by electrophoresis and fluorography. Briefly, reactions were SDS-PAGE analyzed, Blue Coomassie stained, 10% acetic acid destained.
Chapter 2

Materials and Methods

(par. 2.5) and then soaked in Amplify Fluorographic Reagent (NAMP100 – GE Healthcare) for 30 min. Gel was vacuum dried (583 gel dryer, BioRad), exposed to film (Hyperfilm™ ECL - GE Healthcare) at -80 °C and radioactivity was visualized by fluorography.

Methyl group incorporation in “cold” reactions was detected by mass spectrometry (par. 2.10)

2.10 LC-MS analyses of HMGA

HMGA proteins from RP-HPLC purifications and from in vitro modification assays were LC-MS analyzed to check protein molecular masses.

HMGA proteins were made 0.1% TFA and LC-MS analysed with a 1200 Series Capillary HPLC (Agilent) coupled to an HCT Ultra ion trap (Bruker Daltonics). Modified HMGA were separated on a Zorbax 300SB-C18 column (0.5 × 150 mm, 3.5 μm, Agilent Technologies) using the water 0.05% TFA - (A)/acetonitrile 0.05% TFA - (B) gradient reported in par 2.8.3.

Electrospray ionization (ESI) source settings of the ion trap were the following: Capillary: -4150V; Dry gas temperature: 300 °C; Dry gas: 8 L/min; Nebulizer: 8 psi. Ion trap acquisition parameters were set with the Esquire Control software version 6.1 (Bruker Daltonics) using the Smart Parameter Setting option: positive ion mode, MS scan range: 600-1100 m/z; Compound Stability: 100%; Trap drive level: 100%; Target Mass: 800 m/z; Optimize: wide. Reconstructed mass spectra were obtained using the Data Analysis Version 3.4 software (Bruker Daltonics).

2.11 HMGA2 protein phosphate group mapping

2.11.1 Enzymatic digestions

In vitro CK2 phosphorylated and endogenous HMGA2 proteins were RP-HPLC purified and vacuum dried as reported above (par 2.7). Proteins were digested with trypsin (Trypsin Gold, Mass Spectrometry grade - Promega) with a 1/25 (w/w) enzyme/substrate (E/S) ratio in 50 mM ammonium bicarbonate for 16 h at 37 °C. 5 μg of protein were digested in 25 μL reaction.

2.11.2 Phospho-peptides enrichment

Phospho-peptides enrichment was carried out using the PhoshoProfile™ I Phosphopeptide Enrichment Kit (SIGMA). 5 μg of tryptic digest were made 250 mM acetic acid and loaded into batch columns containing a Ga³⁺ derivatized resin (PHOS-SELECT™ Gallium Silica Spin), centrifuged (500 g, RT, 30 sec), and incubated at RT for 15 min. 50μL of bind/wash solution (250mM acetic acid, 30% acetonitrile) were added to the column and eliminated by centrifugation for three times. The column was washed with 50 μL of mQ water. At the end, 25 μL of elution
solution (10% phosphoric acid) were loaded and peptides recovered by centrifugation. This step was repeated twice to obtain a more quantitative recovery of phospho-peptides. Purified phospho-peptides were evaporated and resuspended in 50 μL of 0.1% TFA water.

2.11.3 LC-MS/MS analyses (LC-Tandem Mass Spectrometry analyses)
Either non-enriched and phospho-peptide enriched samples were analyzed by LC-MS/MS using a PerkinElmer Life Sciences apparatus (series 200 LC pump) interfaced to an HCT Ultra ion trap (Bruker Daltonics). HPLC separations were performed with a Waters Symmetry C18 column (1.0 × 150 mm, 3.5 μm) with a water - (A)/acetonitrile - (B) gradient reported below and 0.1% (v/v) formic acid (FA) as a modifier (solvents LC-MS, Chromasolv, Fluka).

<table>
<thead>
<tr>
<th>Step</th>
<th>Time (min)</th>
<th>Flow (mL/min)</th>
<th>% Solvent A</th>
<th>% Solvent B</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.0</td>
<td>0.4</td>
<td>98</td>
<td>2</td>
<td>isocratic</td>
</tr>
<tr>
<td>2</td>
<td>23.0</td>
<td>0.4</td>
<td>77</td>
<td>23</td>
<td>linear</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>0.4</td>
<td>0</td>
<td>100</td>
<td>linear</td>
</tr>
<tr>
<td>4</td>
<td>5.0</td>
<td>0.4</td>
<td>0</td>
<td>100</td>
<td>isocratic</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
<td>0.4</td>
<td>98</td>
<td>2</td>
<td>linear</td>
</tr>
<tr>
<td>6</td>
<td>15.0</td>
<td>0.4</td>
<td>98</td>
<td>2</td>
<td>isocratic</td>
</tr>
</tbody>
</table>

In order to obtain an analytical flow of about 40 μl/min a custom-made splitting system was used. MS/MS parameters were the following: positive ion mode; Full MS scan: 100-1500 m/z; number of precursors selected: 3; fragmentation amplitude: 1.00 V with the Smart Frag active (from 50 to 200% of the fragmentation amplitude selected); MS/MS scan: 100-2300 m/z; active exclusion after accumulation of 2 MS/MS spectra for a period of 2 min; ion charge control (ICC) active allowing the storage of a maximum of 200000 ions in a maximum accumulation time of 200 msec. LC-MS/MS were elaborated with Data Analysis Version 3.4 software (Bruker Daltonics). Molecular masses of tryptic peptides were searched using two strategies: (i) MS data were exported as mgf files and used as input for the MASCOT MS/MS Ion Search option (www.matrixscience.com) and (ii) entire MS analyses were manually inspected to search ion peptide signals.

MASCOT MS/MS Ion Search parameters were the following: Database: SwissProt; Enzyme: trypsin; Allow up to: 3 missed cleavages; Taxonomy: homo sapiens; Fixed modifications: Acetyl N-term (for endogenous HMGA2) – none (for recombinant HMGA2); Variable modifications: Phospho (ST); Peptide tol.: ±1.2 Da; # 13C: 0; MS/MS tol.: ±0.6 Da; Peptide Charge: 2+ and 3+; Monoisotopic; Data format: Mascot generic; Instrument: ESI-TRAP.
Manual inspection of MS/MS spectra was performed with the help of Mascot MS/MS ion search results and Protein Prospector (http://prospector.ucsf.edu) MS-Product option that provides the theoretical values of all fragments that can be generated by CID from a specific peptide. Tryptic peptides mixtures were LC-MS analyzed with the same method in order to obtain complete tryptic maps.

### 2.12 Electrophoretic Mobility Shift Assays (EMSAs)

#### 2.12.1 DNA probes

The DNA probes used were a double-stranded B-type oligonucleotide (E3, a sequence from the human insulin receptor promoter region – Foti et al., 2003) and a four-way junction DNA (4WJ - Bianchi, 1988; Hill et al., 1997).

The sequences of the probes are (for E3 only the upper strand sequence is shown):

- **E3**: 5’-AGAAAAACTCCATCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACA-3’
- **4WJ Leg1**: 5’-CCCTATAACCCCTGCATTGAATTCCAGTCTGATAA-3’
- **4WJ Leg2**: 5’-GTAGTCGTGATAGGTGCAGGGGTTATAGGGG-3’
- **4WJ Leg3**: 5’-AACAGTAGCTCTTTATTCGAGCTCGCGCCCTATCACGACTA-3’
- **4WJ Leg4**: 5’-TTTATCAGACTGGAATTCAAGCGAGCTCGAATAAGAGCTACTGT-3’

#### 2.12.2 [γ-32P] labeling of DNA probe

Single strand labelling was carried out as follows:

4 µl of [γ-32P]ATP (3000 Ci/mmol = 10 µCi/µL – Perkin Elmer) were added to the reaction mix containing 12 µl H2O mQ, 1 µl ssDNA (10 pmol/µl), 1 µl T4-Polynucleotide Kinase (T4-PNK 10 U/µL, GE Healthcare), and 2 µl T4-PNK buffer 10× (GE Healthcare) and incubated at 37°C for 1 h. At the end, T4-PNK was heat inactivated by boiling the reaction solution for 2 min. 2.2 µl of 3 M NaCl were added (to 0.3 M NaCl).

For E3 probe, 2 µl of complementary ssDNA fragment (10 pmol/µl) were added to the reaction mix. For 4WJ probe, the same amount of each strand was added.

The solution was boiled for 10 min and, thereafter, gradually kept at RT to allow a correct base pairing. 80 µl of TEN buffer (10 mM Tris/HCl pH 7.5, 1 mM EDTA pH 8, and 100 mM NaCl) were added and unincorporated [γ-32P]ATP was removed by gel permeation chromatography using Micro Bio-Spin™ P30 columns (BioRad). Purified DNA probe solution (0.1 pmol/µL) was recovered by centrifugation (1000 g, RT, 4 min).
2.12.3 Protein/DNA binding

CK2 phosphorylated and un-phosphorylated HMGA2 FL used for EMSA analyses were prepared starting from a common phosphorylation solution (par 2.9.1) which was split in two equivalent aliquots. In one part the active enzyme was added, whereas in the other one the same amount of a heat-inactivated (3 min at 95 °C) enzyme was added. At the end of the phosphorylation reaction an aliquot was used to check the modification profile of HMGA2 by MS analyses (par 2.10). The remaining solutions were directly used for EMSA experiments. EMSAs were carried out by incubating proteins (1-16 pmol as indicated) with 100 fmol of labeled DNA in 20 µl reactions containing 180 mM NaCl, 1 mM MgCl₂, 0.01% BSA, 8% glycerol, and 10 mM Tris/HCl pH 7.9 at RT for 10 min.

2.12.4 Electrophoretic analysis

After incubation, protein-bound DNA and free DNA were separated on a native polyacrylamide gel run in 0.5× TBE (TBE: 89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3) at 15 V/cm at 4 °C for 2-3 h (4 h for 4WJ). Before sample loading, a pre-run of about one h was performed until constant amperage was reached.

Running gel: polyacrylamide gel (T = 7%, C = 3.3%) in 45 mM Tris base, 44 mM boric acid, 1 mM EDTA, 0.75 mg/mL APS, pH 8.3. Polymerization occurs by addition of TEMED (120 µL for 100 mL solution). Gel thickness: 1.5 mm, length: 18.5 cm and width: 30 cm.

Running buffer: 0.5× TBE.

At the end, gel was incubated in 10% acetic acid for 20 min, washed with water and vacuum dried (583 gel dryer, BioRad). Visualization was achieved by autoradiography (Hyperfilm™ ECL - GE Healthcare) at -80 °C.

2.13 Electrospray Ionization–Mass Spectrometry (ESI-MS) analyses

HMGA proteins, Cytochrome C and Myoglobin (Myoglobin and Cytochrome C were purified proteins purchased from Sigma) were diluted to a final concentration of about 5 µM in 50 mM ammonium acetate / 5% acetonitrile for analyses under native conditions and in 0.1% formic acid / 10% acetonitrile for analyses under denaturing conditions.

Samples were directly infused into the ESI source of an HCT Ultra ion trap (Bruker Daltonics) at a flow rate of 3 µl/min. Interface settings were the following: Capillary: -3650V; Dry gas temperature: 300 °C; Dry gas: 5 L/min; Nebulizer: 10 psi. Ion trap acquisition parameters were set with the Esquire Control software version 6.1 (Bruker Daltonics) using the Smart Parameter Setting option: positive ion mode; Compound Stability: 100%; Trap drive level: 100%; MS scan range:
600-2200 m/z; Target Mass: incremental from 600 to 2200 m/z with increasing steps of 100 m/z after 13 scans; Optimize: wide. Reported mass spectra are the averaged mass spectra of the entire mass spectrometric analysis. MS data were elaborated using the Data Analysis Version 3.4 software (Bruker Daltonics).

2.14 Limited proteolysis analyses

FL and CT HMGA2 forms were subjected to limited proteolysis using trypsin (Trypsin Gold, Mass Spectrometry grade - Promega) in 100 mM ammonium acetate pH 7.0 at an E/S ratio of 1/1000 (w/w). Immediately after the addition of trypsin, HMGA digestion solutions were loaded inside nano-electrospray capillaries to directly follow the ongoing of the proteolytic cleavage by mass spectrometry.

Otherwise, limited proteolysis experiments using Lys-C (Roche) were performed in batch in 80 mM ammonium bicarbonate pH 8.1 at an E/S ratio of 1/5000 (w/w) at 25°C and stopped by acetic acid addition (5% final) after 5 and 15 min.

ESI-MS measurements of limited proteolysis with trypsin and Lys-C were performed on a Q-TOF Micro mass spectrometer (Micromass, UK) equipped with a Z-spray nanoflow electrospray ionization interface. Mass spectra of peptide digests were acquired using the nano electrospray source operating at capillary, cone, and extractor voltages of 1800, 35, and 1 V, respectively (positive ion mode).

Nano electrospray ionization capillaries were custom made from borosilicate glass tubes of 1mm outer diameter and 0.78 mm inner diameter (Harvard Apparatus, Holliston, MA) using a Flaming/Brown P-80 PC micropipette puller (Sutter Instruments, Hercules, CA) and gold coated using an Edwards S-150B sputter coater (Edwards High Vacuum, West Sussex, UK). Instrument control and data acquisition and processing were achieved using the MassLynx software (Micromass, UK).

Limited proteolysis experiments performed with Lys-C endoprotease were also analyzed by LC-MS with a 1200 Series Capillary HPLC (Agilent) coupled to an HCT Ultra ion trap (Bruker Daltonics). Proteins and peptides were separated on a Zorbax 300SB-C18 column (3.5 μm, 150 x 0.5 mm, Agilent) using a water 0.05% TFA - (A)/acetonitrile 0.05% TFA - (B) gradient. The ESI source settings of the HCT Ultra ion trap were the same as described in par. 2.10, with the only difference being the MS scan range: 300-1300 m/z and the Target Mass value: 700 m/z. In these analyses, both total ion count (TIC) and UV chromatograms (Abs 220 nm) were recorded.
2.15 Ion Mobility Separation–Mass Spectrometry (IMS-MS) analyses

FL, CT and CK2 phosphorylated HMGA1 and HMGA2 proteins were diluted in 100 mM ammonium acetate. Samples were infused into the ESI source of a SYNAPT™ HDMS™, a hybrid quadrupole-IMS-orthogonal acceleration time-of-flight mass spectrometer (oa-TOF) (Waters Corp. Manchester, UK) at a flow rate of 10 µl/min. An ESI capillary voltage of 3.2 kV was used, together with a sampling cone voltage of 30 V. The desolvation gas flow (Nitrogen) and temperature were held at 800 l/hr and 300 °C respectively. Throughout the study the source backing pressure was maintained at 5.2 mbar to produce intact gas phase ions from the protein in solution. The electrosprayed ions were transferred non-mass selectively through the quadrupole into the IMS enabling Triwave™ device.

The Triwave consists of three travelling wave (T-wavetm) ion guides. The first, the Trap T-Wave trapped and accumulated ions. Then these stored ions were gated into the IMS T-Wave in which the ion mobility separation occurred. The Transfer T-Wave was then used to transport the separated ions into the orthogonal acceleration time-of-flight mass spectrometer for MS analysis. In these experiments, the pressure in the Trap and Transfer T-Wave regions was about 0.02 mbar of argon and the pressure in the IMS T-Wave was 0.47 mbar of nitrogen. IMS travelling wave velocities of 300 m/s were used, and the wave pulse height was 6.5 V. Recording the temporal arrival profile of ions (mobilogram) was achieved by synchronization of the oa-ToF acquisition with the gated release of ions from the Trap to the IMS T-Wave. Following the gated pulse, the subsequent 200 orthogonal acceleration pushes (mass spectra) of the ToF analyser were recorded, giving an overall mobility recording time of 200 × tpp, where tpp, the pusher period, was equal to 90 µs. After the next gated release of ions, a further 200 mass spectra were acquired and added to the corresponding spectra from the previous acquisition.

Instrument control and data analysis were performed using MassLynx v. 4.1 software (Waters Corp. Manchester, UK). The deconvolution of ESI mass spectra of intact proteins was performed using the MaxEnt 1 algorithm in the MassLynx software. Mobility data were visualized and processed using the driftscope module within MassLynx. Mass accuracy was ensured by calibration on a separate introduction of sodium iodide (2 mg/mL in 1:1 v/v aqueous methanol).

2.16 Immunoprecipitation (IP) of XLHMGA2β-a from PCA extracts

XLHMGA2β-a was PCA extracted from Xenopus laevis embryos (stage 32-33 and 33-34) as previously described in par. 2.3. Since HMGA molecules present in PCA extracts were undetectable by LC-MS because they were expressed at a low level and co-eluting with other
proteins, a protein enrichment by immunoprecipitation (IP) was performed in order to concentrate XLHMGA2β-a.

50 μl of Protein G Sepharose 4 fast-flow (GE Healthcare) were washed three times with Tris HCl 50 mM pH 7. Resin was then collected by centrifugation (1100 g, 4 °C, 5 min) and incubated with 400 μl of a Tris HCl 50 mM pH 7 solution containing an α-HMGA2 antibody (developed against the human HMGA2 protein) for 1 h at 4°C. Resin was washed three times with Tris HCl 50 mM pH 7 and equilibrated once with binding buffer (25mM Hepes pH 7.9, 50 mM NaCl, 1 mM DTT, and 0.01% NP40).

PCA extracts from stage 32-33 and 33-34 embryos were pulled together, conditioned to the final concentration of binding buffer, and incubated with the resin for 3 h at 4°C.

Resin was then collected by centrifugation (1100 g, 4 °C, 5 min), washed four times with binding buffer to completely remove unbound proteins, and finally dried.

Immunoprecipitated proteins were eluted by addition of a 2% TFA water solution, vacuum dried, resuspended in a 0.1% TFA water solution and LC-MS analyzed.

### 2.17 Circular dichroism (CD) measurements

Far-UV CD spectra of FL and CT HMGA2 were recorded at room temperature in 20 mM sodium phosphate, pH 7.4, 30 mM NaCl.
3. Results and Discussion

3.1 Liquid Chromatography-Mass Spectrometry (LC-MS) screening

In order to evaluate which PTMs affect the three HMGA family members in different biological contexts and whether these modifications could be associated with tumorigenic phenotype, we performed a LC-MS screening on twenty cell lines (listed and categorized in table 3.1) in which HMGA expression was expected. These cell lines differ in their origin (kidney, cervix, liver, breast, prostate, and thyroid) and in their tumorigenic potential. With our approach we propose to systematically screen PTMs affecting the bulk of entire HMGA molecules, and therefore we refer to them as macroscopic PTMs.

Proteins were obtained from growing cells by 5% PCA extraction that allow a selective enrichment in linker histone H1 and HMG super-family proteins, i.e. HMGA, HMGB, and HMGN (Zougman and Wisniewski, 2005). Entire proteins were then analyzed by LC-MS, a technique that allowed us to combine reverse phase-high pressure liquid chromatography (RP-HPLC) separation of proteins, UV detection of amide bond at 220 nm, and simultaneous mass spectrometry detection of eluting fractions.

<table>
<thead>
<tr>
<th>Human Cell line</th>
<th>Origin</th>
<th>Characteristics</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEK293</td>
<td>Kidney</td>
<td>Embryo kidney, tumorigenic</td>
<td>+</td>
</tr>
<tr>
<td>C-4 I</td>
<td>Cervix</td>
<td>Carcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>ME-180</td>
<td>Cervix</td>
<td>Epidermoid carcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>HeLa</td>
<td>Liver</td>
<td>Adenocarcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>Chang</td>
<td>Liver</td>
<td>Normal hepatocyte</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>Hep G2</td>
<td>Liver</td>
<td>Hepatocellular carcinoma</td>
<td>+ +</td>
</tr>
<tr>
<td>Hep3B</td>
<td>Liver</td>
<td>Hepatocellular carcinoma, tumorigenic</td>
<td>+ +</td>
</tr>
<tr>
<td>PLC/PRF/5</td>
<td>Liver</td>
<td>Hepatoma, tumorigenic</td>
<td>+ +</td>
</tr>
<tr>
<td>HBL 100</td>
<td>Breast</td>
<td>Non-tumor SV40 immortalized</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>MCF7</td>
<td>Breast</td>
<td>Adenocarcinoma</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>MDA-MB-157</td>
<td>Breast</td>
<td>Medullary carcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>MDA-MB-468</td>
<td>Breast</td>
<td>Adenocarcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>MDA-MB-231</td>
<td>Breast</td>
<td>Adenocarcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>PZ-HPV-7</td>
<td>Prostate</td>
<td>Immortalized normal epithelium</td>
<td>+ +</td>
</tr>
<tr>
<td>PC-3</td>
<td>Prostate</td>
<td>Adenocarcinoma, tumorigenic</td>
<td>+ +</td>
</tr>
<tr>
<td>DU 145</td>
<td>Prostate</td>
<td>Carcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>NIM 1</td>
<td>Thyroid</td>
<td>Papillary carcinoma, low tumorigenic</td>
<td>+ +</td>
</tr>
<tr>
<td>TPC-1</td>
<td>Thyroid</td>
<td>Papillary carcinoma</td>
<td>+ +</td>
</tr>
<tr>
<td>FRO</td>
<td>Thyroid</td>
<td>Anaplastic carcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
<tr>
<td>ARO</td>
<td>Thyroid</td>
<td>Anaplastic carcinoma, tumorigenic</td>
<td>+ n.d.</td>
</tr>
</tbody>
</table>

Table 3.1 Cell lines used in LC-MS screening. HMGA1 and HMGA2 protein expression in each cell line is also indicated. n.d.: not detected in our LC-MS analyses. Cell lines are categorized following ATCC® description.
The experimental conditions adopted for chromatographic analyses were optimized for the separation of the three HMGA family members. In Figure 3.1 the chromatographic separation of PCA extracted proteins from Hep G2 cells, which we chose as representative of all our analyses, is reported. As it is possible to see from the chromatogram we were able to obtain several protein fractions. In the chromatographic region where HMGA are expected to elute (30-50 min) we detected five peaks that we named F1-F5. Mass spectra relative to chromatographic fractions have been deconvoluted to obtain reconstructed mass spectra that gave the molecular masses composition of the analyzed peak. Proteins were then identified by comparison between theoretical and experimental molecular mass values, taking account of possible PTMs addition.

![Figure 3.1 Chromatographic separation of HMG proteins.](image)

Figure 3.1 Chromatographic separation of HMG proteins. Elution profile of HMG and histone H1 proteins PCA extracted from Hep G2 cells and separated by RP-HPLC. F1: HMGN2, F2: HMGN3 and HMGN4, F3: HMGA1b and HMGN1, F4: HMGA2, F5: HMGA1a. Detection was performed measuring absorbance at 220nm.

As regard HMGA, they are known to lose their initial methionine during the translation process and to be constitutively acetylated at N-terminal serine. Therefore, HMGA proteins molecular masses are the following: HMGA1a - 11586.8 Da, HMGA1b - 10589.7 Da, and HMGA2 - 11742.8 Da.
Every additional PTM results in a specific increase in the molecular mass of each protein: for example, 80 Da for phosphorylation, 42 Da for acetylation, and 14 Da for methylation.

The evaluation of the addition and combination of such PTMs allowed identify the different HMG A modified forms present in our samples. The same approach was used to identify the other LC-MS analyzed proteins. F1-F5 turned out to effectively correspond to HMG super-family members. In these five fractions HMG A co-elute together with the highly related HMG N proteins, whereas the extremely abundant histone H1 and the HMGB proteins were efficiently separated from HMG A and elute at the end of the gradient.

Hereafter our investigation focused only on fractions F1-F5 and in table 3.2 the assignment of all proteins identified in each fraction is shown.

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Protein</th>
<th>Modifications</th>
<th>Molecular Mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Theor.</td>
</tr>
<tr>
<td>F1</td>
<td>HMGN2</td>
<td>None</td>
<td>9261.5</td>
</tr>
<tr>
<td>F2</td>
<td>HMGN3</td>
<td>None</td>
<td>10534.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>10614.6</td>
</tr>
<tr>
<td>F2</td>
<td>HMGN4</td>
<td>None</td>
<td>9407.7</td>
</tr>
<tr>
<td>F3</td>
<td>HMGN1</td>
<td>None</td>
<td>10527.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox</td>
<td>10543.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>10607.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox</td>
<td>10623.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P P</td>
<td>10687.7</td>
</tr>
<tr>
<td>F3</td>
<td>HMGA1b</td>
<td>N-term A P P</td>
<td>10749.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P P</td>
<td>10829.7</td>
</tr>
<tr>
<td>F4</td>
<td>HMGA2</td>
<td>N-term A P</td>
<td>11822.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P</td>
<td>11902.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P P</td>
<td>11982.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P P P</td>
<td>12062.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P P P</td>
<td>12142.8</td>
</tr>
<tr>
<td>F5</td>
<td>HMGA1a</td>
<td>N-term A P P</td>
<td>11746.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P M</td>
<td>11760.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P M</td>
<td>11774.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P P</td>
<td>11826.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P P</td>
<td>11840.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-term A P P P</td>
<td>11854.8</td>
</tr>
</tbody>
</table>

**Table 3.2 LC-MS identified HMG A and HMGN proteins and their PTMs.** Differently modified HMG A and HMGN protein forms detected in LC-MS analyses have been identified by comparison between theoretical (Theor., black) and experimental (Exp., blue) molecular mass values. PTMs are schematically represented as coloured boxes.

- Ox - Oxidation (+16 Da)
- N-term A - N-terminal acetylation (+42 Da)
- M - Methylation (+14 Da)
- P - Phosphorylation (+80 Da)
We were able to find and analyze the PTMs of all three HMGA proteins (HMGA1a – F5, HMGA1b – F3, and HMGA2 – F4), which are our major interest, together with the four HMGN proteins (HMGN1 – F3, HMGN2 – F1, HMGN3 – F2, and HMGN4 – F2). As shown in table 3.2 some proteins co-elute in the same chromatographic peak, this is the case of HMGN3 and HMGN4 in F2 and HMGN1 and HMGA1b in F3.

Our analyses revealed also that HMGA1 and HMGN proteins turned out to be expressed in all the cell lines we considered whereas HMGA2 protein is present only in a subset of them (Tab. 3.1).

### 3.2 Identification of HMGA and HMGN post-translational modifications

Since with our LC-MS analyses we looked at full-length proteins, we could assume that the presence of PTMs that introduce charges alterations, such as phosphate groups, does not substantially alter the ionization efficiency of HMG proteins. The intensity of the various reconstructed peaks can therefore be considered as indicative of their relative abundance, at least within the same protein (Diana et al., 2001).

In Figure 3.2 exemplificative reconstructed mass spectra of all HMG proteins that have been identified in F1-F5 fractions from Hep G2 cells are reported. As it is possible to see from this figure and from the scheme of modifications in table 3.2, the three HMGA and the four HMGN have substantially different macroscopic PTMs patterns and, interestingly, HMGA turned out to be more heavily modified than HMGN.

We could detect and compare four different types of macroscopic PTMs:

- **Acetylation.** All HMGA forms are mono-acetylated, as expected. In fact, as mentioned above, they undergo N-terminal acetylation during protein synthesis. Whereas HMGN do not carry this kind of modification;

- **Phosphorylation.** All HMGA are highly phosphorylated proteins, however showing relevant intrafamily differences. HMGA1a and HMGA1b (Figure 3.2, panels F and D, respectively) share the same phosphorylated forms being completely distributed among the di- and tri-phosphorylated forms (2P and 3P), whereas HMGA2 (panel E) shows a broader phosphorylation status from mono- to penta-phosphorylated forms (1P, 2P, 3P, 4P, and 5P). Moreover, the distribution of HMGA1a between the 2P and 3P form is equivalent to that of HMGA1b. These peculiar phosphorylation patterns were found for all HMGA proteins from the different cell lines analyzed in this work, which mass spectra are reported in Figure 3.3 for HMGA1a and HMGA1b and in Figure 3.4 for HMGA2. As regard HMGA2, in some cell lines (for example, Hep3B and TPC-1) also an esa-phosphorylated (6P) form can be detected.
Figure 3.2 Reconstructed mass spectra of HMGA and HMGN proteins from Hep G2 cells. PCA extracted HMGA and HMGN proteins were analyzed by LC-MS. For each peak of the chromatographic separation showed in Figure 3.1 (F1-F5), registered mass spectra were deconvoluted and reconstructed mass spectra obtained for each of the various HMG. The identity of each protein peak was obtained comparing experimental with theoretical molecular masses (Da). A. F1: HMGN2. B and C. F2: HMGN3 and HMGN4. D. F3: HMGN1 and HMGA1b. E. F4: HMGA2. F. F5: HMGA1a. 1P, 2P, 3P, 4P, and 5P: mono-, di-, tri-, tetra-, and penta-phosphorylation.
Figure 3.3 LC-MS screening: reconstructed mass spectra of HMGA1 proteins. PCA extracted HMGA1 proteins were analyzed by LC-MS. For chromatographic fractions F3 and F5 registered mass spectra were deconvoluted and reconstructed mass spectra obtained. The identity of each protein peak was obtained comparing experimental with theoretical molecular masses (Da). P: phosphorylation; M: methylation; Ox: oxidation. HMGA1b is shown together with the co-eluting HMGN1.
Figure 3.4 LC-MS screening: reconstructed mass spectra of HMGA2 protein forms. PCA extracted HMGA2 protein was analyzed by LC-MS. For chromatographic fraction F4 registered mass spectra were deconvoluted and reconstructed mass spectra obtained. The identity of each protein peak was obtained comparing experimental with theoretical molecular masses (Da). P: phosphorylation.
HMGA1 phosphorylations we detected are consistent with previously reported data that clearly demonstrate that HMGA1 proteins are constitutively phosphorylated by CK2 on their acidic C-terminal tail (Zhang and Wang, 2008). Differently, no experimental data regarding HMGA2 in vivo phosphorylation degree and location were previously available.

HMGN are prevalently unphosphorylated proteins; in fact as it is possible to see in Figure 3.2, HMGN2 (panel A) and HMGN4 (panel C) are completely unmodified, whereas only low-abundance of 1P forms of HMGN3 (panel B) and of 1P and 2P forms of HMGN1 (panel D) are detectable. The same kind of HMGN modifications was found in all the twenty cell lines analyzed. Here we show only HMGN1 protein (co-eluting with HMGA1b) which reconstructed mass spectra are reported in HMGA1b panels in Figure 3.3.

- **methyl**ation. HMGA1a is the only HMGA protein that is heavily methylated. In panel F of Figure 3.2 mass peaks of 11760.4 and 11775.9 correspond to mono- and di-methylated forms (1M and 2M) of di-phosphorylated HMGA1a, whereas mass peaks of 11840.3 and 11854.9 correspond to 1M and 2M forms of three-phosphorylated HMGA1a. Figure 3.3 and 3.4 show that in all the cell lines analyzed only HMGA1a is abundantly methylated. In fact, HMGA1b shows only a very slight methylation level in a limited number of cell lines (ME-180, HeLa, Hep G2, PC-3, NIM 1, FRO, and ARO), whereas HMGA2 is always completely un-methylated.

It is significant that in vivo mono-methylation (as well as symmetric and asymmetric di-methylation) of HMGA1a was mapped at R25 embedded in the first DNA-binding domain (Sgarra et al., 2003; Zhang and Wang, 2008). This domain is perfectly conserved both in the HMGA1b isoform and in HMGA2; however here we demonstrate that these proteins are substantially not methylated in vivo. Such a difference could be due to a different docking of the PRMT responsible for R25 methylation (presumably PRMT1, see par. 1.16.3). Indeed, at least for PRMT1, it has been previously demonstrated that positively charged residues distal to the site of methylation are important for the high affinity interaction between this enzyme and its substrates (Osborne et al., 2007). In the HMGA context, these positively charged residues could reasonably be constituted by the R and K residues of the second DNA-binding domain. It is noteworthy that R25 in HMGA1a is 29 amino acid residues from the second DNA-binding domain whereas the corresponding R residues in HMGA1b and HMGA2 are only 18 and 17 positions, respectively, from the second DNA-binding domain. In addition, the amino acid sequences embedded between the modified R residue and the cationic amino acid cluster significantly differ in the three proteins, especially considering their content of the proline residues, which confer specific structural constraints (see Figure
The different size and flexibility of the stretches between the methylation site and the basic clusters could explain the different methylation pattern. However, it remains to be clarified whether PRMT1 is the enzyme responsible for R25 methylation in vivo. Alternatively, the three HMGA proteins could have different chromatin localizations and consequently be differently accessible to PRMTs.

As regard HMGN proteins, they were always found not methylated;

\[
\text{HMGA1a:} \\
\text{KRGR}_{25}\text{GRPRKOPVSPGTLVSQKEPSEVPTPKRPRGPK} \\
\text{HMGA1b:} \\
\text{KRGR}_{25}\text{GRPRKOPKPEVPTPKRPRGPK} \\
\text{HMGA2:} \\
\text{KRGR}_{28}\text{GRPRKQEQEPTGESP}PKRPRGPK
\]

Figure 3.5 Comparison of HMGA1a, HMGA1b and HMGA2 sequences comprising the first two AT-hooks. HMGA sequences from the first to the second AT-hooks are shown. The arginine residue which is in vivo methylated in HMGA1a (R25) is evidenced in both HMGA1a and HMGA1b sequences. The corresponding arginine in HMGA2 sequence (R28) is also evidenced. The DNA-binding domains (AT-hooks) are underlined and the distance in terms of number of amino acid residues between R25 (28) and the first amino acid residue of the basic amino acid cluster of the second DNA-binding domain is indicated. Proline residues are evidenced in bold.

- oxidation. Modified forms which are consistent with a hydroxylation event at the level of a P, K, or D residue were found for HMGN1 protein. In fact, in the reconstructed spectrum reported in panel D of Figure 3.2 the principal peaks corresponding to the 0P (10527.0 Da) and 1P HMGN1 forms (10606.9 Da) are both accompanied by a peak differing +16 Da (10543.7 Da and 10623.0 Da). This modification can be considered peculiar of HMGN1 since none of the other proteins analyzed carries it.

Substantially, no other macroscopic post-translational modifications are detectable in HMGA and HMGN proteins. However, our experiments are not in discordance with literature data describing several in vivo HMGA PTMs not detected in our screening. In fact, previous studies rely on different mass spectrometry-based approaches employing tandem mass spectrometry for the
systematic identification and mapping of PTMs by focusing only on proteolytic peptides obtained from HMGA. This kind of analyses identified numerous phosphorylation, methylation and acetylation events on HMGA proteins, but do not provide information about the relative abundance of each modification in the total population of HMGA molecules. Whereas our approach allows a quantification of each modified form within the entire protein, even if does not provide PTMs site mapping informations.

3.3 Differences among the post-translational modifications of HMGA and HMGN proteins in different cell types

We have considered the possibility of differences in HMG post-translational modifications related to the cell line type, origin and/or phenotype.

No relevant differences were detected on HMGN proteins and therefore these will not be discussed further (data not shown).

Even if it was not possible to individuate a clear correlation between cell line origin/tumorigenic phenotype and HMGA PTM profiles, HMGA proteins showed a high heterogeneity in their PTMs related to the cell type considered.

Panels A, B and C of Figure 3.6 show selected comparisons of HMGA1a proteins from three different cell lines, which were chosen for their marked differences, whereas panel D shows a representative reconstructed mass spectra of HMGA2 protein forms.

Three considerations can be made:

1) The phosphorylation level of HMGA1 proteins is cell type-dependent. For example this is evident comparing the relative abundance of 2P and 3P peaks of HMGA1a in panel A (cervix cancer cells C-4I) and B (cervix cancer cells ME-180) of Figure 3.6. The percentage abundance of each modified form was calculated on the basis of the intensity of each reconstructed peak, considering the sum of the intensities of all the reconstructed peaks deriving from the same protein as 100%. The total abundance of bi-phosphorylated protein form is obtained by the sum of 2P, 2P/M, and 2P/2M forms and, in the same way, the total abundance of three- phosphorylated form is obtained by the sum of 3P, 3P/M, and 3P/2M forms in the spectrum. As it is possible to see, in C-4I cells 44% of HMGA1a molecules were found in the 2P form and the remaining in the 3P form, whereas in ME-180 the phosphorylation level is higher. In fact, 2P forms and 3P forms constitute the 33% and the 67% of HMGA1a molecules, respectively.
The cell type-dependent phosphorylation extent of HMGA proteins is even more evident taking into consideration all HMGA1a and HMGA1b reconstructed mass spectra of Figure 3.3.

Figure 3.6 Exemplificative comparison of HMGA phosphorylation and methylation patterns. A, B and C. Reconstructed mass spectra of HMGA1a from cervix cancer cell line C-4I (A) and ME-180 (B), prostate cancer cell line DU 145 (C). In the upper-left part of each panel the percentage of di-phosphorylation (2P+2P/M & 2M), tri-phosphorylation (3P+3P/M & 2M) and methylation (M_{tot}) is indicated. D, reconstructed mass spectra of HMGA2 from hepatocarcinoma cell line PLC/PRF/5. 1P, 2P, 3P, 4P, 5P and 6P: mono-, di-, tri-, tetra-, penta-, and esaphosphorylation. M and 2M: mono- and di-methylation. Green and grey ovals indicate 2P and 3P HMGA1a forms, respectively; red ovals indicate HMGA1a methylated forms; light-blue shape evidences the wide HMGA2 phosphorylation pattern.

2) The methylation level of HMGA1a is cell type-dependent. This is strikingly evident comparing the different relative abundance of M and 2M peaks in panel B (cervix cancer cells ME-180) and C (prostate cancer cells DU 145) of Figure 3.6. The percentage abundance of the total methylation of a protein is calculated on the basis of the sum of the percentage abundance of each methylated form, independently from the phosphorylation
degree. In ME-180 the methylation is calculated to affect 65% of total HMGA1a molecules, whereas in DU145 methylated forms are just 33% of the total protein. This is even more evident taking into consideration all the HMGA1a reconstructed spectra of Figure 3.3.

3) The HMGA2 protein is not always detectable, its expression being restricted to a relative small subset of cell lines (Table 3.1). This is in agreement with literature data confirming a different and restricted involvement of HMGA2 with respect to HMGA1 proteins in cancer (Fusco and Fedele, 2007). At variance with HMGA1, HMGA2 is always distributed within a wider phosphorylation range, from the mono-phosphorylated form up to the penta-phosphorylated one (Figure 3.6, panel D and Figure 3.4). However, also for HMGA2 the abundance of each specific phosphorylated form varies in a cell type-dependent manner. For instance, as it is possible to see in Figure 3.4, HMGA2 more abundant forms in NIM1 cells are 3P, 4P, and 5P forms, whereas in TPC-1 cells HMGA2 has a broader phosphorylation status being distributed from 1P to 6P forms. However, our LC-MS screening shows that the phosphorylation pattern of HMGA2 is always centered on the most abundant 3P and/or 4P forms.

In order to visualize the prevalent post-translational modification state of HMGA proteins we generated “combined” reconstructed mass spectra for each of the three HMGA proteins which are the mean spectra resulting from all the acquired mass spectra in each different cell type. In addition, we calculated the standard deviation of the abundance of each HMGA post-translationally modified form, which can be taken as an indicator of their fluctuation in the different cellular contexts considered. These combined spectra reassume all the previously described intra-family differences among the HMGA family members and are reported in Figure 3.7. For comparison, we show also the combined spectra of HMGN1, as an example of a protein with no relevant variations in its PTMs, as it is possible to see from the low standard deviation bars. Interestingly, even if HMGA and HMGN are highly related chromatin remodeling factors their modulation by PTMs addition seem to be very different. Therefore, we hypothesize that HMGA general chromatin activities, but not HMGN ones, can be controlled by highly fluctuating macroscopic PTMs.

3.4 HMGA2 post-translational modifications in an animal model

Since our laboratory was involved in a project aimed at studying the role of HMGA proteins during embryonic development using *Xenopus laevis* as a model, we had the availability of *Xenopus*
Chapter 3

Results and Discussion

Embryos (stage 32-33 and 33-34). From our previous data we know that *Xenopus* embryos express a HMGA2 ortholog, called XLHMGA2β-a, that is an essential factor for *Xenopus* development, whereas a HMGA1 ortholog has not been described (Benini et al., 2006; Hock et al., 2006; our unpublished data). Therefore, we decided to investigate *macroscopic* HMGA PTMs also in this biological model. Moreover, as shown by the protein sequences alignment in Figure 3.8 XLHMGA2β-a has a high degree of homology with human HMGA2.

In order to evaluate XLHMGA2β-a PTMs we LC-MS analyzed this protein after an immunoprecipitation enrichment performed on PCA extracts from embryos. XLHMGA2β-a (11470.5 Da - unmodified) was found always mono-acetylated and distributed in a wide phosphorylation range, from 1P to 7P, as it is possible to see from Figure 3.8.

Figure 3.7 Combined mass spectra of HMGA proteins and HMGN1. A. HMGA1a, B. HMGA1b, C. HMGA2, and D. HMGN1 mass spectra resulting from the LC-MS screening of all analyzed cell lines were combined together in order to obtain, for each single protein, the mean spectrum. Red bars indicate standard deviations. 1P, 2P, 3P, 4P, 5P, and 6P: mono-, di-, tri-, tetra-, penta-, and esa-phosphorylation. M and 2M: mono- and di-methylation. Ox: oxidation.
Therefore, *Xenopus* HMGA2 ortholog has a *macropscopic* phosphorylation degree *in vivo* that is almost identical to that of the human protein we detected in cell lines. This, in addition to primary sequence similarities, lets hypothesize conserved functions and modulatory mechanism for these two proteins.

\[
\begin{align*}
\text{hHMGA2} & : \text{SARGEAGQPSTAQGPAPA-\text{-}PQKRGRGPRKQQQEPT-GEPSKPRGRPKGSKNK} & -57 \\
\text{XLHMGA2\text{-}a} & : \text{SSR\text{-}EGARQ-SSVE\text{-}QPASPQSPKRPRKPKEPTAEGPSKPRGRPRPKGSKNK} & -56 \\
\end{align*}
\]

**Figure 3.8 HMGA2 post-translational modifications pattern in Xenopus laevis embryos.** Protein sequences alignment of human HMGA2 (hHMGA2) and Xenopus laevis HMGA2\text{-}a (XLHMGA2\text{-}a) are reported. Identities are evidenced in blue. The reconstructed mass spectra of LC-MS analyzed HMGA2\text{-}a from *Xenopus laevis* embryos (stage 32-33 and 33-34) is reported below. 1P, 2P, 3P, 4P, 5P, 6P, and 7P: mono-, di-, tri-, tetra-, penta-, esa-, epta-phosphorylation. 1A: mono-acetylation.

### 3.5 HMGA2 is multi-phosphorylated by CK2 on the acidic tail

Although the HMGA2 protein derives from another gene than HMGA1a and HMGA1b, it shares a significant homology with the other two family members. However, despite a growing number of evidences for its fundamental role in many important biological processes, such as development, tumorigenesis and senescence, it is still largely uncharacterised from the point of view of its post-translational modifications and their role in modulating HMGA2 function(s).

Our *in vivo* screening evidenced that HMGA2 is widely phosphorylated so we decided to further investigate this modification. Since its discovery it has been demonstrated that human HMGA2 is a phosphoprotein (Giancotti et al., 1991). Subsequently, it was shown that CK2 could be the enzyme responsible for the phosphorylation of its acidic tail (Goodwin, 1998; Schwanbeck et al., 2000). Moreover, also CDK1 and Nek2 kinases have been demonstrated to modify HMGA2 (Schwanbeck et al., 2000; Di Agostino et al., 2004). However, for all the above mentioned enzymes neither the
phosphorylation sites nor the stoichiometry of incorporated phosphates have been unambiguously assigned.

Since HMGA1 proteins belong to the same protein family and are constitutively phosphorylated by CK2 on their acidic C-terminal tail with the addition of up to three phosphate groups (S98, S101, and S102 – HMGA1a isoform – …EEEEG[98]QES[101]S[102]EEEQ) (par. 1.16.1 and Figure 1.12), we decided to focus on this enzyme. HMGA2 C-terminal tail (…EET[96]EE[99]S[100]S[101]QES[104]AEED) differs from that of HMGA1 because there are up to five potentially phosphorylatable residues by CK2 (T96, T99, S100, S101, and S104). Two of them, S100 and S104, are within a canonical consensus sequence for CK2 (S/T-xx-E/D) whereas the other three are embedded in an acidic context (phosphorylated S/T included), therefore becoming potential sites as well (Meggio and Pinna, 2003).

Therefore, we performed in vitro phosphorylation assays using commercially available recombinant CK2 in order to evaluate HMGA2 phosphorylation pattern and phosphates location.

Recombinant full-length HMGA2, C-terminal truncated HMGA2, and HMGA1a used as a positive control, were treated for 16 h with CK2 and reaction solutions were directly LC-MS analyzed in order to visualize the obtained protein forms. As can be noted from the reconstructed mass spectrum of Figure 3.9, panel A, CK2-phosphorylated full-length HMGA2 is present in three different forms. Unmodified HMGA2 has a theoretical molecular mass of 11700.8 Da, so the 11940.7, 12020.8, and 12100.9 Da values obtained are consistent with the addition of three, four, and five phosphate groups, respectively. Moreover, looking at the intensity of each reconstructed peak it is possible to see that the tetra-phosphorylated form is the most abundant one.

In order to ascertain whether all the phosphates are located only on the acidic tail we have also treated with CK2 a truncated form of HMGA2 lacking the entire C-terminal domain (HMGA2 1-93). Panel B of Figure 3.9 clearly shows the complete absence of phosphates in this truncated protein with theoretical molecular mass of 10019.5 Da. This confirms that all the CK2 phosphorylation sites reside on the acidic tail. HMGA1a was used as a positive control for enzyme activity since it is a well known CK2 substrate (Zhang and Wang, 2008). As we expected, CK2 modifies HMGA1a on all its three consensus sites. In fact, the reconstructed peak obtained in panel C has a mass of 11784.3 Da which is consistent with the addition of three phosphate groups to the theoretical molecular mass of unmodified recombinant HMGA1a of 11544.8 Da.

These data demonstrate that all the five S/T residues within the acidic tail of HMGA2 protein are phosphorylatable by CK2 even if some of them do not fulfill the canonical CK2 consensus site.
3.6 HMGA2 phosphate group mapping

3.6.1 Time Course phosphorylation assay

Since our in vivo screening demonstrated that HMGA2 distributes on five differentially phosphorylated forms, from 1P to 5P (Figure 3.4), we proposed to reproduce all these modified forms also in vitro using CK2. Therefore, we set up a time course phosphorylation assay with recombinant CK2 using both HMGA1a and HMGA2 as substrates and we stopped the reaction at 15 min, 30 min, 45 min, 1 h, 2 h, 4 h, and 16 h. We assayed both HMGA proteins in order to compare the rate of entrance of phosphates groups. The phosphorylation degree of HMGA proteins at the various time points was detected by LC-MS analyses. Figure 3.10 shows the reconstructed mass spectra obtained.
Figure 3.10 HMGA1a and HMGA2 CK2 phosphorylation time course analysis. HMGA1a and HMGA2 proteins were phosphorylated for the indicated times by CK2, LC-MS analysed, and corresponding mass spectra reported. Differentially phosphorylated forms positions are indicated by dotted lines numbered with the corresponding phosphorylation degree. 0P: unmodified form; 1P, 2P, 3P, 4P, and 5P: mono-, di-, tri-, tetra-, and penta-phosphorylated forms.
As it is possible to see from these spectra, the phosphorylation profiles of HMGA1 and HMGA2 proteins at each time point are almost identical until 1h of phosphorylation. For example, at 1h they are both distributed from 0P to 3P forms, with 2P form being the most abundant one and the other two forms with similar lower abundances. This means that the two HMGA family members, even if exhibiting different acidic tail sequences, are phosphorylated by CK2 in the same way at the earlier time points having the first three phosphorylation sites equivalent. Instead, the spectra detected from 2h to 16h reveal a different behaviour between the two proteins and this is consistent with the fact that HMGA2 has other two phosphorylatable sites on its acidic tail. In fact, as the reaction proceeds, the HMGA1a protein is found progressively shifted towards the 3P form, whereas the HMGA2 protein is shifted towards the more modified 4P (mainly) and 5P forms. Overall, the differentially modified forms obtained with the time course experiment range over any of the HMGA2 modified forms we found *in vivo* in the endogenous HMGA2. Therefore, we decided to map the phosphorylation sites both in the *in vitro* CK2 phosphorylated protein and in the endogenous one obtained from Hep G2 cells.

We purified the various phosphorylated HMGA2 proteins and subjected them to tryptic digestion and LC-MS analyses. We manually inspected the obtained mass spectra looking for HMGA2 predicted tryptic peptides molecular masses and we obtained a complete coverage of HMGA2 sequence for both the *in vitro* modified HMGA2 and the endogenous HMGA2. Moreover, among all the peptides found, the phosphorylated counterparts were searched by looking at possible 80 Da additions. As it is possible to see from the tryptic map reported in Figure 3.11, we were able to detect the 0P, 1P, 2P, 3P, 4P, and 5P forms of the C-terminal peptide (aa 90-108) from the CK2 *in vitro* modified protein and the 0P, 1P, 2P, and 3P forms from the endogenous one. In addition, only in the endogenous protein, we found a further phosphorylated peptide not belonging to the C-terminal domain, the peptide 33-45 1P. No other phosphorylated peptides were found even if we performed phosphopeptide enrichment by Immobilized Metal Affinity Chromatography (IMAC).

### 3.6.2 Mass sequencing analyses

In order to map HMGA2 phosphorylation sites both *in vitro* and *in vivo* we performed LC-MS/MS sequencing analyses on the same peptide samples analyzed before. Peptide identification and PTMs mapping were carried out by using both Mascot software *MS/MS ion search* option and Protein Prospector software *MS-Product* option that assisted the manual inspection of MS/MS spectra. Identification analyses confirmed the presence of the same phosphorylated peptides obtained before.
Figure 3.11 Tryptic map of peptides obtained from recombinant CK2 phosphorylated HMGA2 and endogenous HMGA2. LC-MS detected peptides of in vitro and in vivo HMGA2 phosphorylated forms are shown. rec.HMGA2: in vitro CK-2 treated recombinant HMGA2; endo.HMGA2: endogenous HMGA2 from HepG2 cells; black bars: unphosphorylated peptides common to both rec. and endo.HMGA2; red bars: phosphorylated peptides found in rec.HMGA2; blue bars: phosphorylated peptides found in endo.HMGA2; 1P, 2P, 3P, 4P, and 5P: mono-, di-, tri-, tetra-, and penta-phosphorylated peptides.

During MS/MS analyses, peptides ions called precursors (generally 2+ and/or 3+ charged) are fragmented by collisions with a neutral gas (collision induced dissociation, CID) and N- and C-terminal fragments generated (called b and y ions, respectively). Amino acid residues identity can be determined from the mass difference of successive fragment ions of the same type (e.g., bₙ and bₙ₋₁) (Paizs and Suhai, 2005).

From the mass sequencing of m/z signals corresponding to the peptide of molecular mass 2123.8 Da, we obtained clear fragmentation spectra (data not shown) from which it is possible to unambiguously assign these m/z values to the un-phosphorylated C-terminal tail of HMGA2 (K₉₀PAQEETETSSQESAEED₁₀₈), both in the recombinant and the endogenous protein. Peptides differing by 80 Da or its multiples from the molecular mass of K₉₀-D₁₀₈ likely represent differently phosphorylated forms. In addition, the loss of phosphoric acid (H₃PO₄) is a typical phenomenon observable during CID fragmentation of phosphopeptides and indeed, in our analyses we systematically observed such event during MS sequencing of di- and tri-charged acidic C-terminal tail phosphopeptides (with the appearance of peaks differing by -49 and -32.6 from the precursor ion, respectively), thus confirming that the fragmented ions are phosphopeptides.

Here we report the fragmentation spectra analysis of C-terminal phosphorylated peptide forms detected in recombinant (Rec) and endogenous (Endo) HMGA2 and of the 33-45 peptide monophosphorylated form detected only in the endogenous protein.
Mono-phosphorylated (1P) C-terminal peptide

The peptide with molecular mass 2203.8 Da matches in weight with the acidic C-terminal tail peptide (aa 90-108: 2123.8 Da) plus one phosphate group.

MS/MS spectra obtained from the bi-charged 90-108 peptide 1P (precursor ion: 1102.92+) deriving from the recombinant *in vitro* phosphorylated protein indicate that the first residue to be phosphorylated is S104. The Rec 1P panel of Figure 3.12 shows the fragmentation spectra of this peptide and Rec 1P panel of Figure 3.13 reports a schematic representation of the fragmentation pattern of this peptide. As it is possible to see, we detected un-phosphorylated b5-b14 fragments, while the b15-b18 fragments that comprise S104 (S15 in this peptide) were consistent with b ions carrying a single phosphate group. This evidence is supported by the presence of phosphorylated y5, y6, and y7 ions.

Differently, the MS/MS spectrum of the endogenous C-terminal peptide 1P shows fragments consistent with phosphorylation both on S104 or S100. Indeed, the fragmentation of tri-charged ion (736.03+) allowed us to detect both phosphorylated and un-phosphorylated b14, y5 and y6 ions (Figure 3.12 and 3.13, Endo 1P). These fragments suggest that *in vivo* S104 is not the only site of modification in a 1P peptide. The b series until b10 was found only matching with un-phosphorylated signals whereas the b11 fragment was found both un-phosphorylated and phosphorylated. Taken together these evidences suggest that both S100 and S104 in the endogenous protein can be the first amino acidic residues modified by CK2 in a mutually exclusive manner.

Di-phosphorylated (2P) C-terminal peptide

The peptide with molecular mass 2283.8 Da matches in weight with the bi-phosphorylated acidic C-terminal tail peptide of HMGA2.

MS/MS spectra obtained from the bi-charged Rec 2P acidic C-terminal peptide ion (1143.102+) clearly indicate that both S100 and S104 are modified. Rec 2P panels of Figures 3.14 and 3.15 show that ions of the b series until b10 are coherent with unmodified fragments, therefore excluding T96 and T99 as possible modification sites. b11-b14 are mono-phosphorylated implying S100 as one of the phosphorylated residues and b16-b18 are di-phosphorylated implying that S104 is the other modified residue. This is also confirmed by the y series where y4 is un-phosphorylated, y5, y7, and y8 are always mono-phosphorylated, and y10-y18 are di-phosphorylated. These data, together with those of the Rec 1P form allow us to establish that S100 is the second residue to be modified.
Figure 3.12 MS/MS fragmentation spectra of mono-phosphorylated HMGA2 C-terminal tail peptide. MS/MS fragmentation spectra of bi-charged mono-phosphorylated (1P) peptide ion (1102.9<sup>2+</sup>) from recombinant in vitro CK2 phosphorylated HMGA2 (Rec) and of tri-charged mono-phosphorylated peptide ion (736.0<sup>3+</sup>) from endogenous HMGA2 (Endo) are shown. b- (red) and y- (blue) series ions that have been found in the spectra are indicated by dotted lines together with their m/z value. -H<sub>3</sub>PO<sub>4</sub>: phosphoric acid loss; -H<sub>2</sub>O: water molecule loss. An asterisk indicates the presence of a phosphate group. The m/z value and the charge of the precursor ion are also indicated.
Figure 3.13 MS/MS fragmentation schemes of mono-phosphorylated HMGA2 C-terminal tail peptide. MS/MS fragmentation schemes of bi-charged mono-phosphorylated (1P) peptide ion ($1102.9^{2+}$) from recombinant in vitro CK2 phosphorylated HMGA2 (Rec) and of tri-charged mono-phosphorylated peptide ion ($736.0^{3+}$) from endogenous HMGA2 (Endo) are shown. b and y ions are indicated above and under the amino acid sequence, respectively. -H$_3$PO$_4$: phosphoric acid loss; -H$_2$O: water molecule loss; dotted lines: bi-charged fragments. An asterisk indicates the presence of a phosphate group. The m/z value and the charge of the precursor ion are also indicated.

Endo 2P MS/MS data on 762.8$^{3+}$ ion give the same result obtained for the Rec 2P peptide indicating that S100 and S104 are modified. As shown in Endo 2P panels (Figure 3.14 and 3.15) b3-b10 signals are un-phosphorylated, b11 and b14 are consistent with phosphorylated fragments and b16-b18 with di-phosphorylated ones. The y series is consistent with the b one, in fact, y5 and y8 are mono-phosphorylated whereas y9-y12 are di-phosphorylated.

**Tri-phosphorylated (3P) C-terminal peptide**

The tri-phosphorylated C-terminal peptide of both recombinant in vitro CK2 phosphorylated and endogenous protein is modified on S100, S101, and S104. Indeed, as it is evident from the Rec 3P and Endo 3P fragmentation spectra of 1183.3$^{3+}$ ion reported in Figures 3.16 and 3.17, b4-b10 fragments are un-phosphorylated (excluding therefore phosphates on T96 and T99), b11 is mono-phosphorylated (S100 phosphorylated), b12-14 are di-phosphorylated (S100 and S101 phosphorylated), and b15-b18 are tri-phosphorylated (S100, S101, S104 phosphorylated). The y series indicates the same. This means that the third residue being modified is S101.

The order we were able to establish (S104 and S100 followed by S101) is consistent with the preferential phosphorylation of CK2 on its canonical consensus sites, in fact S104 and S100 are within an S-xx-E motif, whereas S101 becomes a canonical CK2 site after S104 phosphorylation ($S_{101}$-xx-phospho$S_{104}$).
Figure 3.14 MS/MS fragmentation spectra of di-phosphorylated HMGA2 C-terminal tail peptide. MS/MS fragmentation spectra of bi-charged bi-phosphorylated (2P) peptide ion (1143.1$^{3+}$) from recombinant in vitro CK2 phosphorylated HMGA2 (Rec) and of tri-charged bi-phosphorylated peptide ion (762.8$^{3+}$) from endogenous HMGA2 (Endo) are shown. b- (red) and y- (blue) series ions that have been found in the spectra are indicated by dotted lines together with their m/z value. -H$_3$PO$_4$: phosphoric acid loss; -H$_2$O: water molecule loss. An asterisk indicates the presence of a phosphate group. The m/z value and the charge of the precursor ion are also indicated.
**Figure 3.15** MS/MS fragmentation schemes of bi-phosphorylated HMGA2 C-terminal tail peptide. MS/MS fragmentation schemes of bi-charged bi-phosphorylated (2P) peptide ion (1143.12$^+$) from recombinant *in vitro* CK2 phosphorylated HMGA2 (Rec) and of tri-charged bi-phosphorylated peptide ion (762.83$^+$) from endogenous HMGA2 (Endo) are shown. b and y ions are indicated above and under the amino acid sequence, respectively. -H$_3$PO$_4$: phosphoric acid loss; -H$_2$O: water molecule loss; dotted lines: bi-charged fragments. An asterisk indicates the presence of a phosphate group. The m/z value and the charge of the precursor ion are also indicated.

**Tetra- and penta-phosphorylated (4P and 5P) C-terminal peptides**

From CK2-treated HMGA2 we obtained also the tetra- and penta-phosphorylated C-terminal peptides (2442.7 and 2523.9 Da). The fragmentation pattern of the tetra-phosphorylated peptide clearly indicates that T96 isn’t phosphorylated (b8 and b9 fragments), whereas all the other phosphorylation sites are modified (data not shown). In the penta-phosphorylated form we found fragments coherent with phosphorylation at all S/T residues indicating that all of them can be modified by CK2 (data not shown) even if they do not perfectly fulfil the canonical CK2 consensus site.

We were not able to find tetra- and penta-phosphorylated forms of the acidic tail from the endogenous HMGA2 protein. This could be due either to the low abundance of tetra- and penta-phosphorylated forms or to the presence of phosphate groups outside the C-terminal region.

Data obtained both from the recombinant and the endogenous protein suggest the following phosphorylation order: S104 and S100 are the first residues to be phosphorylated followed by S101, T99, and finally T96 (Figure 3.18).
Figure 3.16 MS/MS fragmentation spectra of tri-phosphorylated HMGA2 C-terminal tail peptide. MS/MS fragmentation spectra of bi-charged tri-phosphorylated (3P) peptide ion (1183.3$^{2+}$) from recombinant in vitro CK2 phosphorylated HMGA2 (Rec) and of bi-charged tri-phosphorylated peptide ion (1183.0$^{2+}$) from endogenous HMGA2 (Endo) are shown. b- (red) and y- (blue) series ions that have been found in the spectra are indicated by dotted lines together with their m/z value. -$\text{H}_2\text{PO}_4$: phosphoric acid loss; -$\text{H}_2\text{O}$: water molecule loss. An asterisk indicates the presence of a phosphate group. The m/z value and the charge of the precursor ion are also indicated.
**Figure 3.17** MS/MS fragmentation schemes of tri-phosphorylated HMGA2 C-terminal tail peptide. MS/MS fragmentation schemes of bi-charged tri-phosphorylated (3P) peptide ion (1183.32+) from recombinant *in vitro* CK2 phosphorylated HMGA2 (Rec) and of tri-charged tri-phosphorylated peptide ion (1183.02+) from endogenous HMGA2 (Endo) are shown. b and y ions are indicated above and under the amino acid sequence, respectively. -HPO₄: phosphoric acid loss; -H₂O: water molecule loss; dotted lines: bi-charged fragments. An asterisk indicates the presence of a phosphate group. The m/z value and the charge of the precursor ion are also indicated.

**Figure 3.18** *In vitro* and *in vivo* HMGA2 acidic C-terminal tail phosphorylations. The scheme indicates the phosphate group mapping results of HMGA2C-terminal tail differentially phosphorylated forms (from 1P to 5P) obtained from both recombinant CK2 phosphorylated HMGA2 (*in vitro*) and endogenous HMGA2 from Hep G2 cells (*in vivo*). HMGA2 acidic C-terminal tail peptide...KPAQET⁹⁶EET⁹⁹S¹⁰⁰S¹⁰¹QES¹⁰⁴A... Phosphate group(s) position is indicated for each of the HMGA2 acidic C-terminal tail phosphorylated forms. P: phosphate group.
**Phosphorylation of endogenous HMGA2 outside the C-terminal tail**

Our LC-MS/MS analyses demonstrated that in the endogenous protein from HepG2 cells also a peptide not belonging to the C-terminal tail can be phosphorylated.

The fragmentation spectrum (Figure 3.19, panel A) of the precursor ion 767.8$^{2+}$ is dominated by a -49 neutral loss indicated in figure with peak number 2: (MH$_3$PO$_4$)$^{2+}$, m/z 718.8 that clearly demonstrates that the precursor ion belong to a phosphopeptide. Mass sequencing allows attribute this tryptic peptide to HMGA2 peptide 33-45 (KQQEPT$^{39}$GEPS$^{43}$PK) that contains two phosphorylable residues: T39 and S43. However, mass spectra inspection unambiguously assigns phosphorylation on S43. Indeed, the fragments of the b series until b9 (Figure 3.19, panel B) are consistent with un-phosphorylated fragments and therefore exclude T39 as the phosphorylation site, whereas b11 and b12, which comprise S43, carry one phosphate group, indicating therefore that this serine is the phosphorylation site. Also the y series confirms the location of the phosphate since y4, y5, and y6 fragments, which only contain S43, carry one phosphate group.

It was previously suggested that in the murine HMGA2 the same residue could be phosphorylated by the cyclin/CDK enzymes at the beginning of S-phase and in the G2/M phases of the cell cycle (Goodwin et al., 1998). It is worthwhile to evidence that S43 is the corresponding residue of T52 in human HMGA1a, which is well known to be phosphorylated by CDK1 (see Figure 1.12 and paragraph 1.16.1). The sequence in which HMGA2 S43 is embedded fits perfectly with the CDK1 consensus S/T-P-x-K/R (Marin et al., 1992) and it was previously demonstrated *in vitro* that CDK1 phosphorylates murine HMGA2 at the level of S43 and S58 resulting in a strong decrease of its DNA binding affinity (Schwanbeck et al., 2000).
Figure 3.19 Serine 43 of HMGA2 protein is in vivo phosphorylated. A. The MS/MS spectrum of precursor ion 767.8^{2+} belonging to the mono-phosphorylated HMGA2 33-45 peptide is reported. y and b ions together with their m/z values are indicated. 1,2,3,4 correspond to precursor ion with losses of water or phosphoric acid. B. a scheme of detected fragments is shown. b and y ions are indicated above and under the amino acid sequence, respectively. \(-\text{H}_3\text{PO}_4\): phosphoric acid loss; \(-\text{H}_2\text{O}\): water molecule loss; dotted lines: bi-charged fragments. An asterisk indicates the presence of a phosphate group.
3.7 Phosphorylation of the HMGA2 acidic C-terminal tail affects its DNA binding properties

Since a peculiarity of HMGA2 is the high degree of phosphorylation of the acidic C-terminal tail, we asked whether this modification could play a role in modulating its DNA binding properties. Therefore, we performed EMSA experiments with typical DNA targets for HMGA proteins. On one hand we considered the ability of HMGA proteins to recognize non canonical DNA structures and we used a Four Way Junction probe (Bianchi, 1988). On the other hand we took into consideration the ability of HMGA to bind to the minor groove of AT-rich stretches of B-type DNA and we used as target DNA a region of the Insulin receptor promoter (E3 probe) (Brunetti et al., 2001).

We set up EMSA experiments in order to have exactly the same protein concentrations and buffer compositions for both un-phosphorylated and CK2 phosphorylated HMGA2 (see paragraph 2.12.3). As it is possible to observe from the mass spectra reported in Figure 3.20, panel A, after CK2 phosphorylation, we obtained a penta-phosphorylated (5P) HMGA2 (and a tetra-phosphorylated minor form, 4P), whereas using a heat-inactivated enzyme we maintained the protein unmodified (0P).

The EMSA experiment performed with Four Way Junction (4WJ) probe clearly shows a change in the DNA binding properties of HMGA2 due to the C-terminal phosphorylation. Indeed, it is possible to observe a decrease in the DNA binding affinity of the phosphorylated protein compared with the un-phosphorylated one (upper part of Figure 3.20, panel B, compare lanes 2-6 with 8-12 of the 4WJ experiment). Moreover, the complexes carrying the phosphorylated protein show less stability during electrophoretic migration as evident from the rather diffused bands of lanes 2 and 3 compared with the well resolved bands of lane 8 and 9. In addition, comparing phosphorylated (lanes 2-5) with un-phosphorylated (lanes 8-11) HMGA protein/DNA bands, it is also possible to observe a change in the mobility of complexes which could reflect alterations in the organization of protein-DNA complexes, a phenomenon which was already described for the HMGA1b isoform (Piekielko et al., 2001).

The EMSA experiment carried out using B-type DNA probe substantially confirm the change in the DNA binding properties of HMGA2 due to C-terminal phosphorylation. A decrease in the DNA binding affinity of the phosphorylated protein compared with the un-phosphorylated one is evident at the level of both HMGA/DNA complexes (lower part of Figure 3.20, panel B, compare lanes 2-6 with 8-12 of the E3 experiment).
Figure 3.20 Acidic tail phosphorylation affects HMGA2 DNA-binding properties. A. HMGA2 was treated with CK2 or heat-inactivated CK2 in order to obtain either a fully-phosphorylated or a non-modified protein. Reconstructed mass spectra of the obtained proteins are shown. B. EMSA experiments were performed using decreasing amounts (16, 8, 4, 2 and 1 picomoles) of CK2- and CK2 inactivated-treated HMGA2 with $^{32}$P labelled Four Way Junction (4WJ) and E3 DNA probes. HMGA/DNA complexes detected are indicated on the left.
3.8 Could HMGA highly phosphorylated acidic tail have a conformational role?

Our LC-MS screening on macroscopic PTMs affecting HMGA family members highlights that C-terminal acidic tail is the domain of HMGA proteins that is more heavily modified. This evidence suggests a role for this protein portion in modulating HMGA activities. On the other hand, previous studies have underlined the importance of HMGA acidic tail in regulating HMGA proliferative and transforming ability. Importantly, truncated HMGA forms, which can be found in several human tumors after chromosomal rearrangements, are known to have enhanced oncogenic properties with respect to full-length ones (see par. 1.18). However, despite these evidences the molecular mechanisms of action underlying HMGA C-terminal tail biological functions have never been dissected. The loss of part or entire protein domain causes for many proteins structural changes responsible for functional alterations (Seidel et al., 2004). We asked ourselves whether HMGA acidic tail could have a conformational role and, therefore, whether the oncogenic differences between full-length and C-terminal truncated HMGA could be at least partially explained by variation in structural conformations.

Since no specific structural information is available for intrinsically disordered (ID) HMGA1 and HMGA2, we started our investigation by comparing the already known structural properties of the different proteins belonging to the HMG super-family. We performed a comparison of HMG’s structural domain organization and a bioinformatic prediction of disordered sequences in HMG proteins with the Predictor Of Naturally Disordered Regions (PONDR) software (Li et al., 1999). As it is possible to see in Figure 3.21, panels A and B, in contrast to HMGB that display order at the level of their DBDs (HMG Boxes), HMGA and HMGN protein sequences are predicted to be totally disordered having PONDR scores greater than 0.5 for their entire length. However, a peculiar characteristic common only to HMGA and HMGB is the presence of a highly negative C-terminal tail (Figure 3.21 A, C) that in both cases is predicted to be ID (Figure 3.21 B).

Several studies have been shown that HMGB1 acidic tail can intra-molecularly shield the highly positively charged HMG Boxes modulating HMGB1-DNA interactions and impairing accessibility to modifying enzymes, such for example CBP (Watson et al., 2007; Wang et al., 2007; Knapp et al., 2004; Pasheva et al., 2004). Similarly, HMGA proteins have highly positively charged DBDs and an acidic C-terminal tail whose negative charge density is comparable to that of HMGB or even higher, being enhanced by constitutive phosphorylation. As shown in Figure 3.21, panel C, the charge density (z density = n° negative charge/n° aa residues) of the long C-terminal tail of HMGB1 is 1, being entirely constituted by E and D residues each carrying one negative charge in physiological conditions. This value is also reached by the shorter and already highly negative HMGA1 and HMGA2 C-terminal tail sequences when fully phosphorylated.
Chapter 3 Results and Discussion

Figure 3.21 Prediction of intrinsically disordered (ID) status of HMG proteins and comparison between HMGA and HMGB1 acidic tails. A. Schematic representation of HMG proteins (HMGA1a, HMGA2, HMGB1, and HMGN1) and their functional domains. B. Prediction of ID regions in HMGA1a, HMGA2, HMGB1, and HMGN1 by PONDR (Predictor Of Natural Disordered Regions – Molecular Kinetics, Inc.) (Li et al., 1999). Scores greater than 0.5 indicate propensity to disorder. C. The amino acid sequences of C-terminal tails of HMGB1, HMGA1a, and HMGA2 are shown. Phosphorylatable residues are showed in red. For each tail the sum of negatively charged amino acid is indicated (z) together with charge density (z density: n° negative charge/n° aa residues). Phosphate group charges have been estimated using pK_a1 = 1.2 and pK_a2 = 6.5 (Halligan et al., 2004) and physiological pH 7.4.
In fact, HMGA1 C-terminal tail z density shifts from 0.53 to 0.93 when all the three S are phosphorylated, whereas HMGA2 C-terminal tail z density shifts from 0.53 to 1.2 when all five S/T residues are phosphorylated. Phosphate group charges have been estimated using pK$_{a1}$ = 1.2 and pK$_{a2}$ = 6.5 (Halligan et al., 2004) and physiological pH 7.4. The fact that our PTMs screening highlighted that HMGA molecules are distributed on differentially phosphorylated forms implies that in vivo the z density of acidic tail can be finely modulated being comprised between 0.5 and 1. These evidences suggest the possibility of a modulable intra-molecular interaction involving HMGA acidic domain and highly basic inner portions. This mechanism could resemble that demonstrated for HMGB, despite the structural differences between HMGA and HMGB proteins. Therefore, in order to investigate the possibility of such a conformational role for HMGA proteins acidic tail we propose to compare the structural properties of full-length (FL) and C-terminal truncated (CT) HMGA proteins, focusing on HMGA2 as representative of all three HMGA proteins.

We performed a classical spectroscopic characterization of HMGA2 conformation. As expected, the far-UV circular dichroism (CD) spectra of FL (aa 1-108) and CT (aa 1-93) HMGA2 forms did not show significant differences, both being characterized by an intense negative minimum centered at 198 nm, which is typical of a random coil conformation (Figure 3.22). These data confirm previous evidences underlining that HMGA lack any secondary structure in solution (see par. 1.12).

Therefore, the intrinsic disorder of HMGA proteins prevents the evaluation of any kind of intra-molecular interaction by conventional strategies. In fact, no data are currently available as to whether they can assume any particular preferred conformation (Sgarra et al., 2010).

In this work we propose to adopt an innovative combination of non-canonical structural techniques aimed at characterizing conformations in protein disordered states.

**3.9 The C-terminal tail of HMGA proteins shields the protein-protein interaction domain**

As a first step in evaluating a possible conformational role for HMGA C-terminal tail, we investigated the structural properties of HMGA by limited proteolysis. This technique, coupled with mass spectrometry detection, is a strategy commonly adopted to highlight those protein regions which are flexible and more exposed to solvent. Indeed, these regions are expected to be more susceptible to proteolytic hydrolysis with respect to protein regions embedded in stable protein secondary or tertiary structures (Fontana e al., 2004). Limited proteolysis has been shown to be particularly useful for the conformational characterization of partly folded states of proteins or molten globule states which are challenging to be analyzed by high resolution techniques (Fontana et al., 1997). Conversely, it can evidence both protein portions not immediately accessible to
proteolysis involved in three-dimensional structures and changes in protein accessibility caused by conformational alterations.

Figure 3.22 Spectroscopic characterization by circular dichroism (CD) of full-length (FL) and C-terminal truncated (CT) HMGA2. A schematic representation indicating length and functional domains of analyzed proteins is reported on the right (yellow boxes: AT-hooks; red box: acidic tail).

The limited proteolysis approach for probing protein conformation implies that the proteolytic event should be dictated by the stereochemistry and flexibility of the protein substrate and not by the specificity of the attacking protease. To this aim, the most suitable proteases are those displaying broad substrate specificity (Fontana et al., 2004). However, HMGA proteins have a peculiar primary structure: about ¼ of their amino acid residues is a K or R and the vast majority of them (K/R-X≠P) are cleavable by trypsin. As it is possible to see from HMGA2 sequence indicated in Figure 3.22, K and R residues are dispersed along the entire protein length. Most of them are located at the level of HMGA DBDs (K25, 33, 45, 52, 73, and 81, and R26, 28, 32, 48, 76, and 80), five K are concentrated between the second and the third DBD (K55, 57, 61, 65, 66), one K is located just before the acidic C-terminal tail (K89), and one R is located within the N-terminus (R3). Therefore, we started our limited proteolysis experiments digesting FL and CT HMGA2 with
trypsin. We performed trypsin digestions inside nano-electrospray ionization capillaries in order to directly follow the ongoing of the proteolytic cleavage by mass spectrometry.

In Figure 3.23 we show a schematic representation of the results obtained. Both FL and CT HMGA2 forms are very rapidly cleaved by trypsin at R26, R28, R32, R48, and R76 (indicated as blue arrows). Surprisingly, all the other K and R residues (indicated as red circles) are initially resistant to cleavage. Therefore, not all potential trypsin proteolytic sites on both HMGA2 forms are randomly cleaved, as would be expected for a fully accessible, flexible, and unstructured protein (Fontana et al., 1997 and 2004). This implies that HMGA2 acquires some bended and hence "protected" conformation. The first sites of cleavage within both FL and CT HMGA2 are located in their DBDs suggesting that these domains have a higher accessibility with respect to the other portions of the protein. On the other hand, the majority of K residues is initially resistant to trypsin hydrolysis.

![Figure 3.23](image_url)

**Figure 3.23 Differential accessibility to proteolytic cleavage of HMGA2 residues.** Schematic representation of the results obtained from full-length (FL) and C-terminal truncated (CT) HMGA2 limited proteolysis experiments performed with trypsin. The sequence and functional domains of HMGA2 are reported. Blue arrows indicate the very first sites cleaved by trypsin; red dots indicate protected or delayed cleaved sites.

The most informative data in proteolysis experiments are obtained by the detection of the very first cleavage events (Fontana et al., 1997 and 2004), since those occurring later involve already digested protein portions. For this reason, the fact that both FL and CT HMGA proteins are cleaved very rapidly and equally by trypsin at the level of R residues within their DBDs suggests that they are not fully-extended proteins but prevents the evaluation of any difference between FL and CT forms. However, given that all K residues are embedded in initially protected protein regions, we decided to perform comparative limited proteolysis experiments with FL and CT HMGA2 changing the enzyme specificity and using endoprotease Lys-C, that specifically cleaves peptide bonds C-terminally at K. We carried out these experiments in batch, stopping the reactions after 5 and 15 min. Obtained peptides were separated by HPLC and detected by both UV absorbance and MS analyses. We adopted this analytical strategy in order to achieve a more sensitive and quantitative peptide detection with respect to digestion reactions performed in capillaries.
In Figure 3.24, panel A, we show the overlaid UV chromatograms obtained injecting equimicrogram quantities of entire CT (blue) and FL (red) HMGA2. The identical chromatographic peaks obtained for the two proteins (both in terms of height and area) assessed the reliability of our protein quantifications. Thereafter, we performed Lys-C limited proteolysis reactions using equimolar quantities (2 nmoles) of CT or FL protein. In Figure 3.24, panel B, we show the overlaid UV chromatograms obtained by injecting FL and CT proteins (200 pmoles each, equivalent to 2 µg and 1.7 µg of protein, respectively) after 5 min of Lys-C digestion. In panels C and D the LC-MS mass spectra and a summary of the identified peptides are reported, respectively.

Comparing both the UV chromatograms and the mass spectra there is a clear difference in peptide signals intensities of FL and CT, being higher in the latter. In fact, comparing the chromatographic profiles of CT an FL in panel B, besides non digested proteins peaks, there is a difference in the abundance of the two chromatographic peaks eluting around min 25. The averaged mass spectrum obtained from the region containing these peaks (evidenced in the insert of panel B) is reported for each protein in panel C, where the different intensities (cps/10^5) of m/z signals obtained for CT and FL protein are compared. Mass spectra deconvolution allowed the assignment of these signals to peptides generated by cleavage at K residues embedded in HMGA protein-protein interaction domain, as indicated in panel D. Since the abundance of these peptides is higher in the CT form we can conclude that the presence of the C-terminal tail partially impairs the Lys-C endoprotease activity. This clearly suggests a shielding effect of the C-terminal tail towards the protein-protein interaction domain. The same results have been obtained from the analysis of 15 min digestions (Figure 3.25), thus confirming the different accessibility of CT and FL proteins. Only N-terminal-containing peptides were evaluated to compare Lys-C cleavage rate between CT and FL forms since C-terminal-containing peptides (eluting at 27-29 min) are obviously different between the two protein forms and cannot be quantitatively compared both in UV and MS detection.

In support of the observed differences between FL and CT HMGA forms, Lys-C digestions, let proceed up to one hour, lead to the complete disappearance of CT m/z signals but not of those of the FL protein (data not shown).

Therefore, by using limited proteolysis we overcame limitations of conventional strategies in obtaining structural data regarding ID proteins. The fact that both trypsin and Lys-C have preferential cleavage sites supports the idea that FL and CT HMGA forms do not assume fully extended and accessible conformations. Moreover, the different rates of Lys-C proteolysis between the two proteins clearly indicate a difference in accessibility, suggesting a shielding effect caused by the presence of the C-terminal tail.
Limited proteolysis by Lys-C cleavage

Figure 3.24 Differential accessibility to Lys-C proteolytic cleavage of CT and FL HMGA2 forms (5 min digestion). A. overlaid view of UV chromatograms (Abs 220 nm) relative to LC-MS separations of equivalent quantities (2 µg) of FL HMGA2 (red) and CT HMGA2 (blue) to assess the reliability of the protein quantification method adopted. B. Overlaid view of UV chromatograms (Abs 220 nm) relative to LC-MS separations of limited proteolysis performed with Lys-C and equimolar quantities of FL (red) and CT (blue) HMGA2 (injected 200 pmoles –
equivalent to 2 µg and 1.7 µg of FL and CT, respectively). Digestion stopped at 5 min. In the grey insert an enlarged view of the chromatographic peaks corresponding to peptides 1-73, 1-65, 1-61, 1-57, and 1-55 is shown. C. Mass spectra of Lys-C generated peptides eluting in the chromatographic region corresponding to the peaks shown in the insert of panel B. D. Schematic representation of the LC-MS identified peptides. Asterisks indicate m/z signals at 639.2 common to both peptides 1-61 and 1-55. The intensity of m/z 639.2 belonging to peptide 1-55 has been arbitrary attributed considering the whole charge state distribution (CSD) of this peptide.

Figure 3.25 Differential accessibility to Lys-C proteolytic cleavage of CT and FL HMGA2 forms (15 min digestion). Overlaid view of UV chromatograms (Abs 220 nm) relative to separations of limited proteolysis experiments performed with Lys-C and equimolar quantities of FL (red) and CT (blue) HMGA2 (injected 200 pmoles – equivalent to 2 µg and 1.7 µg of FL and CT, respectively). Digestion stopped at 15 min. In the grey insert an enlarged view of the chromatographic peaks corresponding to peptides 1-73, 1-65, 1-61, 1-57, and 1-55 is shown. C-terminal peptides deriving from CT form are also indicated, whereas those deriving from FL protein are not visible since coelute together with the entire protein.

3.10 Conformational role of HMGA acidic tail

In order to further explore a conformational role for HMGA C-terminal tail we took advantage of the Ion Mobility Separation-Mass Spectrometry (IMS-MS) technique, which has recently emerged as a very potent tool to probe for the presence of multiple conformational isoforms of proteins or macromolecular complexes. Ion mobility drift cells are able to separate ions on the basis of their size (m), charge (z), and, mostly relevant for a structural point of view, shape. In a simplified vision, ion mobility devices differentiate ions on the basis of the friction they experience traveling within neutral gas molecules under the action of an electric field (Kanu et al., 2008).

An IMS-MS analysis consists of four principal steps: a) native conditions sample preparation; b) ion generation; c) ion mobility separation; d) mass spectrometry analysis. In our experiments pure proteins have been dissolved in native conditions (100 mM ammonium acetate), ions generated by an electrospray ionization (ESI) source and subjected to IMS-MS analyses using a Synapt (Quadrupole-Ion Mobility-Time of Flight) HDMS instrument (Waters, UK) which is equipped with a traveling-wave (T-WAVE) ion mobility separator (Giles, 2004).
The biological relevance of this methodology resides to the fact that it is nowadays recognized that proteins can retain their non-covalent interactions responsible for their native conformations also when they are ionized moving from the liquid to the gas phase (Kalthashov and Abzalimov, 2008). There are two distinct classes of ions to be taken into consideration interpreting an IMS-MS experiment:

- ions belonging to the same molecule but having different m/z ratios. A protein assumes multiple charge states under ESI, acquiring a typical Gaussian charge state distribution (CSD). Therefore, in IMS-MS measurements ions of the same protein which have lower m/z ratios (i.e. a higher charge) will move faster than those having higher m/z ratios (i.e. lower charge);
- ions belonging to the same molecule having the same m/z value, but different shapes. Considering a single charge state of a ionized molecule that can assume two distinct conformations (compact or extended), more compact ions will experience a lower number of collisions with the neutral gas and will therefore be less slowed in their motion with respect to the extended ones. Therefore, when comparing equal m/z values, smaller or more compact ions are transmitted across the drift cell faster (higher mobility) than larger ions (lower mobility) (Kanu et al., 2008).

We IMS-MS analyzed CT, FL, and also FL C-terminal phosphorylated HMGA2 forms. Since C-terminal tail of HMGA proteins is constitutively phosphorylated by CK2 we investigated a possible conformational role for this PTM using in vitro CK2 phosphorylated HMGA2 taking into consideration only the higher phosphorylation states (HMGA2 4P and 5P).

Figure 3.26 A-C shows the driftscope views of IMS-MS separations, in which the drift time is reported against the m/z ratio and the colors indicate the relative intensity of the m/z signals (black, blue, red, and yellow indicate increasing intensities, respectively). Figure 3.26 D-G shows the mobilograms of the 7+ ions of HMGA2 forms analyzed, in which the drift times are reported versus the intensity of m/z signals.

As it is possible to see from the driftscope views and as already evidenced before, ions with low m/z values display a higher speed in comparison with those having high m/z values. Moreover, driftscope views show the existence of two CSDs for each protein, one centered at low m/z values (11+ or 14+), the other at higher ones (8+). In accordance to this, separate electrospray ionization-mass spectrometry (ESI-MS) analyses, performed under near-physiologic conditions, also indicate that HMGA2 can assume both a low- and a high-charged CSD (Figure 3.27). The distribution of charge states observed in mass spectra is an indicator of physical and structural properties of the
protein molecules in solution (Kalthashov and Abzalimov, 2008). In fact, variation of charge state distribution of a given protein in dependence of the analytical conditions adopted is strictly linked to tertiary structure alterations (Konermann and Douglas, 1997). In Figure 3.27 we show HMGA2 protein CSDs compared to those of the two tightly folded proteins Myoglobin and Cytochrome C, both in acidic/denaturing (0.1% formic acid, 10% acetonitrile) and near-physiologic/native (50 mM ammonium acetate, 5% acetonitrile) conditions. Reconstructed mass spectra are reported in panel A, B, and C, whereas in panels A1, B1, and C1 and A2, B2, and C2, m/z spectra obtained in acidic or near-physiologic conditions are shown, respectively. In panel A1 HMGA2 assume a single CSD centered on 16$^+$; differently, in panel A2 it undergoes a shift in the CSD towards less charged states and a second CSD appears, centered on 8$^+$ charge state. A shift in the CSD of a protein from lower m/z values towards higher ones indicates an increased compactness of protein tertiary structure. In fact, a CSD centered at higher m/z values (lower charge states) reflects reduced accessibility to protonation, and therefore a more compact conformation of the protein (Kalthashov and Abzalimov, 2008). Indeed, proteins known to have a defined three-dimensional structure, such as Cytochrome C and Myoglobin only in native conditions distribute on few low charge states (Konermann and Douglas 1997 and 1998), while in denaturing conditions display higher CSDs reflecting unfolded or partially unfolded states (A-A2, B-B2). The presence of a bimodal CSD for ID HMGA2 protein in native conditions (panel A2) is a strong indication that it could assume a more compact conformation distinct from a fully denaturated and extended one, which, on the other hand, became prevalent when denaturing conditions are employed (panel A1). ID proteins are depicted as dynamic ensembles in which the atom positions vary significantly over time with no specific equilibrium values and the energies required to undergo conformational transitions are low (Uversky and Dunker, 2010); this could explain the different behaviour of HMGA2 with respect to structured proteins.

It is relevant to mention that the bimodal CSD of HMGA2 obtained with both Synapt HDMS instrument (IMS-MS measurements) and HCTultra ion trap mass spectrometer (ESI-MS analyses), which are equipped with a conventional orthogonal ESI interface, were reproduced with a Q-TOF Micro mass spectrometer equipped with a Z-spray nanoflow ESI interface (data not shown), underlining that these observations are instrumentation-independent.

Since protein low-charged CSDs are representative of protein compact/native conformation, while high-charged CSDs are usually attributed to extended/denatured conformations only ions carrying a low number of charges were chosen for the IMS-MS comparison. In the mobilograms of 7$^+$ ions of the different HMGA2 forms (CT, FL, FL 4P and 5P; Figure 3.26 D-G) all the proteins show two distinct ion mobility peaks, indicating that all of them are present in at least two different
conformations. Assuming that these HMGA2 forms share the same charge and shape, they should have different drift times depending on their masses. Smaller ions move faster than bigger ones because of the higher acceleration they experience and the lower friction they encounter in the drift tube (Figure 3.28 A). Therefore, assuming that CT and FL HMGA2 forms have the same shape, the lighter 7+ CT ions should have a lower drift time than 7+ FL ions. As shown in panels D and E, this is not the case. CT HMGA2 ions are distributed between two distinct conformations, one of which with a drift time even higher than that of FL ions (S: slow mobility). This peculiar behaviour indicates that the effect of shape is predominant with respect to the contribution of the mass of the protein. This clearly points out that the CT form assumes a conformation with a steric hindrance comparable or even higher to that of the longer FL (see Figure 3.28 B-D for a schematic representation of this phenomenon).

Since CT and FL have a low CSD (see driftscope views reported in panel A and B) we can consider that both forms assume a compact conformation, but the lower mobility of 7+ CT ions underlined that the presence of the acidic C-terminal tail is responsible for further compacting the protein with respect to the CT form.

As mentioned above, FL form (Figure 3.26 E) is present in two distinct conformations (M: medium mobility and F: fast mobility). Considering FL 4P and FL 5P (Figure 3.26 F, G), it is clear that the abundance of the F form increases with the presence and number of phosphate groups. These observations indicate that the acidic C-terminal tail allows HMGA2 to assume the more compact conformation (F) and that the equilibrium between F and M forms shifts towards the F form when the acidic C-terminal tail is phosphorylated.

Same kind of conclusions can be draw looking at the 8+ ions (Figure 3.26 H-K), even if with a lower resolution, and at HMGA1a IMS-MS analyses (Figure 3.29).

The conformational effect of the acidic tail highlighted by limited proteolysis experiments is thus strongly supported by IMS-MS analyses. In fact, we suggest that the shielding effect of the C-terminal tail, in the context of the FL protein, causes a compaction of the whole molecule. Moreover, IMS-MS data strongly suggest that acidic C-terminal tail phosphorylations could be involved in the modulation of these intramolecular interactions enhancing protein compaction.
Figure 3.26 Ion Mobility Separation-Mass Spectrometry (IMS-MS) analyses reveal a conformational role for HMGA2 acidic tail. A, B, and C. Driftscope views of C-terminal truncated (CT), full-length (FL), and C-terminal tail phosphorylated (FL 4P & FL 5P) HMGA2, respectively. The drift time (abscissa) is plotted vs. the m/z (ordinate); the
abundances of ions are given by a color graduation starting from black (no signal), to blue, red, and yellow (high intensity). The $7^+$ charge states (signals corresponding to compact HMGA2 proteins chosen for ion mobility measurement shown in panels D-G) are showed in white. $8^+$ and $11^+/14^+$ signals correspond to the center of compact and extended HMGA2 m/z distributions, respectively. **D-G and H-K** Arrival time distribution (mobilogram) for $7^+$ and $8^+$ ions belonging to CT, FL, FL 4P, and FL 5P HMGA2 forms. The drift time (abscissa) is plotted vs. the relative intensity (ordinate); The dashed lines allow a better alignment visualization of the various HMGA1a forms detected: F (green), M (blue), and S (red) indicate ions with a fast, medium and slow mobility, respectively.

---

**Figure 3.27** Electrospray Ionization-Mass Spectrometry (ESI-MS) analyses reveal a compact conformation for HMGA proteins. HMGA2 full-length (FL) has been ESI-MS analyzed together with Cytochrome C (Cyt C) and Myoglobin (Myo). **A-C.** Reconstructed mass spectra of analyzed proteins. **A1-C1.** m/z spectra obtained under acidic conditions. **A2-C2.** m/z spectra obtained under near-physiological conditions. Charge state distributions (CSDs) attributed to a native or “compact” protein conformation are evidenced by a red dotted line whereas those attributed to a denatured or “extended” protein conformation by a blue dotted line. Partially unfolded CSD of Cytochrome C in denaturing conditions is evidenced in green. As regards Myoglobin, the reconstructed mass spectrum reported (E) is relative to apo-Myoglobin (m/z spectrum in E1), whereas the m/z spectrum of Myoglobin obtained under native condition (E2) is relative to Myoglobin associated to the heme group (molecular mass: 17567.2 Da).
Figure 3.28 Schematic representation of Ion Mobility Separation-Mass Spectrometry (IMS-MS) results. A. According to IMS-MS theory, acidic tail C-terminal truncated (CT) HMGA2 should have a lower drift time with respect to full-length (FL) HMGA2 protein. B. Schematic representation of the expected relative positions of CT and FL forms in an overlaid mobilogram. C. Schematic representation of the experimentally obtained relative positions of CT and FL forms in an overlaid mobilogram. D. Interpretation of the obtained IMS-MS results in which the higher velocity and lower steric hindrance of FL forms with respect to CT is attributed to a more compact conformation.
Figure 3.29 Ion Mobility Separation-Mass Spectrometry (IMS-MS) analyses reveal a conformational role of HMGA1a acidic tail. Arrival time distributions (mobilograms) for 8+ and 10+ ions belonging to HMGA1a 1-89, HMGA1a, and HMGA1a 3P (C-terminally phosphorylated) forms. The dashed lines allow a better alignment visualization of the various HMGA1a forms detected. Blue and red lines indicate ions with fast and slow mobility, respectively.
3.11 C-terminal truncation of HMGA proteins makes them more susceptible to PTMs

In order to provide a functional significance to the structural features showed above and considering that HMGA activities are finely modulated by several PTMs, we asked ourselves whether there is a structure/PTM relationship involving the acidic C-terminal tail. For this purpose, we in vitro modified HMGA with protein arginine methyltransferase 1 (PRMT1) and 6 (PRMT6) and cyclin dependent kinase 1 (CDK1) in order to compare FL and CT forms with respect to their susceptibility to enzyme activities.

As shown in Figure 3.30, PRMT1 and PRMT6 are known to methylate HMGA1a at the level of several R residues embedded in its AT-hooks, while CDK1 phosphorylates both HMGA1a and HMGA2 at the level of S/T residues located before the second AT-hook and inside the protein-protein interaction domain (our data; Schwanbeck, 2000; Zhang and Wang, 2007 and 2008).

For our comparative experiments we performed enzymatic reactions incubating in the same batch equimolar amounts of FL and CT HMGA1 or HMGA2 protein forms with the selected modifying enzyme. The methyl-transferase activity towards HMGA1a proteins has been checked by LC-MS analyses (Figure 3.31) and radiolabeling experiments (Figure 3.32). Instead, time course experiments in combination with LC-MS detection have been performed to follow the kinetic of HMGA1 and HMGA2 phosphorylation (Figure 3.33).

Figure 3.31A and B shows the reconstructed mass spectra of FL and CT HMGA1a forms modified by recombinant GST-fused PRMT6 and PRMT1, respectively. Both PRMTs are able to methylate CT HMGA1a much more efficiently (up to 4 methyl groups by PRMT6 and up to 8 methyl groups by PRMT1) with respect to FL HMGA1a. These data were confirmed by parallel experiments performed with [$^3$H]-SAM and visualized by fluorography (Figure 3.32).
Figure 3.31 Loss of HMGA1a acidic tail increases PRMTs methylation – LC-MS detection of methyl groups incorporation. Reconstructed mass spectra, obtained by LC-MS analyses, of C-terminal tail truncated (CT) and full-length (FL) HMGA1a forms in vitro methylated by PRMT6 (A) or by PRMT1 (B). M indicates the addition of a methyl group.

Figure 3.32 Loss of HMGA1a acidic tail increases PRMTs methylation – Fluorographic detection of radioactive methyl groups incorporation. Equimolar quantities of HMGA1a full-length and C-terminal truncated forms have been subjected in the same batch to in vitro methylation assays with GST-PRMT1, GST-PRMT6, and GST alone in the presence of [3H]-SAM. Proteins have been separated by SDS-PAGE (T=15%) and blue coomassie stained (panel A). The same gel was subjected to fluorographic detection to visualize [3H]-methyl groups incorporation (panel B). Protein molecular weight markers are indicated (kDa).
Figure 3.33 shows the reconstructed mass spectra of FL and CT HMGA1 (panels A-E) and HMGA2 (panels F-J) forms modified by commercially available CDK1 during a time course experiment of five time points (0.5, 2.5, 5, and 16 h). As it is possible to see, both CT HMGA1a and CT HMGA2 forms are more rapidly phosphorylated than FL proteins. In fact, at each time point CT HMGA forms have a higher phosphorylation degree with respect to FL proteins. Looking for example at reconstructed mass spectra shown in panel G, after 1h of phosphorylation CT HMGA2 protein is fully phosphorylated being distributed in the 1P (mainly) and 2P forms, whereas its FL counterpart is not entirely modified and is distributed mainly in the 0P and 1P forms.

![Figure 3.33 Loss of HMGA acidic tail increases CDK1 phosphorylation rate](image)

Reconstructed mass spectra, obtained by LC-MS analyses, of time course CDK1 phosphorylation assays (0.5, 1, 2.5, 5, and 16 hours) performed with C-
terminal tail truncated forms (CT) and full-length (FL) HMGA proteins (HMGA1a and HMGA2, panels A, B, C, D, and E and F, G, H, I, and J, respectively). P indicates the addition of a phosphate group. Un-, mono-, bi-, and tri-phosphorylated HMGA peaks are highlighted in blue, violet, pink, and green, respectively.

Given the different behaviour of FL and CT HMGA proteins with respect to enzymatic activities it is possible to conclude that the HMGA C-terminal tail is involved in hampering the accessibility of the modifying enzymes to their specific consensus sites.
4. Conclusions

HMGAs belong to the HMG super-family of architectural transcription factors. They play a role in the transcription of many genes involved in the neoplastic transformation process and have been proposed to be diagnostic markers for both tumor genesis and metastatic progression as well as potential therapeutic targets for cancer chemotherapy (Sgarra et al., 2004, Fusco and Fedele, 2007). However, no useful structural informations for drug development are available for HMGAs since they constitute prototypes of intrinsically disordered proteins (IDPs) lacking a unique and fixed protein conformation. Moreover, HMGAs can be affected by a big number of post-translational modifications, both constitutive and dynamic. The HMGa protein family is composed of three members sharing a high degree of sequence and protein domain homology. However, several experimental evidences suggest that these proteins may have non-overlapping functions. Differences in HMGa DNA binding properties (Maher and Nathans, 1996), their non-overlapping biological functions (Fedele et al., 2006; Zhou et al., 1995), and their different expression patterns (Reeves et al., 2001, Fusco and Fedele, 2007) have already been pointed out, and it was also suggested that their activity could be differentially modulated by post-translational modifications (Edberg et al., 2005; Zou and Wang, 2007). HMGAs are highly abundant chromatin proteins, whose expression level, in particular during neoplastic transformation, can reach a 1/10 ratio with respect to histone H1 (Giancotti et al., 1993). This means that, beside their involvement in regulating the expression of specific genes they can play a widespread influence on chromatin structure (Sgarra et al., 2004; Fusco and Fedele, 2007). This has, for example, been underlined by their involvement in formation of senescence-associated heterochromatic foci (SAHFs) (Narita et al., 2006).

Therefore, we asked whether beside HMGa family members’ differences in the primary structure other differences can affect the bulk of HMGa molecules being likely involved in modulating HMGa general chromatin activities. Our LC-MS screening, performed on a relevant number of cell lines differing both in tumorigenic phenotype and in their origin, evidenced macroscopic differences between the post-translational modifications (PTMs) affecting the three HMGa (reassumed in figure 4.1) that are linked to the cellular context. These data are significant since cell-type dependent changes in PTMs are specific for all three HMGa architectural factors, but are not observable for the highly related HMGN proteins. Moreover, the HMGa PTMs fluctuation that we evidenced suggests, even if doesn’t directly demonstrate, that the activities of HMGa proteins could be mainly and generally modulated by macroscopic PTMs. These observations suggest that
the three HMGA proteins may participate in different cellular processes being HMGA1a, HMGA1b and HMGA2 affected by unique PTM patterns.

The acidic C-terminal tail phosphorylation has up to now been considered the only constitutive modification of HMGA protein, but our screening evidenced, at least for the HMGA1a isoform, that methylation (which has always been mapped at the level of Arginine 25) could also be considered an HMGA constitutive modification, which function, however, is completely unknown. R25 is embedded in the first HMGA1a DBD and can be mono- (Sgarra et al., 2003 and 2003bis) or di-methylated, both symmetrically and asymmetrically (Zou and Wang, 2005). This kind of modification does not alter R side chain charge, but can increase its steric hindrance and affect hydrogen bonding interaction(s) (McBride and Silver, 2001). From our unpublished data, R25 methylation is not involved in modulating HMGA1a binding affinity towards DNA. Nevertheless, recent issues demonstrated that the first DBD is involved in HMGA1a interaction also with RNA (Norseen et al., 2008; Manabe et al., 2003), thus suggesting that R25 methylation could be implicated in its modulation. However, we cannot also exclude a role of this PTM in modulating interaction with protein molecules, even if HMGA protein-protein interaction domain have been mapped in a different protein region.

Our PTM data regarding C-terminal phosphorylation and R25 methylation are consistent with previous findings regarding HMGA1a post-translational modifications in human breast tumor specimens (Zou and Wang, 2007). Instead, our LC-MS and LC-MS/MS data regarding in vivo HMGA2 PTMs give completely new information. These data support the evidence of non overlapping biological functions among HMGA proteins. Indeed, the phosphorylation pattern of HMGA2 is completely different from that of HMGA1 proteins. This is primarily linked to the different C-terminal amino acid composition of the two proteins; in fact HMGA2 possesses five CK2 phosphorylatable amino acids (S/T) compared to HMGA1a which has only three. HMGA phosphorylation at the C-terminal acidic tail suggests a modulatory role for this domain in regulating HMGA activity. In the past, several lines of evidence underlined the functional importance of the C-terminal domain for both HMGA1 and HMGA2 phosphoproteins (Fusco and Fedele, 2007). Cells expressing C-terminal tail truncated (CT) HMGA forms have been shown to display a more aggressive neoplastic phenotype with respect to those expressing full-length (FL) HMGA. In fact, CT HMGA forms either show an increased transformation potential (Fedele et al., 1998; Li et al., 2007) or confer to cells an increased growth rate (Pierantoni et al., 2003bis) with respect to their full-length counterparts. Therefore, the current idea is that the acidic C-terminal domain plays a negative modulatory role in the regulation of proliferative and transforming ability of HMGA proteins. Moreover, very recently different biological effects of CT HMGA1 with
respect to FL form have been observed in transgenic mice (Fedele et al., 2011). In addition, a gene-specific transactivating function has been assigned to the acidic tail of HMGA2 (Thuault et al., 2008) and gene expression analyses showed that overexpression of CT HMGA2 in mesenchymal stem-like cells affects the expression of a higher number of genes than FL HMGA2 (Henriksen et al., 2010). All these observations suggest a functional and modulatory role of the C-terminal domain towards HMGA activities.

Structural data are normally required to unravel the function of specific domains within a given protein, unveiling their three-dimensional localization and relationship within the whole protein and suggesting possible molecular mechanisms. These approaches were not however suitable for HMGA given their intrinsically disordered (ID) status. These proteins are described as multifunctional, with no specific three-dimensional organization, and displaying two main biochemical properties: the ability of binding with a “relaxed” specificity to DNA and of interacting with a huge number of other molecular partners (Sgarra et al., 2010). Importantly, these properties are also finely modulated by several PTMs, adding a further level of complexity to HMGA activities/functions.

Taking into account both the primary sequence and the contribution of PTMs, we demonstrated that although ID, HMGA proteins are not extended but, on the contrary, acquire a compact conformation, requiring a more sophisticated interpretation of the term “intrinsically disordered”. Indeed, this term suggests the lack of a fixed three-dimensional structure but it does not exclude that the polypeptide chain preferentially acquires some bended conformations which are favoured by the formation of intramolecular interactions. All the biochemical and MS techniques used to assay HMGA conformations support this view. The most relevant result of this work is that we provide a link between a biochemical property of HMGA proteins and their structural features, i.e. the accessibility with respect to a set of modifying enzymes is linked to the presence of their acidic tail. Our data support a model in which the tail masks the protein-protein interaction domain causing a protein conformation alteration that impairs enzymatic accessibility. Moreover, IMS-MS analyses evidence a role for constitutive phosphorylation of the C-terminal tail in enhancing the compactness level of HMGA proteins. In figure 4.2 we present the model of HMGA conformational transitions that our MS-based approach allowed to depict. The obtained results allow us to better interpret our previous data demonstrating that truncated HMGA2 has altered DNA- and protein-protein interaction properties (Noro et al., 2003; Sgarra et al., 2005) and the findings reported in this thesis showing that HMGA2 acidic tail phosphorylation modulates DNA-binding affinity. In analogy with the HMGB1 protein (Kawase et al., 2008; Watson et al., 2007; Wang et al., 2007; Knapp et al., 2004; Pasheva et al., 2004), we thus propose a model in which intra-molecular interactions
occurring between the tail and inner positive portions of HMGA proteins lead to a compaction of the whole molecule. The resulting conformational alterations could be responsible for the different transforming ability of CT form with respect to FL HMGA proteins. Obviously, the functional difference between CT and FL forms could be also due to other additional mechanisms. One possibility is that the acidic C-terminal tail constitutes a steric hindrance or an electrostatic repeller with respect to HMGA interactions. We cannot furthermore exclude that the C-terminal tail could have a role in modulating HMGA protein stability, as was previously demonstrated for HMGB proteins (Knapp et al., 2004).

Moreover, the influence of an acidic and phosphorylated C-terminal domain on the biological activity of a protein seems to be a rather frequent mechanism of protein function regulation. For example, the C-terminal domain of the tumor suppressor PTEN (Phosphatase and Tensin homologue) acts as an autoinhibitory domain of the catalytic activity of PTEN itself (Odriozola et al., 2007; Raftopoulou et al., 2004) and the acidic C-terminal tail of the ssDNA-binding protein of bacteriophage T7 shields, intramolecularly, the positively charged DNA-binding domain (Marintcheva et al., 2008). Regarding the latter example, this mechanism has been proposed to be a general tool to prevent random bindings between positively and negatively charged protein/nucleic acid surfaces (Marintcheva et al., 2008).

Our conformational data constitute a proof of concept for ID HMGA as compact proteins, implying that not all the amino acid residues are equally accessible. This evidence could be important information for the development of strategies in which HMGA activities are targeted by means of interacting molecules. This is in the light of recent findings showing that molecules able to bind to HMGA1a with high affinity and to compete with HMGA1a cellular partners counteract its oncogenic activity. HMGA-interacting L-RNA oligonucleotides (Spiegelmers) have for example been demonstrated to act as efficient therapeutic agents in a pancreatic adenocarcinoma xenograft mouse model (Maasch et al., 2010). Our data suggest that the region comprised between the second and the third DBD could be a potential target to interfere with HMGA activities. Indeed, this region seems to be the one most protected by the C-terminal tail and thus most linked to the negative modulatory role of the tail with respect to HMGA oncogenic activities. It is striking that the target emerging from this work overlaps almost perfectly with the HMGA protein-protein interaction domain (Sgarra et al., 2010).

To our knowledge, this is the first example of a characterization, obtained by IMS-MS, of changes within the conformational equilibria of an ID protein due to PTMs or sequence deletion. ID proteins constitute a growing protein family playing crucial biological functions (Uversky and Dunker, 2010). Since this technique preserve native structures and is also used to study protein complexes, it
could be adopted as a screening method to assess the influence of both PTMs and small interacting molecules on ID proteins conformations, thus revealing their efficiency in perturbing a “native” condition and, hopefully, protein functions.

Figure 4.1 HMGA macroscopic PTMs. HMGA1a, HMGA1b and HMGA2 macroscopic PTMs that have been detected with our LC-MS screening are indicated.

Figure 4.2 A model for HMGA conformational transitions. ESI-MS and limited proteolysis data suggest that HMGA proteins could be in equilibrium between a fully extended and a compact conformation (1). IMS-MS suggests that phosphorylation at the acid tail could enhance such compactness (2) and that truncated HMGA forms, on the contrary, could assume a more relaxed structure (3). These structural differences are responsible for a different accessibility to modifying enzymes (4) highlighting that there could be a structural/functional relationship at the basis of the enhanced transforming ability of truncated HMGA (5).
Collaborations

I would like to thank Dr. Barbara Spolaore from CRIBI Biotechnology Centre, University of Padua for her expertise and collaboration in the set up of limited proteolysis experiments and for having performed trypsin digestion experiments inside nano-electrospray capillaries.

I would like to thank Liam Brady and Laetitia Cravello from Waters corporation in Manchester, UK for having performed ion mobility separation-mass spectrometry measurements.
References

- Bianchi, ME. Interaction of a protein from rat liver nuclei with cruciform DNA. EMBO J. 1988, 7, 843-49.
- Bullerdiek J, Rommel B. Comment re: HMGA2 is a negative regulator of DNA-PK pathway. Cancer Res. 2010, 70, 1742; author reply 1742.
References

- Petit MM, Mols R, Schoenmakers EF, Mandahl N, Van de Ven WJ. LPP, the preferred fusion partner gene of HMGIC in lipomas, is a novel member of the LIM protein gene family. Genomics. 1996, 36, 118-29.
M, Umbas R, Karthaus HF, Debruyne FM, Schalken JA. Increased expression of high
ed" proteins unstructured under physiologic conditions?  
Sgarra R, Diana F, Bellarosa C, Dekleva V, Rustighi A, Toller M, Manfioletti G, Giancotti V. During apoptosis of tumor cells
HMGA1a protein undergoes methylation: identification of the modification site by mass spectrometry. Biochemistry. 2003,
42, 3575-85.

Sgarra R, Diana F, Rustighi A, Manfioletti G, Giancotti V. Increase of HMGA1a protein methylation is a distinctive


Sgarra R, Lee J, Tessari MA, Altamura S, Spolaore B, Giancotti V, Bedford MT, Manfioletti, G. The AT-hook of the
chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine


Sgarra R, Zammiti S, Lo Sardo A, Maurizio E, Arnoldo L, Pegoraro S, Giancotti V, Manfioletti G. HMGA molecular network:

Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z,


Summer H, Li O, Bao Q, Zhan L, Peter S, Sathiyanaathan P, Henderson D, Klonisch T,Goodman SD, Dröge P. HMGA2
exhibits DRP/AP site cleavage activity and protects cancer cells from DNA-damage-induced cytotoxicity during
chemotherapy. Nucleic Acids Res. 2009, 37, 4371-84.

Syme CD, Blanch EW, Holt C, Jakes R, Goedert M, Hecht L, Barron LD. A Raman optical activity study of rhamphornorphism in

Tait S, Dutta K, Cowburn D, Warwicker J, Doig AJ, McCarthy JE. Local control of a disorder-order transition in 4E-BP1

Tamimi Y, van der Poel HG, Denyn MM, Embas R, Karthaus HF, Debruyne FM, Schalken JA. Increased expression of high


Giancotti V, Manfioletti G. Transcriptional activation of the cyclin A gene by the architectural transcription factor

Thanos D, Maniatis T. The high mobility group protein HMG I(Y) is required for NF-kappa B-dependent virus induction of the

HMGA1a and the origin recognition complex creates site-specific replication origins. Proc Natl Acad Sci U S A. 2008,
105, 1692-7.

Thuault S, Tan EJ, Peinado H, Cano A, Heldin CH, Moustakas A. HMGA2 and Smads co-regulate SNAI1 expression during

Thuault S, Valscourt M, Petersen M, Manfioletti G, Heldin CH, Moustakas A. Transforming growth factor-beta employs


Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions?

Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu

Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell


References
References

- Zhao K, Käs E, Gonzalez E, Laemmli UK. SAR-dependent mobilization of histone H1 by HMGI-I/Y in vitro: HMGI-I/Y is enriched in H1-depleted chromatin. EMBO J. 1993, 12, 3237-47.
- Zheng J, Machida K, Antoku S, Ng KY, Claffey KP, Mayer BJ. Proteins that bind the Src homology 3 domain of Crkl have distinct roles in Crk transformation. Oncogene. 2010, 29, 6378-89.


The work presented in this thesis resulted in one published article, one submitted manuscript and two congress communications.

**Publication:**

**Submitted manuscript:**
- **Maurizio, E.;** Cravello, L.; Brady L.; Spolaore, B.; Arnoldo, L.; Giancotti, V.; Manfioletti, G. and Riccardo Sgarra. *A conformational role for the C-terminal tail of the intrinsically disordered High Mobility Group A (HMGA) chromatin factors.*

**Congress communications:**

During my PhD period I’ve been involved in other collateral projects not inserted in this thesis work.

**Non-thesis related publications:**
Non-thesis related publications:
- Malini E., **Maurizio E.**, Bembich S., Sgarra R., Edomi P., and Manfioletti G. *HMGA interactome: new insights from phage display technology*. Just Accepted by Biochemistry.

Non-thesis related submitted manuscript:
- Sgarra R., Arnoldo L., Pegoraro S., **Maurizio E.**, Giancotti V., and Manfioletti G. *The histone chaperone Nucleophosmin modulates DNA interactions of the architectural transcription factors HMGA.*

Non-thesis related manuscripts in preparation:

Non-thesis related congress communications: