PhD degree in Molecular Medicine

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Faculty of Medicine

Settore disciplinare: BIO/10

A biochemical and structural study of the kinetochore - centromere interface

Federica Basilico

IFOM-IEO Campus, Milan

Matricola n. R08904

Supervisor: Prof. Andrea Musacchio

IFOM-IEO Campus, Milan

Anno accademico 2012-2013

Table of contents

List of abbreviations

a.u.: arbitrary unit

CA-MN: CENP-A-containing mononucleosomes

CCAN: constitutive centromere-associated network

CENP: centromere protein

GDP: guanosine diphosphate

GTP: guanosine triphosphate

H3-MN: H3-containing mononucleosomes

KMN: Knl1 complex, Mis12 complex, Ndc80 complex

SEC: size-exclusion chromatography

wt: wild type

List of figures

List of tables

Abstract

Faithful chromosome segregation during mitosis requires the dynamic interaction between spindle microtubules and kinetochores, multiprotein complexes built on chromosomal *loci* named centromeres.

A group of kinetochore proteins associates with centromeres throughout the cell cycle and is thus named constitutive centromere-associated network (CCAN). Biochemical and functional analyses indicate that, rather than forming a single stable complex, CCAN proteins are organized in sub-complexes. However, the exact protein composition and organization of these sub-complexes has not been fully elucidated to date. Moreover, investigations on kinetochore localization dependencies of CCAN proteins have revealed an extremely complicated network of relationships.

A thorough mapping of the direct interactions that take place among CCAN proteins is an essential condition to deconvolute such an intricate picture.

The aim of my project has been the biochemical reconstitution of CCAN sub-complexes and their structural and functional characterization. I have been interested in investigating interactions among different CCAN sub-complexes, other kinetochore proteins and centromeric chromatin. In particular, this dissertation dwells upon the results I have obtained regarding three different but intrinsically related topics.

First, I present a biochemical and structural characterization of the CCAN protein CENP-M (centromere protein M), which displays the fold, but not the enzymatic activity of a G protein. In addition, I disclose its unprecedented role in the context of a quaternary complex with CENP-H, CENP-K and CENP-I and provide information about the spatial organization of this complex. The first steps towards an *in vivo* validation of these results are also described.

Second, I report the discovery of a direct interaction of CENP-H / CENP-K complex with CENP-C and discuss its implications for the kinetochore recruitment of these proteins.

Third, I have been involved in establishing in the laboratory techniques for the *in vitro* reconstitution of recombinant nucleosomes. The production of material of good quality and quantity has recently been achieved, supporting the analysis of *in vitro* interactions between nucleosomes and kinetochore components. Specifically, I illustrate some preliminary observations concerning a direct interaction of Mis12 complex with nucleosomes.

1 Introduction

1.1 Overview of kinetochore organization and function

Mitosis is the process through which a eukaryotic mother cell equally distributes its duplicated chromosomes to two daughter cells, so that each of them is provided with a full complement of the genome. The accuracy of this process is crucial for cellular viability as its perturbation leads to aberrations in chromosome number, called aneuploidies, which often result in birth defects and / or cancers[1,2].

Faithful chromosome segregation during mitosis requires the dynamic interaction between spindle microtubules and kinetochores, multiprotein complexes built on chromosomal *loci* named centromeres[3]. We can thus state that the primary function of kinetochores is to create load-bearing attachments between chromosomes and microtubules in a dividing mother cell.

The first insights into the organization of vertebrate kinetochores came from electron microscopy studies[4], which revealed that kinetochores have a trilaminar morphology (Figure 1), with electron dense inner and outer plates and an electron lucent middle layer. In the absence of microtubules, a fibrous corona is visible as a dense array of fibres extending outward from the outer plate. Although such distinct layers are not as apparent using modern sample preparation techniques[5], the concept of an overall layered geometry has been supported by studies showing an arrangement of protein complexes along the inner-outer kinetochore axis.

This organization reflects the main functions of kinetochores, which can be grouped under four modules. The first module, in the inner kinetochore plate, forms an interface with centromeric chromatin[6]. The second module, in the outer kinetochore plate, contributes a microtubule-binding interface[7]. The fibrous corona contains microtubule motors and components of the spindle assembly checkpoint[8], which represents the third module. This provides a feedback control mechanism that monitors the state of kinetochore-microtubule attachments and coordinates it with the progression of the cell cycle. In particular, it delays the entry into anaphase until all sister chromatids are properly aligned along the metaphase plate and attached to opposite spindle poles. This condition, called bi-orientation or amphitelic orientation, is crucial for the equal distribution of sister chromatids to the two daughter cells at anaphase. Finally, the fourth module discerns correct

from improper kinetochore-microtubule attachments, selectively stabilizing the former and favouring the correction of the latter[9,10].

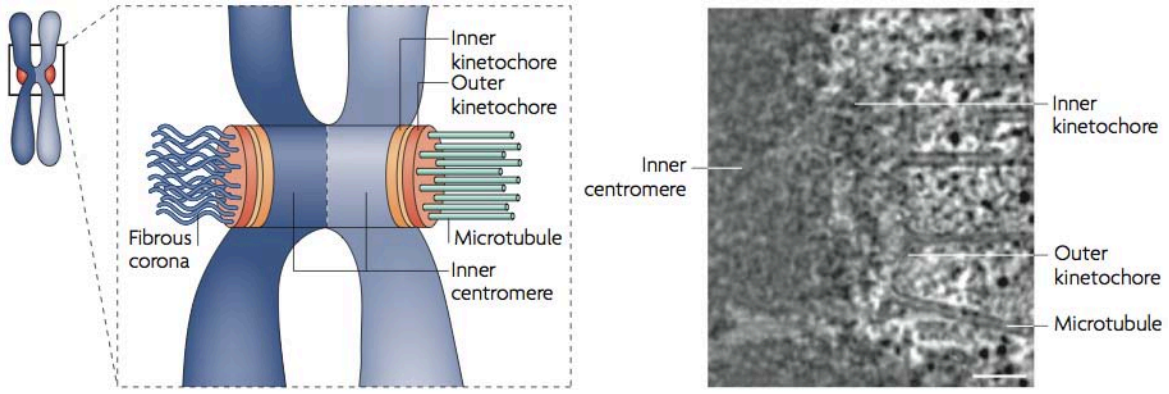


Figure 1 - Vertebrate kinetochore ultrastructure.

Left panel: schematic representation of a mitotic chromosome. The sister chromatid on the right is attached to microtubules, while the sister chromatid on the left is unattached. Right panel: electron micrograph of a human kinetochore (image courtesy of Dong Y. and McEwen B., State University of New York at Albany, USA). The micrograph represents a single slice from a tomographic volume of a high-pressure frozen mitotic cell. Scale bar: 100 nm. Adapted from Cheeseman I. M. and Desai A., *Nat Rev Mol Cell Biol*, 2008[11].

1.2 Conserved and divergent features of kinetochores across evolution

The aforementioned major themes in kinetochore organization and function are conserved throughout eukaryotes, but significant differences in kinetochore arrangement exist among species. Given the relevance of the topic in the context of this dissertation, an overview of conserved and divergent features of kinetochores across evolution is provided below.

Microtubules are polymers of α/β-tubulin dimers. They are extremely conserved cytoskeletal elements and the amino acidic sequences of α and β tubulin are over 80 % identical from yeast to human. The number of microtubules attached to each kinetochore instead varies among species: from 1 in budding yeast, to 2 - 4 in fission yeast, to about 25 - 40 in human. Despite these differences, the microtubule-binding interface of the kinetochore is largely conserved at the molecular level. It principally comprises the so-called KMN network, where the acronym stands for Knl1 complex, Mis12 complex and Ndc80 complex. Ndc80 complex directly contacts microtubules through its Ndc80 and Nuf2 subunits[12]. Although the amino acidic sequences of Ndc80 are only 22

% identical from yeast to human, human Ndc80 can complement the deletion of its yeast ortholog[13]. In addition, structural studies have revealed a conserved calponin-homology domain in the N-terminal region of Ndc80[14], further highlighting its structural and functional conservation despite low sequence identity.

In the past few decades, a number of complementary approaches have been employed to elucidate kinetochore composition in human and in model organisms, resulting in the identification of over 80 kinetochore proteins in human and over 60 in budding yeast. The majority of kinetochore proteins are organized into multiprotein sub-complexes[15-20], which can be isolated from cellular extracts as intact entities using affinity purifications and can be produced as soluble complexes by coexpressing their subunits in heterologous systems such as bacteria or insect cells. The kinetochore localization codependencies of these proteins usually reflect such physical interactions. These complexes appear to be conserved from budding yeast to human, apart from few exceptions (e.g. most notably the CBF3 and Dam1 complexes, which are present in budding yeast but not in human)[21-23]. Systematic sequence analyses of kinetochore proteins have led to the identification of numerous orthologs across species[24,25]. This work is complicated by the fact that most kinetochore proteins appear to have structural roles, rather than catalytic ones, which allows a more rapid evolution of protein sequence. As previously stated, structural and functional conservation, which are often not immediately evident from the amino acidic sequence, are usually the norm[26]. In this respect, biochemical and structural studies are extremely valuable to elucidate kinetochore organization at various levels of complexity and detail, ranging from its overall conformation to high-resolution atomic contacts.

The basis on which the kinetochore is built, the centromeric chromatin, represents a fascinating paradox in genetic inheritance: centromeres possess an extremely conserved function in chromosome segregation, while displaying a highly divergent organization and a rapidly evolving DNA sequence[27]. This topic is more extensively addressed in the next paragraph.

1.3 Centromere structure and specification

On the basis of their overall structure, centromeres can be classified into three categories. Point centromeres represent the simplest case and are limited to a subset of fungi, among which

Saccharomyces cerevisiae. They consist of a defined DNA sequence of about 125 bp, which wraps around a single centromeric nucleosome and is sufficient to specify kinetochore formation. Kinetochores built on point centromeres bind to a single microtubule[21]. Conversely, regional centromeres occupy much longer DNA regions, ranging from about 10 - 40 kb in *Schizosaccharomyces pombe* to millions of bases in human. They assemble kinetochores that bind to multiple microtubules, ranging from 2 - 4 in *S. pombe* to about 25 - 40 in human[28]. Finally, holocentric chromosomes exist in a few organisms, including *Caenorhabditis elegans*. They are characterized by centromeres, and thus kinetochores, that extend all along chromosome length[29].

A more detailed description of centromeric DNA sequences is here provided (Figure 2). Centromere function in *S. cerevisiae,* as already said, is specified by a 125 bp DNA sequence[21]. This comprises three centromere DNA elements (CDEs): CDEI, CDEII and CDEIII. CDEIII recruits a complex, called Cbf3 complex, containing sequence-specific DNA-binding proteins (Cep3, Ctf13, Ndc10 and Skp1). This complex in turn determines the assembly of a single centromeric nucleosome, containing the CENP-A ortholog Cse4, over the AT-rich CDEII. Additionally, at least one Cbf3 subunit, Ndc10, is also found in association with CDEII. CDEI recruits a dimer of Cbf1, a helix-turn-helix protein that runs a parallel life as a transcription factor.

Regional centromeres instead comprise long arrays of repetitive DNA sequences[30]. Centromeres of *S. pombe* extend for about 10 - 40 kb. They consist of a central domain, with a non-conserved core sequence situated between a pair of inverted repeats (called imr, for innermost repeats). The central domain is flanked by outer repeats assembled into heterochromatin. Human centromeric regions comprise megabase arrays of repetitive α-satellite DNA. In particular, this consists of a core of thousands of copies of a 171 bp sequence known as αI-satellite, framed on either side by divergent repetitive sequences and retrotransposons, referred to as αII-satellite DNA. At the outskirts, centromeric chromatin becomes rich in long interspersed element 1 (LINE-1) repeats. The αI-satellite DNA contains a 17 bp sequence called CENP-B box, which binds in a sequence-specific manner to the CENP-B protein and facilitates, but is not strictly required for, kinetochore formation. Usually, human centromeres form on a small subdomain of the αI-satellite DNA, but centromeric DNA is neither necessary nor sufficient to support centromere function. In fact, occasionally, so-called neo-centromeres form on DNA devoid of α-satellite repeats[31]. Also, in stably transmitted dicentric chromosomes, one of the two centromeres is functionally inactivated, indicating that the presence of centromeric DNA does not necessarily lead by itself to the formation

of a functional centromere[32]. The main implication of these observations is that centromere specification does not strictly depend on the underlying DNA sequence. This in turn hints at the existence of epigenetic mechanisms regulating the establishment and maintenance of centromere identity and, therefore, the formation of kinetochores.

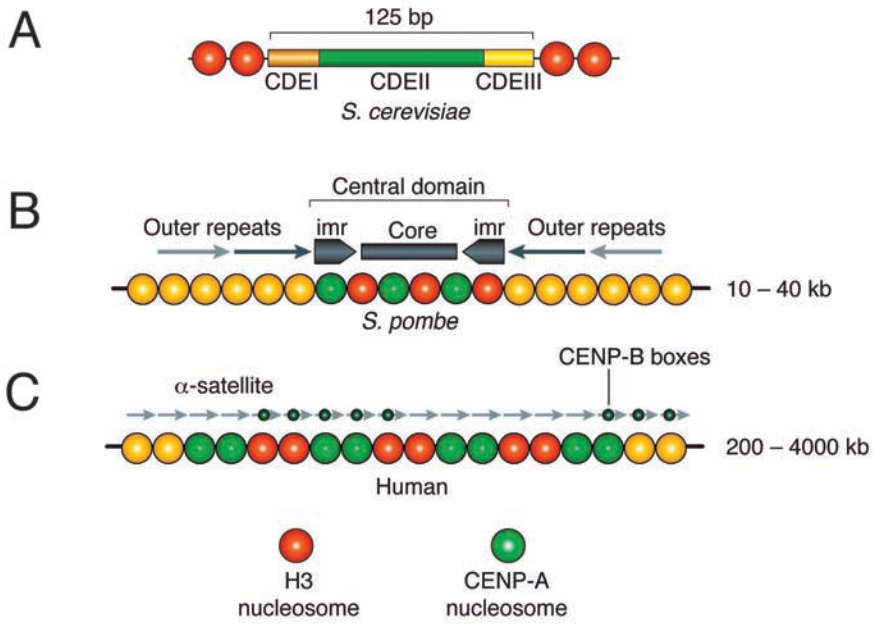


Figure 2 - Organization of centromeres across evolution.

A) Centromere function in *S. cerevisiae* is specified by only 125 bp of DNA. This comprises three centromere DNA elements (CDEs): CDEI, CDEII and CDEIII. A single centromeric nucleosome, containing the CENP-A ortholog Cse4, is assembled over the AT-rich CDEII. B) Centromeres of *S. pombe* extend for about 10 - 40 kb. They consist of a central domain, with a non-conserved core sequence situated between a pair of inverted repeats (called imr, for innermost repeats). The central domain is flanked by outer repeats assembled into heterochromatin. C) Human centromeric regions comprise megabase arrays of repetitive α-satellite DNA. The α-satellite DNA contains a 17 bp sequence called CENP-B box, which binds in a sequence-specific manner to the CENP-B protein. Regional centromeres, such as the ones present in *S. pombe* and human, contain multiple CENP-A nucleosomes interspersed with H3 nucleosomes. Adapted from Santaguida S. and Musacchio A., *EMBO J*, 2009[3].

Extensive work over the past years has defined CENP-A, a centromere-specific histone H3 variant, as the primary epigenetic determinant of centromeric chromatin. Point centromeres in *S. cerevisiae* include a single CENP-A nucleosome[33], called Cse4 in budding yeast, while regional centromeres contain multiple CENP-A nucleosomes interspersed with H3 nucleosomes[34]. Despite this difference, CENP-A invariably represents a hallmark of active centromeres, while it is not present at inactive ones[35]. Furthermore, deletion of CENP-A is lethal[36-39]. In particular, in the absence of CENP-A, cells fail to localize kinetochore proteins and show evident chromosome segregation

defects, indicating that CENP-A is necessary for kinetochore formation. Overexpression of CENP-A in human cells[40] results in its faulty incorporation into widely dispersed non-centromeric sites and in the recruitment of a subset of kinetochore components, including CENP-C. In fact, CENP-A seems to reside at the top of the centromere and kinetochore assembly hierarchy.

1.4 Characteristic features of CENP-A nucleosomes

CENP-A is a highly divergent histone H3 variant (Figure 3). It possesses a variable N-terminal tail, which ranges from 20 to 200 amino acids in size and shows essentially no sequence homology across eukaryotes, with the exception of some arginine-rich motifs, or to the N-terminal tail of histone H3[41]. Also, the histone fold domain, which is an ancient structural element comprising three α-helices connected by loops[42], shares only an average 48 % identity across evolution and to histone H3[41]. Moreover, CENP-A displays a poorly conserved C-terminal tail[43]. These observations indicate that CENP-A evolves rapidly, in contrast with the high conservation of histone H3 and the other canonical histones.

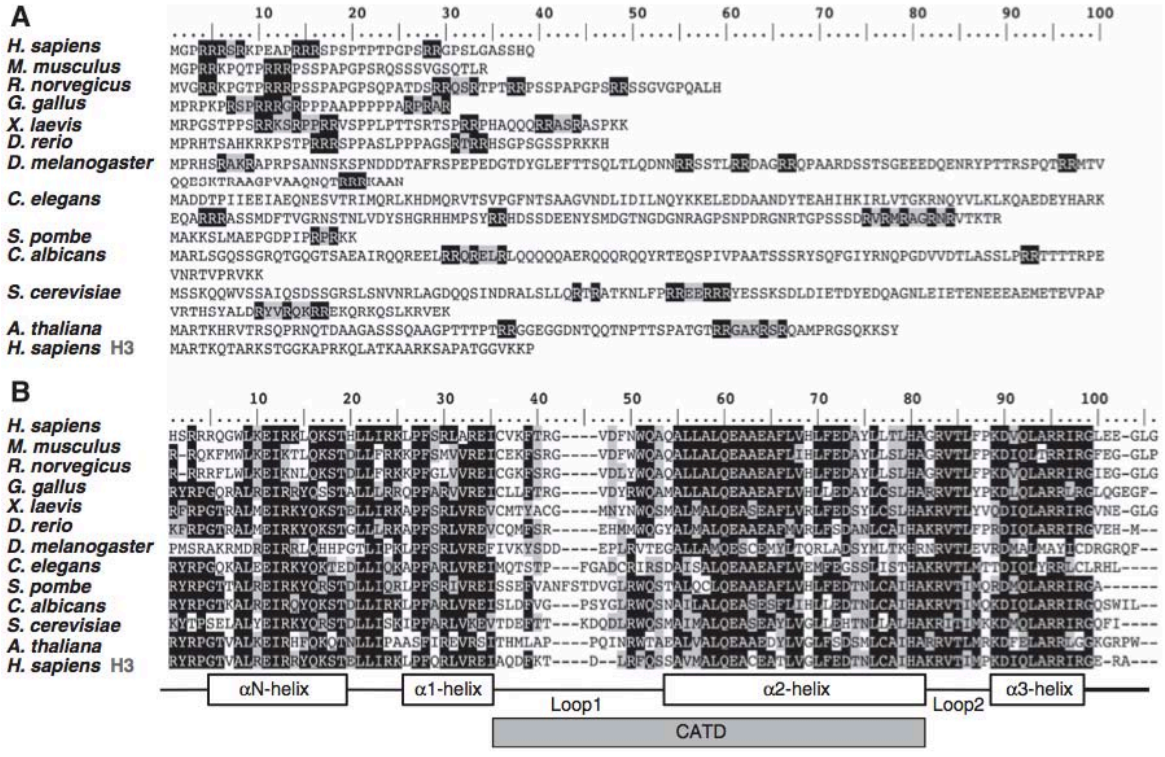


Figure 3 - CENP-A is a highly divergent histone H3 variant.

Sequence comparison of the N-terminal tail (A) and the histone fold domain (B) of CENP-A proteins from different species. The sequence of human histone H3 is shown at the bottom for comparison. Arginine-rich motifs are highlighted in A. Secondary structure elements of the histone fold domain are indicated in B. The position of the so-called CATD (CENP-A targeting domain) is also displayed. Adapted from Torras-Llort M. et al., *EMBO J*, 2009[41].

Despite sequence variability, CENP-A possesses some peculiar characteristics that are conserved across evolution. In particular, the crystal structure of an *in vitro* reconstituted human CENP-A nucleosome[44] revealed that, although its overall arrangement resembles that of the canonical H3 nucleosome[45], in the CENP-A nucleosome structure only the central 121 bp of the DNA are visible and 13 bp from both ends of the DNA are disordered, while in the H3 nucleosome structure 146 bp of DNA are visible. Thus, the DNA regions at the entrance and exit of the CENP-A nucleosome lack a fixed conformation. This is consistent with previous biochemical studies performed in various organisms, which showed that the DNA segments at the entrance and exit of the CENP-A nucleosome are more flexible and accessible to nuclease digestion than those of the canonical H3 nucleosome[46,47].

Moreover, a characteristic property of CENP-A is the so-called CENP-A targeting domain (CATD)[48-50]. It comprises loop 1 and helix α2, in the histone fold domain. The CATD contains the cis-acting determinants for CENP-A centromeric deposition. In fact, when grafted onto H3, the CATD is sufficient to restrict the localization of the $H3^{CATD}$ chimera to centromeres and to complement the lethality of CENP-A depletion[50,51]. Nevertheless, the CATD is not sufficient to assemble a fully functional kinetochore and determine a long-term rescue[52]. The incorporation of CENP-A specifically in the centromeric chromatin and in a cell cycle-dependent manner requires HJURP (Holliday junction recognition protein), a CENP-A-specific histone chaperone[53]. The CATD is crucial for CENP-A binding to HJURP. The crystal structure of HJURP bound to a CENP-A / H4 heterodimer has provided high-resolution insights into this interaction[54]. Also, the crystal structure of the corresponding budding yeast complex, consisting of Scm3 bound to a CENP-A / H4 heterodimer, has revealed a substantial conservation, hinting at a common mechanism by which CENP-A is recognized by its specific chaperons and deposited at centromeres[55]. In addition, the CCAN protein CENP-N has been reported to directly bind to CENP-A nucleosomes, but not H3 nucleosomes[56]. The CATD appears to mediate this interaction because the $H3^{CATD}$ chimera is sufficient for CENP-N binding[57].

A consistent feature of CENP-A is the presence of a 2 - 6 amino acid insertion in loop 1 of the

histone fold domain, which is part of the CATD. The crystal structure of human CENP-A nucleosome[44] shows that CENP-A loop 1 protrudes from the nucleosome and that the 2 amino acid insertion (Arg 80 and Gly 81) is located at the tip of the loop, in a solvent-accessible region. Experiments of deletion and mutation of these residues indicate that they are critical for stable CENP-A retention at centromeres and suggest that they may represent a binding site for trans-acting factors.

Despite the aforementioned high divergence of the CENP-A N-terminal tail in terms of length and sequence, an interesting conserved feature is a significant enrichment in arginines with respect to histone H3[41]. This region is not required for centromere targeting of CENP-A[49,58], while it was shown to be involved in proper loading of the DNA binding protein CENP-B[52]. Also, recent investigations on post-translational modifications of the CENP-A N-terminal tail have proposed a role for methylated Arg 37 in *S. cerevisiae* in the recruitment of kinetochore components[59].

The CCAN protein CENP-C binds directly and specifically to CENP-A nucleosomes[57]. Recent high-resolution structural studies employing both nuclear magnetic resonance (NMR) and crystallography have defined the molecular determinants of this interaction[43]. In particular, two unstructured regions of CENP-C, the so-called central region (residues 426 - 537 of human CENP-C) and the CENP-C motif (residues 736 - 758 of human CENP-C), display the ability to bind to CENP-A nucleosomes. They both employ the same recognition mode, which involves docking onto the acidic patch of histone H2A and H2B through electrostatic interactions and binding to the CENP-A C-terminal tail through hydrophobic interactions. The central region of CENP-C is conserved in most mammals and the CENP-C motif is conserved from budding yeast to human. On the contrary, the C-terminal tail of CENP-A is divergent. Nevertheless, it retains a common property across evolution: it displays a remarkably higher hydrophobicity than the C-terminal tail of H3. In essence, it has been proposed that the interaction of CENP-C with CENP-A nucleosomes occurs via a conserved mechanism, in which disordered CENP-C peptides (the central region and the CENP-C motif) bind to the CENP-A C-terminal tail through hydrophobic interactions facilitated by an electrostatic docking onto the nucleosome core.

1.5 Debated nature of CENP-A nucleosomes

During DNA replication in S phase, parental H3 nucleosomes are distributed randomly to daughter strands and new H3 nucleosomes are deposited on both strands, enabling the propagation of parental chromatin structures to both daughter cells[60]. Studies in human cells have shown that CENP-A nucleosomes are also equally partitioned to sister centromeres during S phase, but the incorporation of new CENP-A nucleosomes is uncoupled from DNA replication[61] and, specifically, it occurs in late mitosis (telophase) and subsequent G1 phase[62]. Consistent with this, the deposition of new CENP-A nucleosomes takes place during anaphase in *Drosophila* syncytial embryonic nuclear divisions, which cycle between S and M phases without G1 or G2 phases[63].

The incorporation of new CENP-A nucleosomes during late mitosis and G1 phase has a significant impact on our understanding of centromeric chromatin[30]. First, it implies that the amount of CENP-A at centromeres is halved after DNA replication and through G2 and M phases. This raises questions about the composition of centromeric chromatin during this period, which includes the time of kinetochore assembly and function. One possibility is that the CENP-A gaps generated during DNA replication and nucleosome segregation in S phase are initially filled with, for example, H3 nucleosomes. At the end of mitosis and in G1 phase such H3 nucleosomes might be again replaced with CENP-A nucleosomes. Another option is that the CENP-A gaps remain nucleosome-free. Alternatively, CENP-A nucleosomes might be split during replication, forming half-nucleosomes that are reconstituted as octamers during late mitosis and G1 phase.

In vitro reconstitution of both human and yeast CENP-A nucleosomes yields a standard assembly, composed by a histone octamer containing two copies of CENP-A, H2A, H2B and H4 histones, with DNA wrapped in a left-handed orientation around the octamer[44,64].

Canonical octameric nucleosome has been described as the predominant form of CENP-A nucleosome in budding yeast[64], *Drosophila*[65] and human[66,67] centromeric chromatin *in vivo*.

Despite strong evidence in support of octameric CENP-A nucleosomes, tetrameric nucleosomes containing only one copy of CENP-A, H2A, H2B and H4 histones have been proposed to exist in *Drosophila*[68] and human cells[69]. The existence of a tetrameric nucleosome structure was essentially derived from three observations: failure to detect species consistent with an octameric histone arrangement after cross-linking treatment; protection of only approximately 120 bp of DNA from nuclease digestion; evidence from atomic force microscopy for a structure half the height of

typical nucleosomes. It has been recently suggested[70] that the failure to cross-link an octamer could be simply due to the fact that the cross-linker which was used specifically reacts with primary amines, which are absent from the region that is predicted to hold together the two halves of a putative *Drosophila* octameric CENP-A nucleosome. Moreover, as already mentioned, in the crystal structure of an *in vitro* reconstituted human CENP-A nucleosome[44], containing an octameric histone assembly, only the central 121 bp of the DNA are visible, suggesting that the DNA segments at the entrance and exit of the CENP-A nucleosome are flexible and more accessible to nuclease digestion than those of the canonical H3 nucleosome. Finally, it was recently shown that octameric CENP-A nucleosomes assembled *in vitro* exhibit reduced heights in atomic force microscopy measurements with respect to canonical octameric H3 nucleosomes, indicating that they possess distinct physical properties[71]. These considerations would eliminate the need to invoke the presence of tetrameric nucleosomes to explain the aforementioned observations. On the other hand, the possibility that CENP-A nucleosomes undergo cell cycle-dependent structural transitions *in vivo* has recently been proposed[72,73], which would go some way to reconcile the contrasting data reported in the literature about the structure and composition of CENP-A nucleosomes.

Moreover, the ability of Scm3, a CENP-A-specific histone chaperone in budding yeast, to displace H2A / H2B dimers from CENP-A-containing histone octamers *in vitro* led to the hypothesis that the centromeric nucleosome could possess an hexameric core containing two copies of CENP-A, H4 and Scm3[74]. As already mentioned, the crystal structure of Scm3 bound to a CENP-A / H4 heterodimer[55], together with the crystal structure of the corresponding human complex comprising HJURP bound to a CENP-A / H4 heterodimer[54], provided high-resolution insights into this interaction. Specifically, structural data would exclude the hexameric model, as Scm3 binding to CENP-A / H4 heterodimer yields a trimeric complex that is incompatible with CENP-A / H4 tetramerization and with DNA binding to the histones. The same holds true for the HJURP / CENP-A / H4 trimeric complex. This in turn suggests that the various chaperone-bound species described *in vivo*, rather than representing stable nucleosome structures, most likely depict intermediates in the nucleosome assembly process[67].

1.6 The constitutive centromere-associated network (CCAN)

As discussed in the previous sections, CENP-A has been defined as the primary epigenetic determinant of centromeric chromatin and, as such, it possesses the ability to specify centromere identity across the cell cycle and to ensure its faithful inheritance through successive cell generations by self-templating the deposition of newly expressed CENP-A during each cell cycle.

Following CENP-A incorporation into chromatin, the information encoded by this mark must be recognized and translated into the generation of a functional centromere. In this respect, a crucial role is performed by the so-called constitutive centromere-associated network (CCAN).

The centromere-associated protein complex was first described about ten years ago by a study which included the isolation of CENP-A chromatin from HeLa interphase nuclei through chromatin immunoprecipitation using a monoclonal antibody against CENP-A, followed by a systematic analysis of its components by mass spectrometry. This led to the identification of approximately 40 proteins, including previously reported centromere proteins as well as proteins of unknown function[16]. Stemming from this finding, a few years later, three almost simultaneous studies employed different approaches in human and chicken cells to isolate additional proteins that associate with centromeres throughout the cell cycle[17,18,75]. At the same time, bioinformatic analyses identified potential orthologs of many CCAN components across evolution, suggesting that this protein complex is conserved[24].

Studies of localization dependencies show that all CCAN proteins are recruited downstream of CENP-A deposition at centromeres[17,76,77]. This poses CENP-A at the top of the centromere and kinetochore assembly hierarchy. On the other hand, there are several studies that suggest the involvement of at least some CCAN components in the establishment of CENP-A chromatin[18,78,79].

Biochemical and functional analyses indicate that, rather than forming a single stable complex, CCAN proteins are organized in sub-complexes (Figure 4). These include: CENP-C, CENP-L / CENP-N group, CENP-H / CENP-I / CENP-K group, CENP-M, CENP-T / CENP-W / CENP-S / CENP-X group, CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group. A brief overview of each of them is provided below.

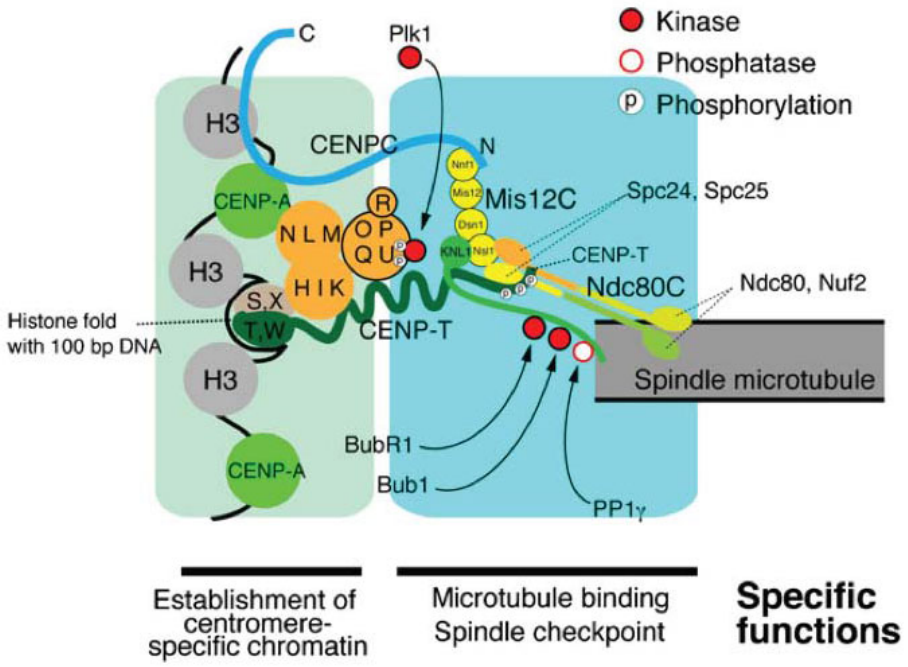


Figure 4 - Model of the molecular organization of kinetochores in metazoans.

Centromeric chromatin, comprising both CENP-A and H3 nucleosomes, lies at the basis of kinetochore. CENP-A nucleosomes function as epigenetic marks for centromere specification. CCAN components are involved in recognizing the information encoded by this mark and generating a functional centromere. CCAN proteins are organized in sub-complexes. The CCAN, in the inner kinetochore (light green), contacts the outer kinetochore (light blue), which comprises components of the KMN network, that contributes a microtubule-binding interface, and of the spindle assembly checkpoint. Adapted from Hori T. and Fukagawa T., *Chromosome Res*, 2012[80].

Many CCAN proteins are not essential in yeast. Moreover, some organisms, like *Caenorhabditis elegans* and *Drosophila melanogaster*, appear to have lost most CCAN subunits and to rely solely on CENP-A and CENP-C as inner kinetochore components. This suggests that these organisms might assemble simpler versions of kinetochores that retain the basic chromosome segregation function, whereas robustness and high-fidelity are achieved in higher eukaryotes through the assembly of a more complex and redundant machinery[22].

As already mentioned, the CCAN comprises proteins that associate with the centromere throughout the cell cycle, as opposed to kinetochore components that are only transiently recruited during mitosis, such as the microtubule-binding KMN network. Recent evidence supports a more dynamic view of the CCAN as a group of proteins that are regulated in a cell cycle-dependent manner and respond to the mechano-structural state of the kinetochore. For example, CENP-I and a fraction of CENP-H are stably bound to the centromere during the whole cell cycle[81]. CENP-U is instead absent from the kinetochore in anaphase because it is targeted for degradation in a polo-like

kinase 1 (Plk1)-dependent manner[82]. Moreover, kinetochore levels of CENP-O are reduced by 40 % in metaphase compared to interphase[83]. Also, CENP-N is bound to kinetochores during S phase and G2, but largely absent during mitosis and G1[84]. The dynamic behaviour of some CCAN subunits adds a new layer of complexity and definitely requires further investigation to get a better understanding of kinetochore biology.

1.6.1 CENP-C

CENP-C was the first protein discovered to be part of the kinetochore by immunoelectron microscopy[85] and then shown by antibody microinjection to be required for kinetochore assembly[86]. CENP-C orthologs are found widely across evolution[87-90], including *S. cerevisiae* (where it is named Mif2), *S. pombe* (where it is named Cnp3), *D. melanogaster* and *C. elegans* (where it is named HCP4).

Depletion studies in different organisms have located CENP-C quite high in the centromere and kinetochore assembly hierarchy. Indeed, it is required for the recruitment of various CCAN proteins (such as CENP-K[91] and CENP-L[92]) and also of components of the outer kinetochore (such as Mis12 complex[76,91,93] and Ndc80 complex[76,93]) and of the spindle assembly checkpoint (such as Mad1[76], Mad2[76,91,93], Bub1[76], BubR1[76,93]). CENP-C recruitment seems to be mainly dependent on CENP-A[40]. A role of the CENP-H / CENP-I / CENP-K group in CENP-C localization has also been suggested[93] (Table 1).

Direct interacting partners of CENP-C described in the literature so far are Mis12 complex, CENP-A nucleosomes and CENP-C itself. In particular, CENP-C is predicted to be mostly unstructured and some relevant known features of this protein are the N-terminal region (residues 1 - 21), which is sufficient for direct binding to Mis12 complex[94,95]; the central domain (residues 422 - 537) and the so-called CENP-C motif (residues 736 - 758), which are involved in direct binding to CENP-A nucleosomes[43,57], as described in a previous section of this dissertation; and finally the C-terminal region (residues 890 - 943), which is responsible for protein dimerization[96] and comprises a cupin domain, whose crystal structure has been determined for the budding yeast ortholog of CENP-C[97] (Figure 5).

Recent evidence points to CENP-C and CENP-T as major players in kinetochore assembly, connecting inner and outer kinetochore. In particular, experiments of ectopic targeting of these proteins in human and chicken cells have shown their ability to generate kinetochore-like *foci* that

can recapitulate various functions of normal kinetochores, such as the recruitment of various inner and outer kinetochore components, the presence of microtubule attachments and the microtubule-sensitive recruitment of spindle assembly checkpoint proteins[79,98].

1.6.2 CENP-L / CENP-N group

CENP-N and CENP-L were first identified as members of the CENP-A nucleosome associated complex (NAC) and CENP-A nucleosome distal (CAD) complex, respectively[17].

Orthologs of these proteins have been found in yeast[24,25]. In particular, CENP-L is called Iml3 or Mcm19 in *S. cerevisiae* and Fta1 in *S. pombe*, while CENP-N is called Chl4 in *S. cerevisiae* and Mis15 in *S. pombe.*

As already mentioned, CENP-N has been reported to directly bind *in vitro* to CENP-A nucleosomes, but not H3 nucleosomes[56]. CENP-N and CENP-C recognize different regions of the CENP-A nucleosome and their binding is not competitive[57].

The C-terminus of CENP-N, which is not required for the interaction with CENP-A nucleosomes, directly binds to CENP-L[56]. A recent study has determined the crystal structures of an Iml3 (CENP-L) homodimer and of a heterodimer of Iml3 (CENP-L) and the C-terminal region of Chl4 (CENP-N) from *S. cerevisiae*[99]. The structure of Iml3 displays a central β-sheet with ten β-strands that wrap around two α-helices, while four α-helices decorate the outside of this core. In the Iml3 homodimer, the β-sheets of two Iml3 molecules associate to form a continuous β-sheet comprising twenty β-strands. The C-terminal Chl4 construct includes four β-strands and two α-helices. The Iml3 / Chl4 heterodimer (Figure 5) relies on β-augmentation at the same surface that mediates Iml3 homodimerization; homodimer and heterodimer formations are therefore mutually exclusive.

Moreover, the same study has shown that the budding yeast Iml3 (CENP-L) / Chl4 (CENP-N) complex directly binds to Mif2 (CENP-C). This interaction requires the central region of Mif2 and the N-terminal domain of Chl4, which is sufficient to support Mif2 binding[99]. In addition, Fta1, the fission yeast ortholog of CENP-L, has been suggested to directly interact with Cnp3, the fission yeast ortholog of CENP-C[92]. If conserved, this molecular link between CENP-L / CENP-N complex and CENP-C would provide a possible explanation to the fact that CENP-C localization is reduced upon CENP-N depletion in HeLa cells[56] and the observation that CENP-N is found at kinetochore-like *foci* specified by the ectopic targeting of CENP-C and CENP-T, in the absence of CENP-A[98].

The CENP-L / CENP-N group is also closely related to the CENP-H / CENP-I / CENP-K group and

to CENP-M. Indeed, they are often found associated in immunoprecipitations against one of them[17,18] and show mutual localization dependencies[17,18,56,100] (Table 1). A better understanding of the direct interactions among these proteins is crucial to shed light on the molecular mechanisms that underlie such a complex network of relationships.

1.6.3 CENP-H / CENP-I / CENP-K group

CENP-H was initially identified as a constitutive centromeric protein because it was observed to colocalize with known centromeric components, such as CENP-A and CENP-C, throughout the cell cycle[101,102]. In the same studies, it was predicted to be mainly composed of coiled-coils. Orthologs of CENP-H have been found in yeast[24]. In particular, CENP-H is called Mcm16 in *S. cerevisiae* and Fta3 in *S. pombe*.

CENP-I was first described as a constitutive component of the centromere because of its ability to colocalize with CENP-A, CENP-C and CENP-H throughout the cell cycle[103]. In the same study, yeast orthologs of CENP-I were identified in the *S. cerevisiae* protein Ctf3 and the *S. pombe* protein Mis6.

CENP-K was initially found to be a member of the CENP-A nucleosome distal (CAD) complex[17]. Like CENP-H, it is predicted to be enriched in coiled-coils. Yeast orthologs of CENP-K have been identified in the *S. cerevisiae* protein Mcm22 and the *S. pombe* protein Sim4[25].

CENP-H, CENP-I and CENP-K were soon defined as a CCAN sub-group because they are usually found closely associated in immunoprecipitation experiments and because they share a similar pattern of localization dependencies[18] (Table 1). In particular, their kinetochore localization is downstream of CENP-A. In turn, although their depletion does not affect CENP-A molecules already present at the centromere, they seem to have a role in the incorporation of newly synthesized CENP-A[18]. Also, as already mentioned, the CENP-H / CENP-I / CENP-K group, the CENP-L / CENP-N group and CENP-M display mutual localization dependencies[17,18,56,100]. Contrasting data can be found in the literature regarding the role of CENP-T[77] and CENP-C[57,91] in the kinetochore recruitment of the CENP-H / CENP-I / CENP-K group. In turn, a role of the CENP-H / CENP-I / CENP-K group in CENP-C localization has also been suggested[93]. Moreover, the centromeric localization of the CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group of proteins was shown to be dependent on the CENP-H / CENP-I / CENP-K group, but not the opposite[18,100,104]. In addition, the CENP-H / CENP-I / CENP-K group of proteins contributes to the

recruitment of outer kinetochore components[18,93,104,105].

The only direct interaction described so far involving these proteins is the formation of a 1 : 1 complex between CENP-H and CENP-K[106].

1.6.4 CENP-M

CENP-M was initially discovered as proliferation-associated nuclear element 1 (PANE1), highly expressed in proliferating cells such as activated lymphoid cells and tumours[107,108]. A few years later, CENP-M was identified as inner kinetochore component in three studies, which, in particular, defined it as interphase centromere complex (ICEN) component ICEN39[75], as well as a member of the CENP-A nucleosome associated complex (NAC)[17], and finally as part of the CENP-H / CENP-I-associated group of proteins[18].

No CENP-M orthologs have been identified so far in yeast[25], *Drosophila* and *C. elegans*.

In terms of localization dependencies, CENP-M influences the kinetochore recruitment of the CENP-H / CENP-I / CENP-K group[17,18,75], the CENP-L / CENP-N group[17,18] and the CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group[18]. In turn, CENP-M localization has been reported to depend on all of them[17,18] and on CENP-T[17] (Table 1).

No direct binding partners of CENP-M have been described to date. It is worth pointing out once more that only a thorough mapping of the direct interactions involving CENP-M and the aforementioned groups of proteins will allow an understanding of the intricate localization dependencies that have been observed.

1.6.5 CENP-T / CENP-W / CENP-S / CENP-X group

CENP-T was originally identified as a component of the CENP-A nucleosome associated complex (NAC)[17]. Subsequently, CENP-W was discovered as a CENP-T-associated CCAN protein[77]. CENP-S was initially described as a member of the CENP-A nucleosome distal (CAD) complex[17]. CENP-X was afterwards identified as a CENP-S-associated CCAN component[109]. CENP-S and CENP-X, also named MHF1 and MHF2, run a parallel life as part of the Fanconi Anaemia nuclear core complex, which is involved in DNA repair[110]. The relation, if any, between this role of CENP-S and CENP-X and their function in the context of the kinetochore is currently unknown.

Orthologs of these proteins have been identified in yeast[25]. In particular, CENP-T is Cnn1 in *S.*

cerevisiae and SPBC800.13/Cnp20 in *S. pombe*, CENP-W is YDR374W-A in *S. cerevisiae* and SPAC17G8.15 in *S. pombe*, CENP-S is YOL086W-A in *S. cerevisiae* and SPBC2D10.16 in *S. pombe* and CENP-X is YDL160C-A in *S. cerevisiae* and SPCC576.12c in *S. pombe*.

CENP-T possesses a long N-terminal tail and a histone fold domain at its C-terminus[77]. CENP-W, CENP-S and CENP-X are small proteins entirely represented by a histone fold domain[77,111]. The presence of a histone fold domain in non-histone proteins is not unprecedented: some examples include members of the TFIID complex[112], a major interacting partner of RNA polymerase II, and transcription factors, like negative cofactor 2 (NC2a/NC2b)[113]. CENP-T and CENP-W associate into a 1 : 1 dimer, while CENP-S and CENP-X bind as a 2 : 2 tetramer, but, when incubated in equimolar amounts with CENP-T and CENP-W, they form a 1 : 1 : 1 : 1 CENP-T / CENP-W / CENP-S / CENP-X tetramer. The crystal structure of these three complexes has recently been determined[111] (Figure 5).

Like canonical histones, these proteins have retained the ability to interact with DNA and CENP-T / CENP-W / CENP-S / CENP-X complex was shown to protect approximately 100 bp of DNA from MNase digestion *in vitro*. Stemming from this observation, it was proposed that this complex forms a novel nucleosome-like structure at the centromere[111]. Additional biochemical and structural characterization of the association of CENP-T / CENP-W / CENP-S / CENP-X complex with DNA, as well as an assessment of its interactions with centromeric components, both *in vitro* and *in vivo*, are required to clarify this point further. This is especially important in light of a previous report showing that CENP-T / CENP-W complex associates with H3 nucleosomes, but not with CENP-A nucleosomes, in centromeric regions[77].

The formation of the CENP-T / CENP-W heterodimer through their histone fold domains is required for their kinetochore recruitment both in human and yeast[25,111]. Also, CENP-T possesses two α-helices that extend beyond the histone fold domain and are positioned away from the predicted DNA-interacting interface. The sequence of these two α-helices is highly conserved, suggesting a functional relevance of this region. Indeed, deletion of these helices abolishes centromere localization of CENP-T[111]. The kinetochore localization of CENP-S and CENP-X is interdependent and also requires CENP-T / CENP-W[109]. On the contrary, the formation of the CENP-T / CENP-W / CENP-S / CENP-X heterotetramer is not strictly required for the recruitment of CENP-T, as it is present at kinetochores in the absence of CENP-S. However, the observation that CENP-S- and CENP-X-deficient cells display reduced Ndc80 levels at kinetochores and a perturbed outer plate

structure suggests that CENP-T is not fully functional in these cells[109]. In the kinetochore assembly hierarchy, it seems that CENP-T / CENP-W is upstream[77], while CENP-S / CENP-X is downstream[109] of the CENP-H / CENP-I / CENP-K group. Centromeric targeting of CENP-S / CENP-X and CENP-C appear to be independent, as CENP-S / CENP-X localization is not affected in the absence of CENP-C[77] and vice versa[109]. The relationship between CENP-T / CENP-W and CENP-C is instead a matter of debate. On one hand, knock-out analyses in chicken DT40 cells showed that CENP-T / CENP-W and CENP-C localize independently to centromeres[77]; on the other hand, CENP-T recruitment was reported to occur downstream of CENP-C based on RNAi experiments in human cells[57]. A summary of localization dependencies of CCAN components is presented in table 1.

Like CENP-C, CENP-T also functions as a bridge between the centromere and the outer kinetochore. In fact, while the C-terminal histone fold domain of CENP-T is embedded in the centromeric chromatin, the N-terminal region of the protein directly binds to the Ndc80 complex. This interaction is dependent on the phosphorylation of CENP-T by cyclin-dependent kinase 1 (CDK1)[98,114]. Also in budding yeast, the CENP-T ortholog Cnn1 was shown to be a centromere receptor for the Ndc80 complex[25,115].

1.6.6 CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group

These proteins were first identified as part of the CENP-A nucleosome associated complex (NAC) and CENP-A nucleosome distal (CAD) complex[17]. CENP-U and CENP-O were also defined as ICEN (Interphase Centromere Complex) components, with the names of ICEN24 and ICEN36, respectively[75].

Yeast orthologs have been found for all of them, except for CENP-R[24,25,100]. In particular, CENP-O is Mcm21 in *S. cerevisiae* and Mal2 in *S. pombe*, CENP-P is Ctf19 in *S. cerevisiae* and Fta2 in *S. pombe*, CENP-Q is Okp1 in *S. cerevisiae* and Fta7 in *S. pombe*, CENP-U is Ame1 in *S. cerevisiae* and Mis17 in *S. pombe.*

Coexpression in bacteria followed by an affinity purification step and a size-exclusion chromatography showed that CENP-O, CENP-P, CENP-Q and CENP-U form a complex and can associate with CENP-R[116]. Recently, the crystal structure of a sub-complex comprising a Mcm21 (CENP-O) / Ctf19 (CENP-P) heterodimer from *Kluyveromyces lactis* has revealed that these two proteins possess similar structures, as they both contain double-RWD domains, which together

form a Y-shaped assembly with flexible N-terminal extensions[117] (Figure 5). The RWD domain was named after three major RWD-containing proteins: RING finger-containing proteins, WD-repeat-containing proteins, and yeast DEAD (DEXD)-like helicases. The first high-resolution structure determination of an RWD domain revealed that it consists of a $\alpha + \beta$ sandwich fold, comprising a four-stranded antiparallel β-sheet flanked by three α-helices, with a $\alpha\beta\beta\beta\beta\alpha\alpha$ topology[118]. The RWD domain has been so far observed in six kinetochore proteins, namely Mcm21 (CENP-O), Ctf19 (CENP-P), Csm1 (a subunit of the *S. cerevisiae* monopolin complex)[119], Spc24, Spc25[14,120], and Mad1[121], suggesting that this fold is a recurring feature of kinetochore molecular architecture. CENP-Q, CENP-U and CENP-R are predicted to be mainly composed by α-helices forming coiled-coils. In CENP-U, this region seems to be preceded by an N-terminal disordered tail[22].

Knock-outs of each of these proteins in DT40 cells are all viable and show mild mitotic defects, including a slower proliferation and a less organized metaphase plate with respect to control cells. CENP-O, CENP-P, CENP-Q and CENP-U are interdependent for their centromere localization, while CENP-R appears to be located downstream in the kinetochore assembly hierarchy[18,116]. Also, their recruitment requires the CENP-H / CENP-I / CENP-K group, while the contrary is not true[18,100,104], and shows interdependency with CENP-M[18] (Table 1).

A direct interaction between CENP-U and Ndc80 was recently described[122]. In addition, CENP-U was observed to bind to microtubules directly and also to display cooperative microtubule binding with Ndc80 *in vitro*[122]. Moreover, CENP-U phosphorylation by Aurora B was shown to negatively regulate CENP-U binding to Ndc80 *in vitro* and kinetochore-microtubule interactions *in vivo*, suggesting that this mechanism contributes to the error correction function of Aurora B[122]. CENP-Q was also reported to directly bind to microtubules *in vitro*[123].

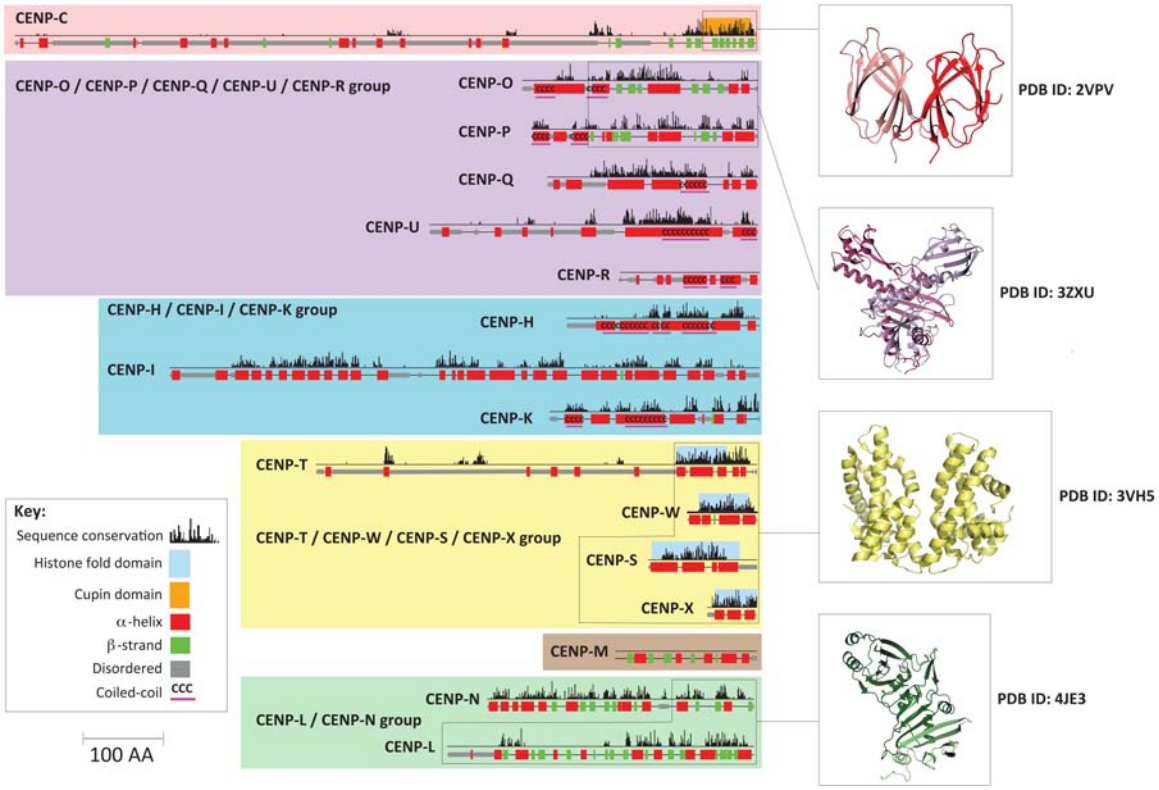


Figure 5 - Conservation and structural features of CCAN proteins.

Left panels: schematic representations of CCAN proteins indicate the distribution of predicted secondary structure elements (JPRED), disordered regions (DISOPRED) and coiled-coil domains (COILS). Conservation scores between human and yeast proteins, based on multiple sequence alignments, are plotted above the schemes. The cupin domain of CENP-C and the histone fold domains of CENP-T, CENP-W, CENP-S and CENP-X are highlighted. Right panels: cartoon representations and respective PDB IDs of crystal structures of CCAN proteins, which have been described in the main text. Adapted from Westermann S. and Schleiffer A., *Trends in Cell Biology*, 2013[22].

CCAN sub-groups	Centromeric recruitment depends on
CENP-C	CENP-L / CENP-N group[56] CENP-H / CENP-I / CENP-K group[93]
CENP-L / CENP-N group	CENP-C[92,98] CENP-H / CENP-I / CENP-K group[17,18] CENP-M[17,18]
CENP-H / CENP-I / CENP-K group	CENP-H / CENP-I / CENP-K are interdependent[18] CENP-C[57,91] CENP-L / CENP-N group[17,18] CENP-M[17,18] CENP-T[77]
CENP-M	CENP-L / CENP-N group[17,18] CENP-H / CENP-I / CENP-K group[17,18] CENP-T[17] CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group[18]
CENP-T / CENP-W / CENP-S / CENP-X group CENP-T / CENP-W CENP-S / CENP-X	 CENP-T and CENP-W are interdependent[25,111], while the interaction with CENP-S / CENP-X is not strictly required[109] CENP-C (debated[57,77]) CENP-S and CENP-X are interdependent and dependent on CENP-T / CENP-W[109] CENP-H / CENP-I / CENP-K group[109]
CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group	CENP-O / CENP-P / CENP-Q / CENP-U are interdependent, while CENP-R appears to be downstream[18,116] CENP-H / CENP-I / CENP-K group[18,100,104] CENP-M[18]

Table 1 - Summary of dependencies of CCAN components for their recruitment at centromeres.

For each CCAN sub-group (left column), a synthetic list of CCAN components that are involved in its centromeric recruitment is provided (right column). In addition, it has been shown that all CCAN proteins are recruited downstream of CENP-A deposition at centromeres[17,76,77]. A more detailed description of the localization dependencies of CCAN components is provided in the main text.

1.7 The Knl1 complex, Mis12 complex, Ndc80 complex (KMN) network

After the centromere has been specified by CENP-A and centromeric chromatin has been established by the CCAN, the kinetochore must build robust interactions with spindle microtubules to achieve accurate chromosome segregation during mitosis. In particular, the microtubule-binding interface is provided by the outer kinetochore plate, which is primarily composed of a highly conserved group of 10 kinetochore proteins referred to as the KMN network (Figure 6). This acronym stands for Knl1 complex, Mis12 complex and Ndc80 complex. KMN network components start to localize to kinetochores around the G2 phase and dissociate at the end of mitosis[11].

Ndc80 complex comprises four proteins, namely Ndc80, Nuf2, Spc24 and Spc25[19]. The crystal structure of an engineered version of the complex revealed that it has a dumbbell-like shape with a long coiled-coil region and globular domains at both ends[14]. The globular domains of Spc24 and Spc25 tightly interact with Mis12 complex[20], while those of Ndc80 and Nuf2 directly bind to microtubules[12]. Microtubule binding is mediated by electrostatic interactions and is negatively regulated by phosphorylation of the N-terminal region of Ndc80 by Aurora B, consistently with the role of this kinase in de-stabilizing improper kinetochore-microtubule attachments[124].

Mis12 complex also contains four proteins, specifically Nnf1, Mis12, Dsn1 and Nsl1. Recent biochemical and structural analyses showed that the four subunits of Mis12 complex arrange as a rod-shaped unit[20]. Mis12 complex has been defined as a protein interaction hub crucial for kinetochore assembly. In fact, a direct interaction between Mis12 complex and the N-terminal region of CENP-C has recently been described[94,95]. In addition, as already mentioned, it binds to Ndc80 complex[20]. Furthermore, it interacts, through the C-terminal tail of its Nsl1 subunit, with the C-terminal region of Knl1[20].

Knl1 is a large protein (~300 kDa in human), predicted to be mostly unstructured. The C-terminal region of Knl1 binds to Zwint[125] and to Mis12 complex[20], while a direct interaction between Knl1 and Ndc80 complex has not been described to date. Knl1 has also been shown to possess microtubule-binding activity *in vitro*[124,126]. In addition, the N-terminal region of Knl1 binds to protein phosphatase 1 (PP1). PP1 recruitment at the kinetochore is required to counteract Aurora B activity by de-phosphorylating its substrates and, consequently, to stabilize kinetochore-microtubule attachments[127]. Moreover, the N-terminus of Knl1 has also been shown to interact with the spindle assembly checkpoint kinases Bub1 and BubR1[125,128].

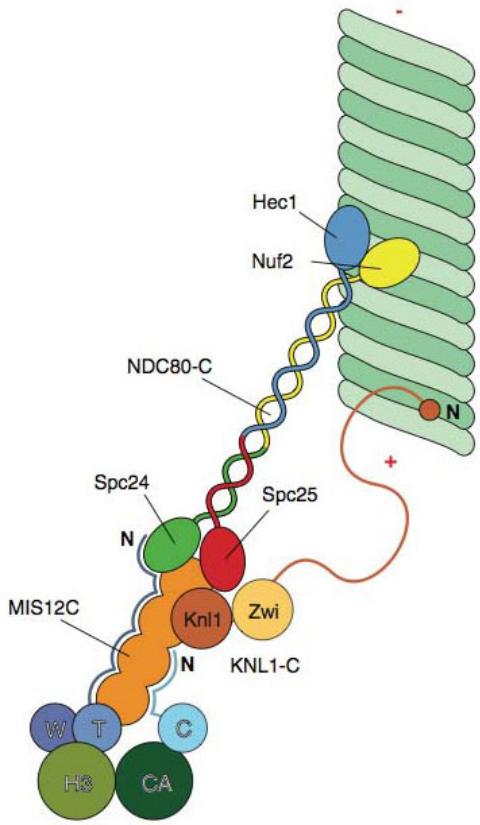


Figure 6 - Model of the molecular architecture of the KMN network.

With the exception of CENP-T / CENP-W complex (abbreviated as T and W) and CENP-C (indicated as C), all subunits of the CCAN have been omitted. CENP-T / CENP-W complex associates with H3 nucleosomes (indicated as H3), whereas CENP-C associates with CENP-A nucleosomes (abbreviated as CA). The N-terminal region of CENP-T is an extended, largely disordered polypeptide chain that makes contacts with Mis12 complex (MIS12-C) and Ndc80 complex (NDC80-C). The N-terminal region of CENP-C is probably also disordered and binds to MIS12-C. Knl1 complex (KNL1-C), which comprises Knl1 and Zwint (indicated as Zwi), might contain a microtubule-binding site in the N-terminal region of Knl1. The C-terminal region of Knl1 directly interacts with MIS12-C. The Spc24 and Spc25 subunits of NDC80-C interact with MIS12-C, whereas the Ndc80 (also called Hec1) and Nuf2 subunits bind to microtubules. Adapted from Deluca J. G. and Musacchio A., *Curr Opin Cell Biol*, 2012[7].

2 Results

2.1 Section 1 – CENP-M

2.1.1 Determination of CENP-M$^{1-171}$ crystal structure

Given that the structure of CENP-M was unknown, I sought to determine it by X-ray crystallography. Human CENP-M was produced by recombinant expression in *Escherichia coli*. Affinity purification through a cleavable GST-tag and additional chromatographic steps led to the purification of an apparently monodisperse sample (Figure 7 A). Unfortunately, all my attempts to crystallize this full-length construct were unsuccessful. According to the results of a limited proteolysis experiment, I then generated a construct lacking nine amino acids at the C-terminus of the protein (CENP-M$^{1-171}$) for crystallographic purposes. As CENP-M full-length, also CENP-M$^{1-171}$ was expressed in bacteria and purified to homogeneity (Figure 7 B). I was able to obtain and optimize crystals of this protein construct. In particular, CENP-M$^{1-171}$ (~10 mg/ml) was crystallized by sitting drop vapour diffusion. Numerous commercially available crystallization screens were tested, but protein crystals were obtained only in one condition (JCSG+ suite condition F6: 0.1 M Bicine pH 9.0, 10 % MPD) (Molecular Dimensions). Diffraction-quality crystals were produced by optimizing the initial condition in hanging drops. Typically, crystals formed within 48 h and reached full-size in about one week. Most of them grew as bundles, tangled up with precipitated protein, and only a minority of single rod-shaped crystals were present. Single crystals were transferred to a cryobuffer containing the reservoir liquor supplemented with 15 % Glycerol and were flash-frozen in liquid nitrogen. A selenomethionine (SeMet)-derivative of the protein was crystallized under similar conditions. X-ray diffraction data were collected with synchrotron radiation yielding high-resolution datasets. The structure of CENP-M$^{1-171}$ was solved by SAD (Single-wavelength Anomalous Diffraction) phasing and then refined at 1.5 Å resolution. It represented a particularly challenging case due to a problem of merohedral twinning affecting all analysed crystals. For this reason, it is worth spending a few words on how I succeeded in determining CENP-M$^{1-171}$ structure.

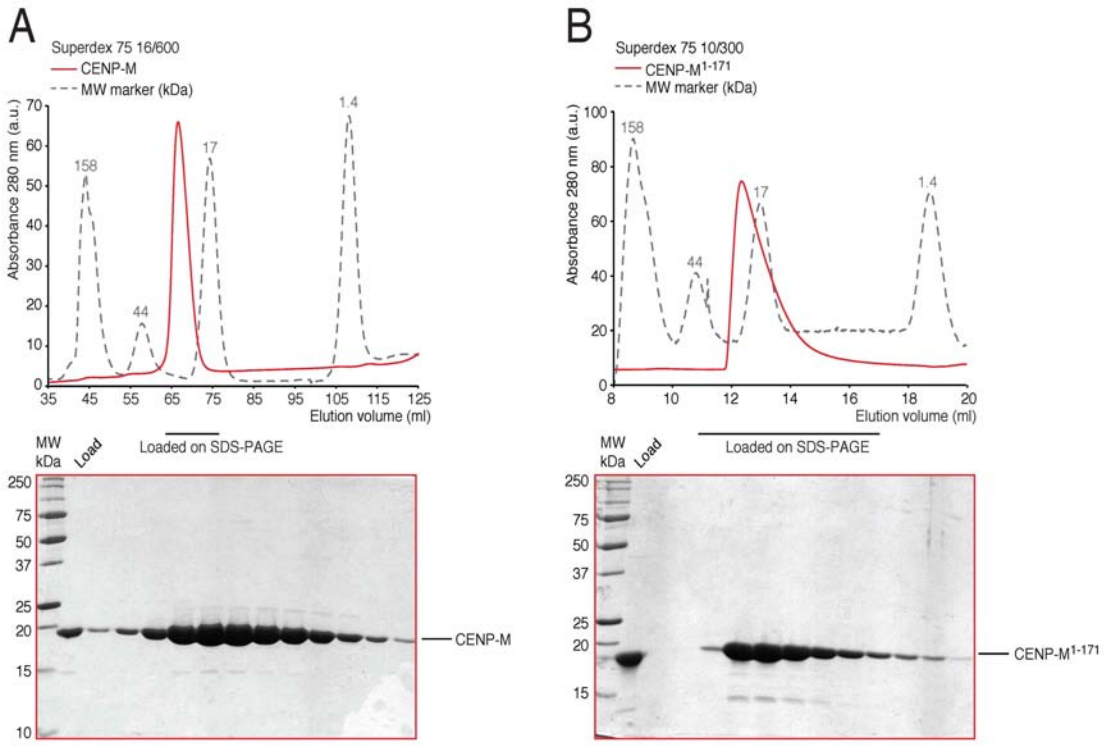


Figure 7 - Final step of CENP-M and CENP-M$^{1-171}$ purifications.

Size-exclusion chromatography (SEC) elution profiles from Superdex 75 columns and SDS-PAGE analyses of purified recombinant CENP-M (A) and CENP-M$^{1-171}$ (B). In both cases, the protein migrates as a monomer and a good level of purity (> 95 %) has been achieved. Bands visible on gels below purified CENP-M or CENP-M$^{1-171}$ represent degradation forms.

All analysed crystals displayed an apparent hexagonal crystal system but an evaluation of data quality and crystal defects using phenix.xtriage[129] revealed that they were all affected by twinning, a pathological situation often hampering crystal structure solution. A conventional non-twinned crystal consists of the repetition of the same unit cell, with the same orientation, in the three dimensions. A twinned crystal instead consists of the repetition of the same unit cell, with different orientations, in the three dimensions. Each domain of a twinned crystal gives rise to its own diffraction pattern and what is measured is a superposition of the various patterns, with diffraction intensities weighted according to the fraction of the crystal contributing to each domain[130]. In particular, all CENP-M$^{1-171}$ crystals displayed the so-called merohedral twinning, implying that reflections from different twin domains perfectly overlap[131]. In order to deconvolute the diffraction pattern of a twinned crystal, two parameters need to be taken into account: the twin law, which is the symmetry operator relating twin domains, and the twin fraction, which is the fraction of the crystal contributing to each domain. Concerning the twin law, in our case we needed to discern, from the apparent six-fold symmetry of crystals, if this was due to a crystallographic two-fold axis plus a three-fold axis as twin law or the

opposite.

In order to tackle the problem, I first tried to avoid or at least reduce twinning occurrence during the crystallogenesis phase. A number of crystallization optimization attempts were performed with this aim. In particular, I tested different crystallization temperatures, vapour diffusion rate control with oils, additive screens[132], lysine methylation[133], addition of the starting crystallization condition to screens, microseeding and in-screen microseeding[134]. Unfortunately, no improvements were obtained concerning twinning.

Given that I did not succeed in eliminating the problem, I then sought to solve the structure of CENP-M$^{1-171}$ using twinned datasets and try to overcome the obstacle from a computational point of view. The main challenge I had to face was phasing. One possibility is to detwin the diffraction data and then use such detwinned data for structure solution. However, the extent to which this deconvolution procedure is advantageous is matter of debate[131], as it may introduce significant errors. Moreover, potential improvements significantly depend on the accuracy of the originally measured intensities as well as on the correctness of the choice of the twin law and on the accuracy of the estimation of the twin fraction. In my case, this approach was not beneficial.

Phasing by molecular replacement was attempted, employing a model of human CENP-M structure predicted by the server Phyre2[135] with high confidence. Nevertheless, the model did not yield any phasing solution.

As I was also able to crystallize a SeMet-derivative of the protein, I tested a number of these crystals, until I could collect a good SAD dataset (Table 2). Specifically, it displayed high completeness, redundancy and I/sigma (signal to noise ratio), low Rsym (indicating a good agreement among the independent measurements of symmetry-related reflections in a crystallographic dataset, where symmetry-related reflections should have identical intensities) and lower estimated twin fraction with respect to the majority of the datasets that I had previously obtained. On this dataset I tried several indexing possibilities. In particular:

- P6, the apparent spacegroup, given the apparent six-fold symmetry.
- P2, a plausible real spacegroup. This would imply an interpretation of the apparent six-fold symmetry as determined by a crystallographic two-fold axis plus a three-fold axis as twin law.
- P3, another plausible real spacegroup. This would imply an interpretation of the apparent six-fold symmetry as determined by a crystallographic three-fold axis plus a two-fold axis

as twin law.

P3 proved to be the correct spacegroup. Indeed, it resulted in the identification of heavy atom sites. The subsequent steps of phasing and density modification produced an interpretable electron density map. Model building was then carried out combining automatic building of protein fragments by various programs and manual building. This preliminary model was then used for molecular replacement with a native dataset at higher-resolution (1.5 Å). Iterative model building and refinement yielded a final model covering the full asymmetric unit. During refinement, model parameters are modified in order to describe the experimental data as accurately as possible. In order to do so, twinning has to be taken into account. Most of the currently available refinement programs, such as phenix.refine[129], are able to deal with twinning and refine model parameters, including the twin fraction, once provided with the appropriate twin law. Spacegroup P3 can give rise to twinning through three possible symmetry operations: (-h, -k, l), (k, h, -l), (-k, -h, -l). In my case, I reasoned that the twin law was (-h, -k, l), because it represents a two-fold axis parallel to the three-fold axis of P3 and thus results in an apparent six-fold symmetry, precisely the situation observed with my crystals. Moreover, programs for analysis of data quality and crystal defects, such as phenix.xtriage[129], suggested that (-h, -k, l) was the most likely twin law, while the other two possible ones for P3 were estimated as negligible. The refinement process resulted in a final model displaying good statistics (Table 2), suggesting that twinning was properly handled.

Data collection	Native	Derivative
Beamline	ESRF ID14-4	SLS X06DA (PXIII)
Spacegroup	P3	P3
Unit cell parameters (Å)	a = b = 104.50, c = 33.59	a = b = 104.03, c = 33.56
(°)	α = β = 90, γ = 120	α = β = 90, γ = 120
Wavelength (Å)	0.91970	0.97942
Resolution limits (Å)	52.25 – 1.49 (1.54 – 1.49)*	31.45 – 2.00 (2.06 – 2.00)*
Reflections observed / unique	607786 / 64606	419456 / 27271
Completeness (%)	98.3 (96.3)*	99.9 (98.9)*
R_{sym} § (%)	5.6 (36.1)*	9.4 (75.2)*
<I> / <σI>	26.3 (6.4)*	24.1 (3.9)*
Redundancy	9.4 (8.8)*	15.4 (13.9)*
SAD phasing		
BAYES-CC		49.1 + / - 18.5
Se sites found / expected		5 / 6
FOM before solvent flattening and density modification		0.35
FOM after solvent flattening and density modification		0.69
Refinement		
Resolution limits (Å)	52.25 – 1.49 (1.52 – 1.49)*	
Reflections for R_{cryst} / for R_{free}	59806 / 4800	
R_{cryst} ‡ (%)	12.4 (20.6)*	
R_{free} ‡ (%)	16.4 (23.0)*	
No. of protein atoms / water atoms	2307 / 297	
Average B factor protein atoms / water atoms	21.67 / 35.34	
RMSD bond lengths (Å)	0.005	
RMSD bond angles (°)	0.854	
Twin fraction (operator -h, -k, l)	0.49	
Ramachandran Plot Statistics#		
Favoured region (%)	99.3	
Outliers (%)	0.0	

BAYES-CC: Bayesian estimate of the correlation coefficient (CC) between the experimental map and an ideal map, reported as CC * 100 + / - 2 standard deviations

FOM: figure of merit

RMSD: root mean square deviation

* Values in parentheses refer to the highest resolution shell

§ $R_{sym} = \Sigma_h \Sigma_i | I_{h,i} - < I_h > | / \Sigma_h \Sigma_i I_{h,i}$

‡ R_{cryst} and $R_{free} = \Sigma | F_{obs} - F_{calc} | / \Sigma F_{obs}$; R_{free} calculated for a 7.4 % subset of reflections not used in the refinement

Calculated using MOLPROBITY within the PHENIX suite

Table 2 - Data collection, phasing and refinement statistics.

2.1.2 CENP-M displays the fold of a small G protein

The structure of CENP-M$^{1-171}$ exhibits a globular assembly, comprising a core of five parallel β-strands forming a β-sheet surrounded by six α-helices. A search of the protein structures database (PDB, Protein Data Bank) using the Dali server[136], which aims to identify similarities in protein topology between a query structure and known structures, revealed a structural similarity of CENP-M to members of the Ras superfamily of small GTPases.

To explore this in detail, I employed the PDBeFold server[137] together with manual inspection and performed a multiple 3D structure alignment of CENP-M$^{1-171}$ and known members of the Ras superfamily of small GTPases. In particular, I chose H-Ras (PDB ID: 2CLD), as the chief representative of the superfamily; Rad (PDB ID: 2DPX), the best structure hit from the Dali server; Arl2 (PDB ID: 1KSH), which I then employed for biochemical comparisons with CENP-M; and Ran (Ras-related nuclear protein) (PDB ID: 3GJ0), which constitutes a subfamily involved in nuclear transport, control of DNA synthesis and cell cycle progression and which I then employed for functional comparisons with CENP-M.

In figure 8, the cartoon representations and the topology diagrams of the structures of CENP-M$^{1-171}$ (left) and H-Ras/GDP (right) emphasize the similarity of protein folds. Secondary structure elements that are only present in CENP-M$^{1-171}$ but not in H-Ras/GDP are highlighted in red. These include an extremely short central α-helix and a C-terminal one, which is situated beyond the C-terminus of H-Ras/GDP crystal structure. Conversely, secondary structure elements that are present in H-Ras/GDP but not in CENP-M$^{1-171}$ are indicated in green. In particular, strand β2 is missing in CENP-M$^{1-171}$ and helix α2 is not visible, probably because of the flexibility of this region in CENP-M$^{1-171}$ crystals.

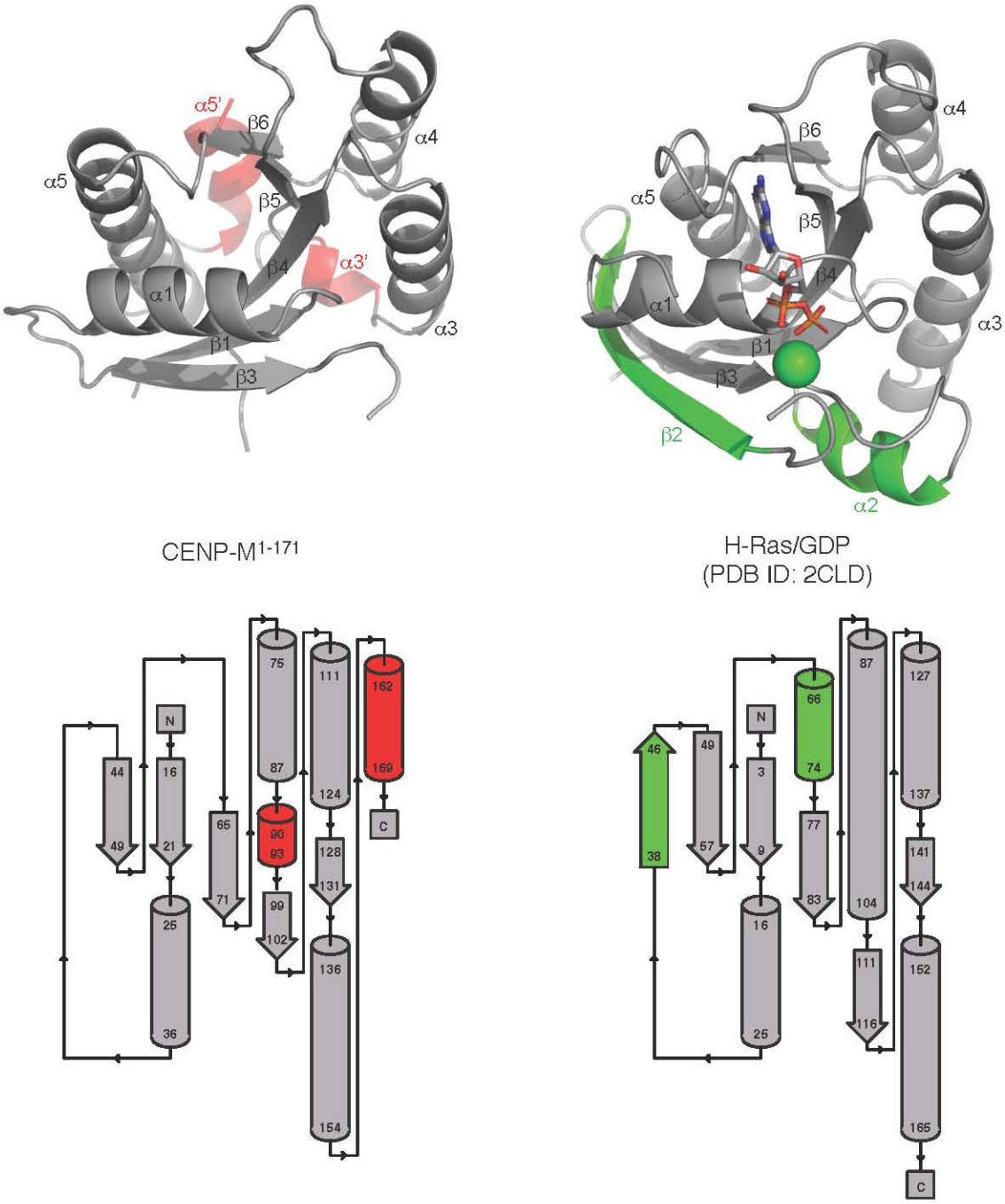


Figure 8 - CENP-M displays the fold of a small G protein.

Cartoon representations (upper panel) and topology diagrams (lower panel) of CENP-M$^{1-171}$ (left) and H-Ras/GDP (right) aligned structures. Secondary structure elements that are present in CENP-M$^{1-171}$ but not in H-Ras/GDP are highlighted in red, while those present in H-Ras/GDP but not in CENP-M$^{1-171}$ are indicated in green.

2.1.3 CENP-M lacks the characteristic motifs of the G domain and does not bind to adenine and guanine nucleotides *in vitro*

Despite overall low sequence similarity, G proteins are characterized by five conserved fingerprint sequence motifs[138], which are indicated in the structure alignment shown in figure 9 A. In particular, the sequence GxxxxGKS/T represents the so-called P-loop, which is involved in phosphate binding; T and DxxGQ/H/T represent switch I and II, respectively, which make contacts with the γ-phosphate of GTP and can undergo important conformational changes upon hydrolysis of GTP to GDP; N/TKxD and S/CAK/L/I are the major determinants of guanine base binding specificity. From the structure alignment is evident that CENP-M lacks all these functionally relevant conserved motifs, suggesting that it does not possess GTPase activity or even GTP binding activity.

To test this possibility, I monitored CENP-M binding to adenine and guanine nucleotides *in vitro* (Figure 9 B). N-methylanthraniloyl (MANT)-labelled nucleotides (ADP, ATP, GDP, GTP) were employed[139]. Protein-nucleotide interactions were detected exploiting the environmental sensitivity of MANT, as its fluorescence quantum yield increases in nonpolar solvents and upon binding to proteins. Arl2, a member of the Ras superfamily of small GTPases, was used as control. In this assay, CENP-M did not bind to any of the tested nucleotides, in accordance with the fact that CENP-M lacks the characteristic motifs of the G domain.

Therefore, we conclude that CENP-M displays the fold, but not the enzymatic activity of a G protein.

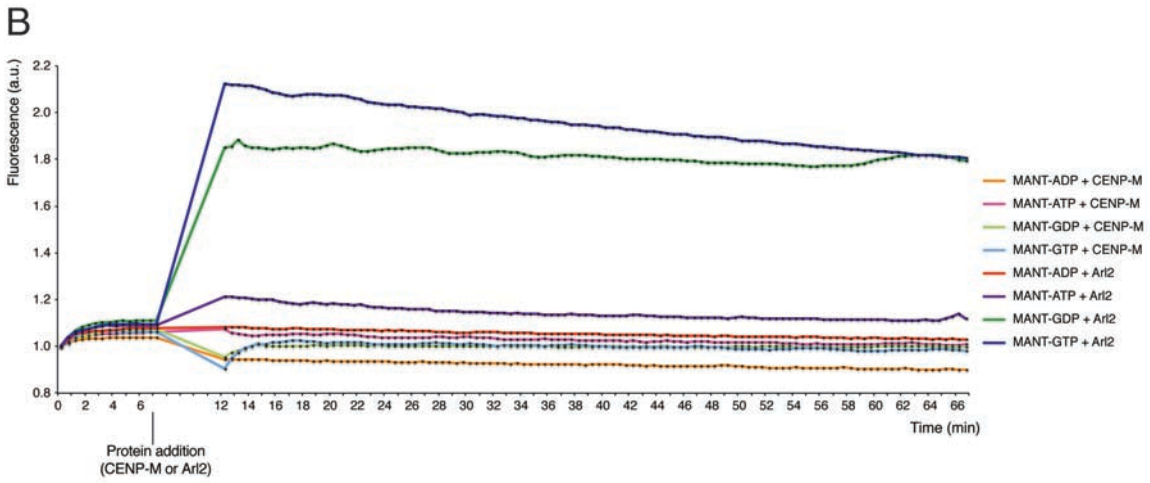


Figure 9 - CENP-M does not display the enzymatic activity of a G protein.

A) CENP-M lacks the characteristic motifs of the G domain. Multiple 3D structure alignment of CENP-M and members of the Ras superfamily of small GTPases, such as H-Ras (2CLD), Ran (3GJ0), Rad (2DPX) and Arl2 (1KSH). Colour by percentage identity is applied; the characteristic motifs of the G domain are indicated below the alignment. B) CENP-M does not bind to adenine and guanine nucleotides *in vitro*. CENP-M binding to adenine and guanine nucleotides was tested *in vitro* employing N-methylanthraniloyl (MANT)-labelled nucleotides (ADP, ATP, GDP, GTP). Arl2 was used as control. CENP-M did not display binding to any of the tested nucleotides.

2.1.4 CENP-H / CENP-K complex is a 1 : 1 assembly predicted to possess an elongated structure enriched in coiled-coils

As already introduced, CENP-M appears to play a central role in the centromeric localization of a number of CCAN components, including the CENP-H / CENP-I / CENP-K group[17,18,75], the CENP-L / CENP-N group[17,18] and the CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group[18]. In turn, CENP-M localization has been shown to depend on all of these protein groups[17,18] and on CENP-T[17]. Stemming from these observations, I sought to gain a better understanding of the molecular mechanisms underlying these relationships by investigating direct interactions involving CENP-M and different CCAN subunits.

I was first able to obtain human CENP-H / CENP-K complex[106] by recombinant coexpression in insect cells and to purify it to homogeneity by an affinity purification through a cleavable His-tag on the CENP-K subunit and additional chromatographic steps (Figure 10). The complex, whose theoretical molecular weight is about 60 kDa, migrates in SEC in close proximity to the 158 kDa marker protein, either suggesting that it forms oligomers or that it has an elongated structure. A

static light scattering analysis estimated the molecular weight of the complex to be compatible with a 1 : 1 assembly (data not shown), thus excluding the first hypothesis. In favour of the second possibility is the fact that protein structure prediction servers, such as I-TASSER[140] and Phyre2[135], predict CENP-H and CENP-K to possess an elongated structure, enriched in coiled-coils. This is in agreement with the output of the program COILS[141], designed for the prediction of coiled-coil regions in proteins.

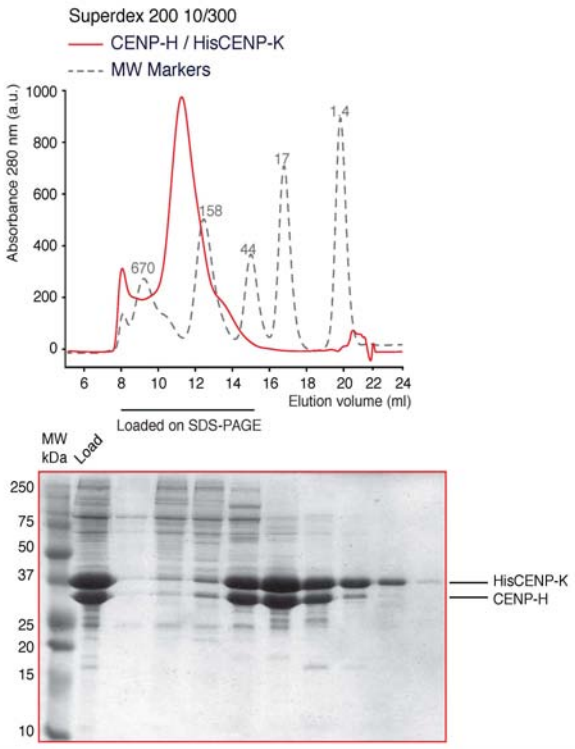


Figure 10 - Final step of CENP-H / HisCENP-K complex purification.

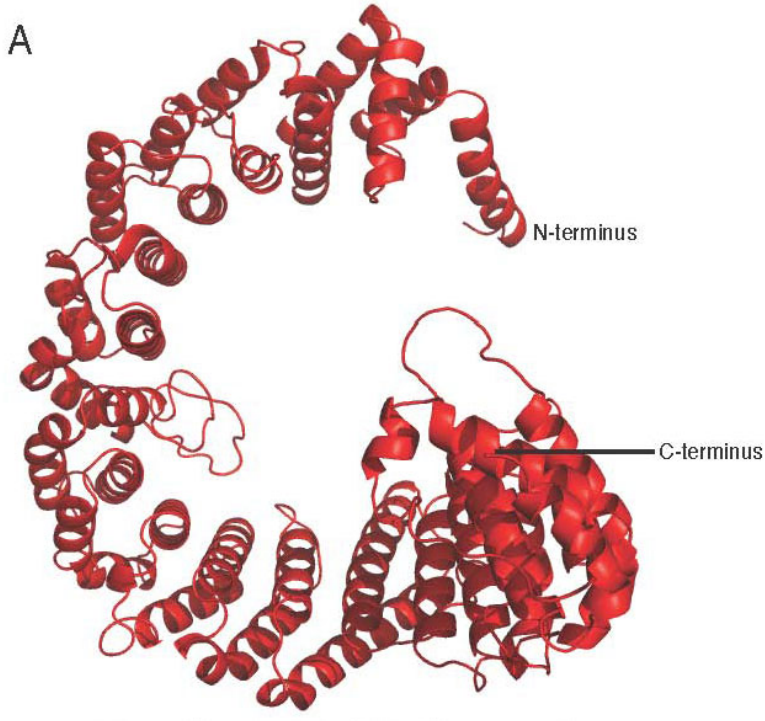
SEC elution profile from a Superdex 200 column and SDS-PAGE analysis of purified recombinant CENP-H / HisCENP-K complex. The complex, whose theoretical molecular weight is about 60 kDa, migrates in close proximity to the 158 kDa marker protein. The most likely reason is that it possesses an elongated structure.

2.1.5 CENP-I is predicted to possess an α-solenoid fold, similarly to the members of the Importin-β family of proteins

Various protein structure prediction servers (such as I-TASSER[140], Phyre2[135] and Robetta[142]) model CENP-I structure (Figure 11 A) as displaying an α-solenoid fold, which consists of α-helices arranged in a curved pattern, and invariably use structures of so-called karyopherins as threading

templates. Karyopherins represent a class of proteins involved in the nucleocytoplasmic transport of molecules through the nuclear pore complex (NPC) in eukaryotic cells[143]. Remarkably, structures including the karyopherin Importin-β rank among the top templates chosen by those structure prediction servers. Importin-β has been previously shown to interact with the small G protein Ran[144]. The structural similarity of CENP-I with Importin-β and of CENP-M with Ran raised the hypothesis that CENP-M could directly interact with CENP-I (Figure 11 B) in a similar fashion to Ran interacting with Importin-β (Figure 11 C).

The structure of members of the Importin-β family consists of a tandem series of HEAT repeats, which derive their name from the approximately forty-residue tandem sequence repeats first recognized in Huntingtin, elongation factor 3, PR65/A subunit of protein phosphatase 2A and kinase TOR. Each HEAT repeat comprises two antiparallel α-helices linked by a turn. HEAT repeats stack together to form two C-shaped arches that ultimately generate a helicoidal molecule. The relative orientation of the two arches differs among members of the family and is also influenced by the binding of interacting partners[145].



CENP-I (I-TASSER model)

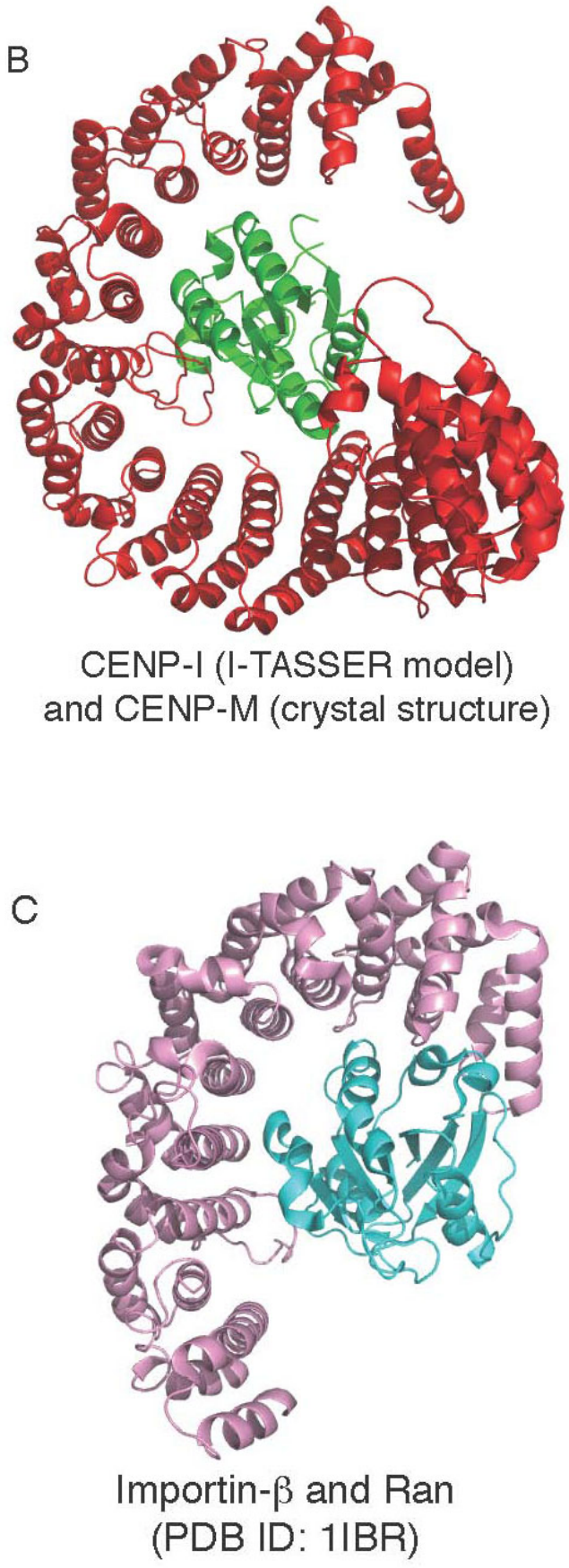


Figure 11 - CENP-M might interact with CENP-I in a similar fashion to Ran interacting with Importin-β.

A) I-TASSER model of CENP-I structure (red). CENP-I is predicted to possess an α-solenoid fold, which consists of α-helices arranged in a curved pattern. To estimate the accuracy of structure predictions, a scoring function (C-score) is

associated with I-TASSER models. C-score is typically in a range from -5 to 2, where a higher score reflects a model of better quality. Both false positive and false negative rates are estimated to be below 0.1 when a C-score > -1.5 is displayed[140]. Specifically, the CENP-I model shown above was associated with a C-score of -1. B) I-TASSER model of CENP-I structure (red) and CENP-M crystal structure (green). The position of CENP-I and CENP-M was modelled by superposition on the structure of Importin-β (pink) and Ran (light blue), which is depicted in C. This was performed using the Secondary Structure Matching (SSM) superposition tool in Coot[137]. All protein structures are displayed as cartoon representations.

2.1.6 CENP-I$^{57-281}$ and CENP-H / CENP-K assemble into a ternary complex, which does not bind to CENP-M

In order to investigate the hypothesis that CENP-M directly interacts with CENP-I, I first tried to express and purify recombinant human CENP-I. Either the full-length version or various truncated constructs, designed according to structure predictions, were tested. Also, different expression systems, such as bacteria and insect cells, were employed. The only available CENP-I cDNA clone contained the 5' moiety of the coding region, but was lacking the segment encoding amino acids 523 - 756 (C-terminus). Thus, I ordered a synthetic full-length CENP-I cDNA, codon-optimized for expression in bacteria as well as insect cells. After a number of attempts, the only CENP-I construct that could be produced in a soluble and well-behaved fashion was the one comprising residues 57 - 281. In particular, it was produced as a C-terminal fusion to a non-cleavable His-tag. This construct is predicted to comprise a first domain of HEAT repeats in the models of CENP-I structure.

Analytical size-exclusion chromatography (SEC) migration shift assays were performed to test *in vitro* the binding of this CENP-I construct to other inner kinetochore proteins that we were able to recombinantly produce and purify in the laboratory. In particular, CENP-I$^{57-281}$ was incubated with approximately stoichiometric amounts of candidate binding partners and the migration of the resulting species in SEC was observed, as a read-out of the formation of larger protein assemblies. In the context of this dissertation, it is relevant to point out that CENP-I$^{57-281}$ did not bind to CENP-M (Figure 12 A), while it formed an apparently stoichiometric ternary complex with CENP-H / CENP-K (Figure 12 B). Also, CENP-M did not interact with CENP-H / CENP-K complex (Figure 12 C).

The subsequent step was the attempt to coexpress in insect cells CENP-H / CENP-K and various

constructs of CENP-I, hoping that this approach could stabilize longer CENP-I fragments. Unfortunately, this was not the case and CENP-H / CENP-K / CENP-I$^{57-281}$ was the only soluble and well-behaved construct that could be produced. Further analytical SEC migration shift assays confirmed that, even when assembled in a ternary complex, CENP-H / CENP-K / CENP-I$^{57-281}$ do not display any binding to CENP-M (Figure 12 D).

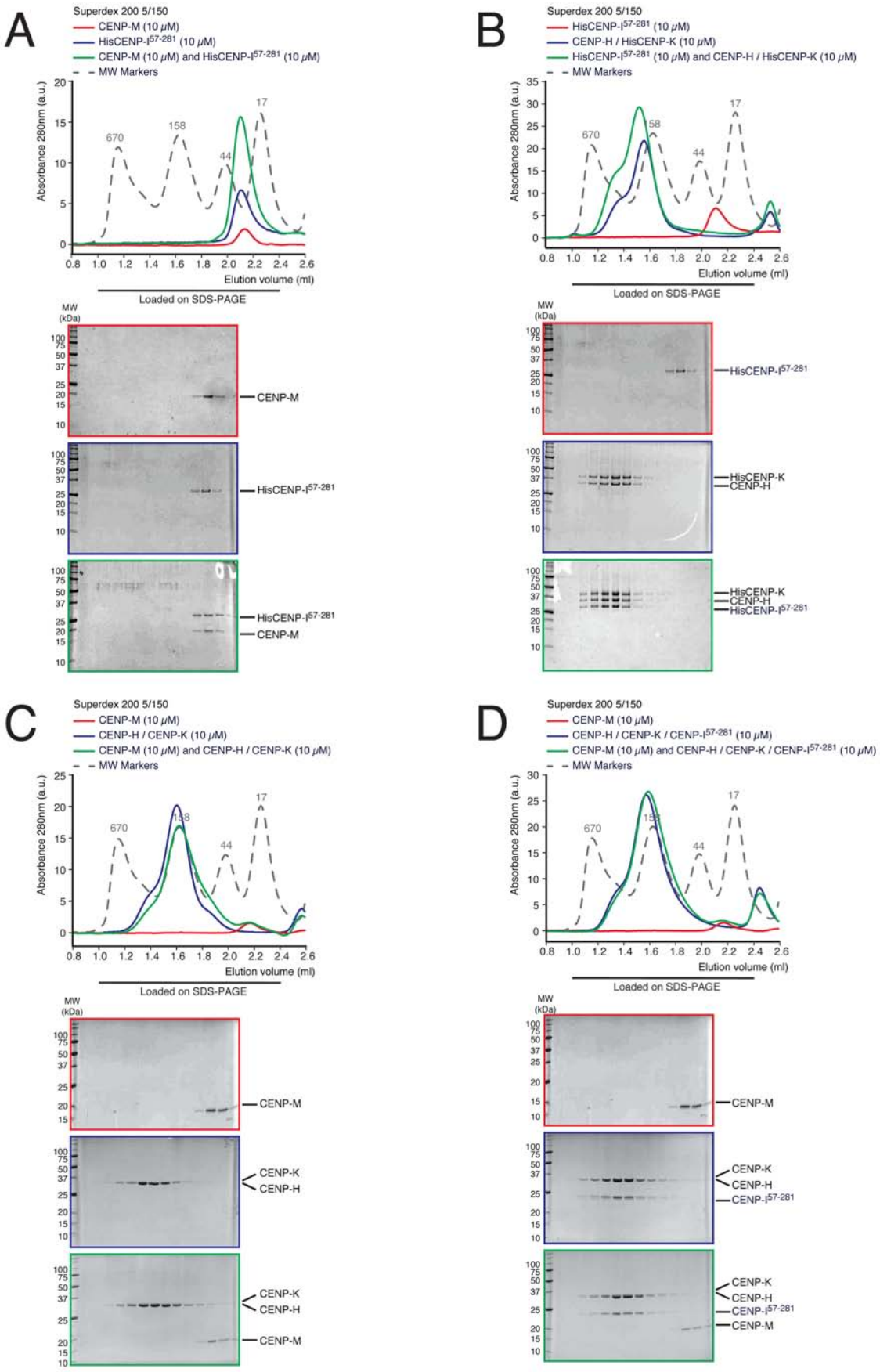


Figure 12 - CENP-I$^{57-281}$ and CENP-H / CENP-K assemble into a ternary complex, which does not bind to CENP-M.

A) SEC elution profiles and SDS-PAGE analyses of CENP-M (red), HisCENP-I$^{57-281}$ (blue) and their stoichiometric combination (green). No direct binding of these two proteins is observed. B) Same experimental setting as in A, but using

HisCENP-I$^{57-281}$ and CENP-H / HisCENP-K complex. Upon their combination, the formation of a ternary complex occurs. C) No interaction between CENP-M and CENP-H / CENP-K complex is detected. D) CENP-M is not incorporated in CENP-H / CENP-K / CENP-I$^{57-281}$ complex.

2.1.7 CENP-H, CENP-K, CENP-I^{57-C} and CENP-M assemble into a quaternary complex

Another series of coexpression tests in insect cells was then performed, including also CENP-M. Coexpression with GST-CENP-M resulted in the solubilisation of a longer CENP-I construct, comprising residues 57 - C-terminus. Coinfection with a virus expressing His-tagged CENP-K and CENP-H proved beneficial for further stabilizing this protein assembly and yielded a well-behaved quaternary CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex. This discovery was followed by the generation of a unique virus encoding these four proteins and the scale up of expression and purification efforts (Figure 13).

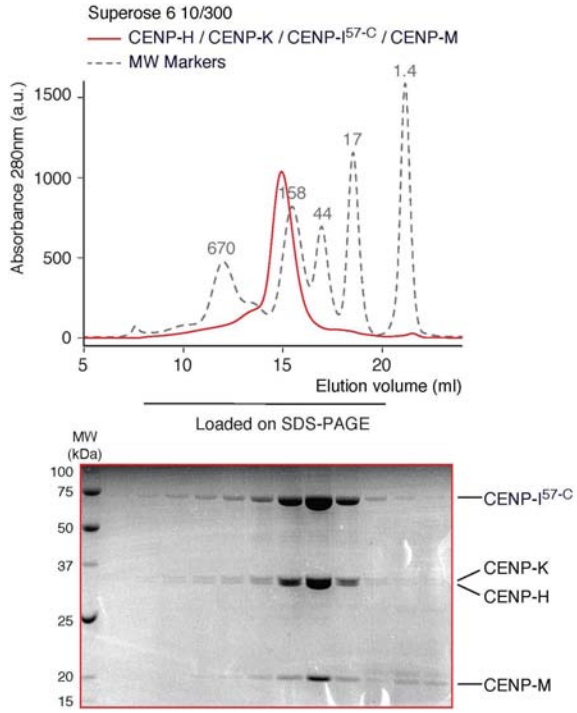
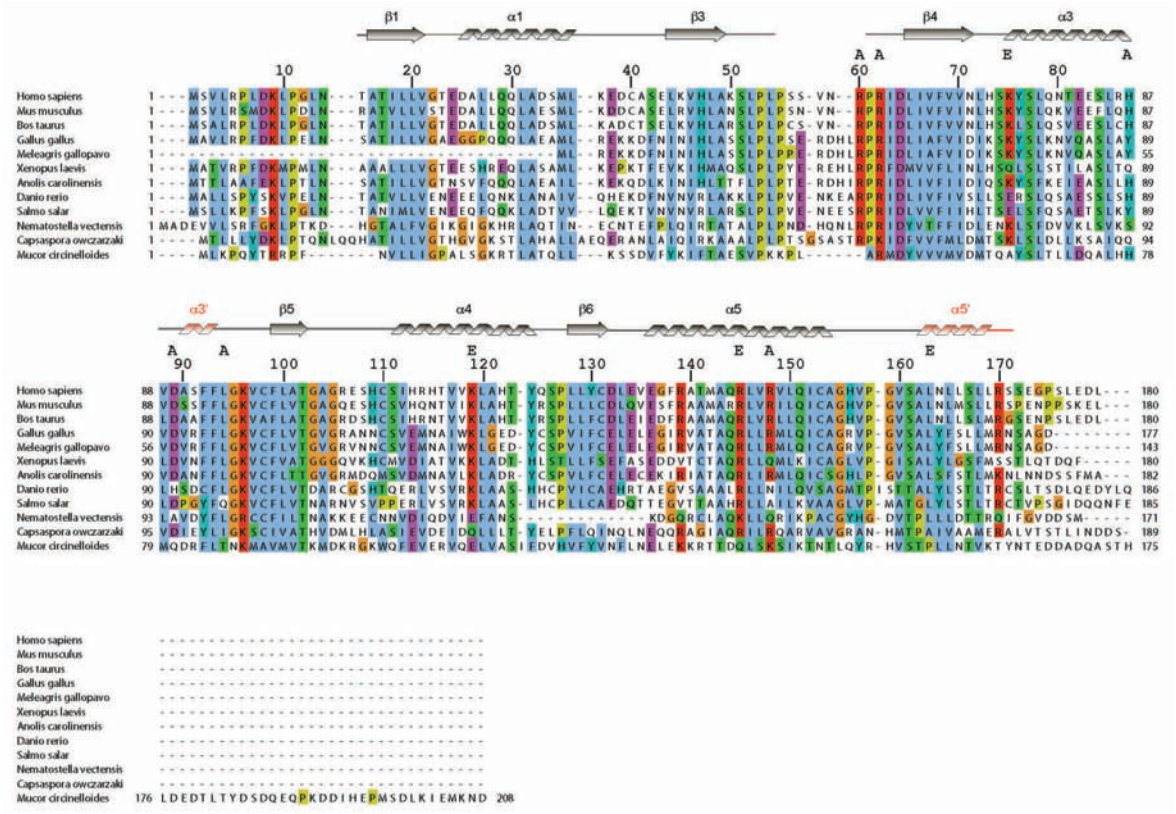


Figure 13 - Final step of CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex purification.

SEC elution profile from a Superose 6 column and SDS-PAGE analysis of purified recombinant CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex. The migration of the complex in SEC suggests that it is a 1 : 1 : 1 : 1 globular assembly.

2.1.8 Effects of point mutations in CENP-M sequence on its interaction with CENP-I

In order to validate the direct interaction between CENP-M and CENP-I *in vitro* and to elucidate which residues of CENP-M mediate the binding, I tested the effects of several point mutations in CENP-M sequence. Mutations were designed taking into account both sequence conservation and structural information. In particular, my rationale was to mutate highly conserved residues exposed on the protein surface and rather preserve amino acids buried in the protein core and most likely involved in determining the protein fold. Also, I tried to hit various areas of the protein and decided for single or double alanine or charge reversal mutations depending on the chemical surroundings of the selected residues. Specifically, the following CENP-M mutants were devised: R60A + R62A, L94A + L163E, R145E + R148A, K119E, K75E, H87A + D89A. Their position along CENP-M sequence and their evolutionary conservation, as well as their location in CENP-M structure are depicted in Figure 14.



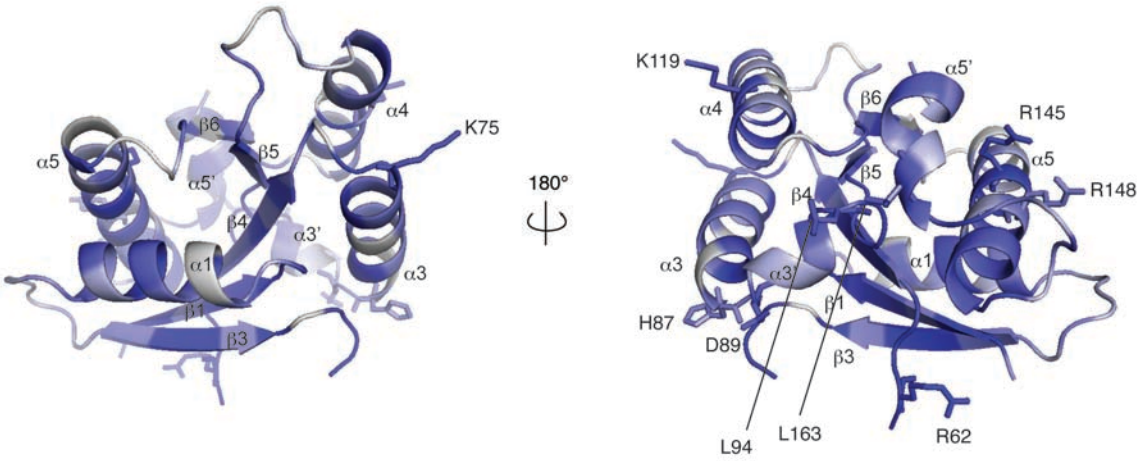


Figure 14 - Overview of CENP-M mutants.

Upper panel: the MUSCLE server[146] was employed to perform a multiple sequence alignment of CENP-M orthologs. Default Clustal colour scheme is applied. Residue numbers above the alignment refer to the human CENP-M sequence. Secondary structure elements are illustrated. CENP-M mutations are indicated above the alignment. Lower panel: the ConSurf server[147] was employed to display evolutionary conservation on CENP-M structure. Colour by percentage identity is applied and five discrete conservation bins (from white to blue) are defined. CENP-M structure is represented as a cartoon, where residues selected for the mutational analysis are highlighted as sticks.

First of all, I ascertained that CENP-M mutants did not affect the protein structural integrity. With this aim, CENP-M wild type (wt) and mutants were all expressed in bacteria and purified employing the same procedure. Afterwards, approximately the same amount of each protein was separately loaded on an analytical SEC column and the various chromatographic profiles were compared (Figure 15). CENP-M wt and mutants displayed a uniform behaviour, suggesting that in all these cases the protein fold is preserved.

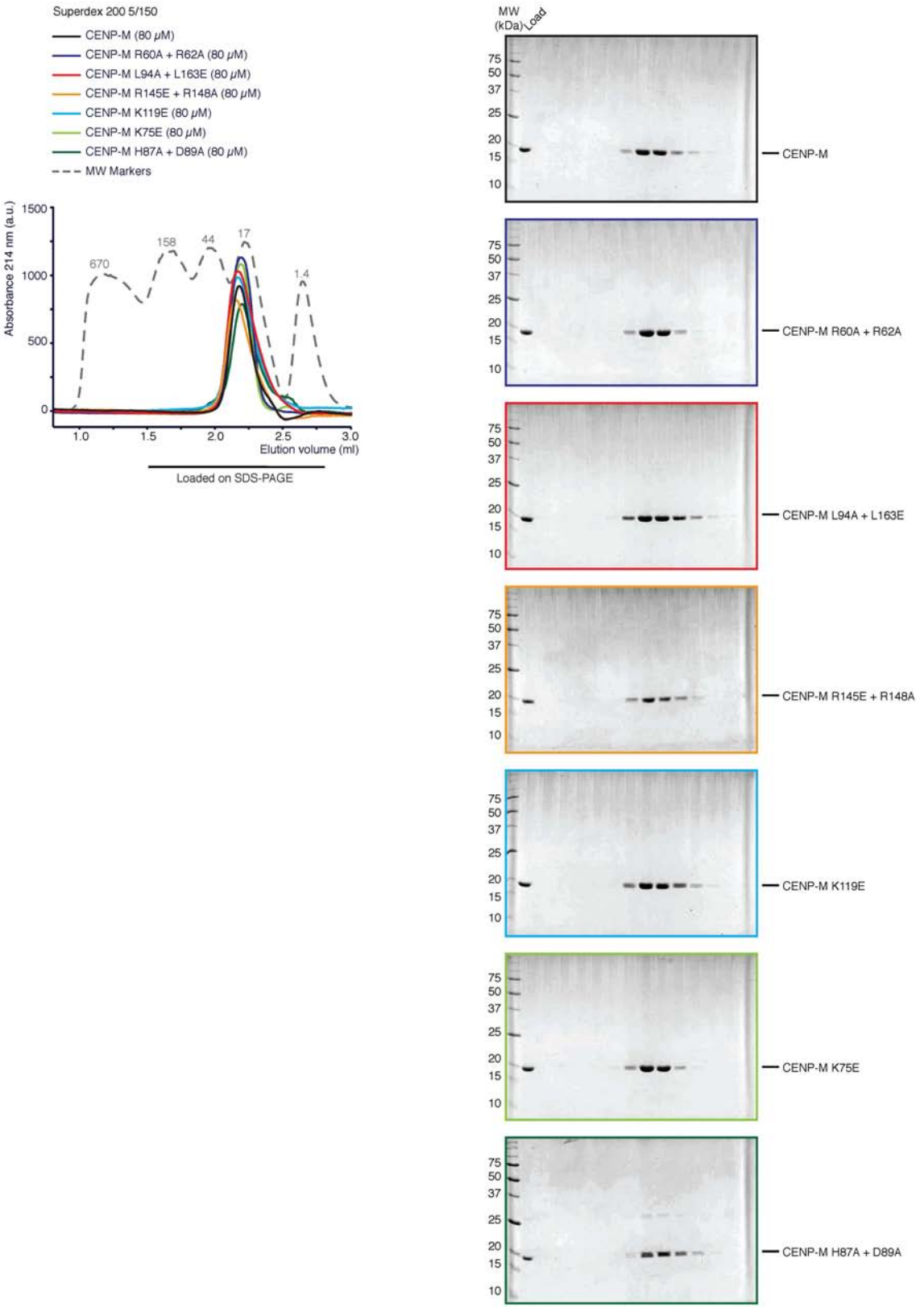


Figure 15 - CENP-M mutants maintain the protein structural integrity.

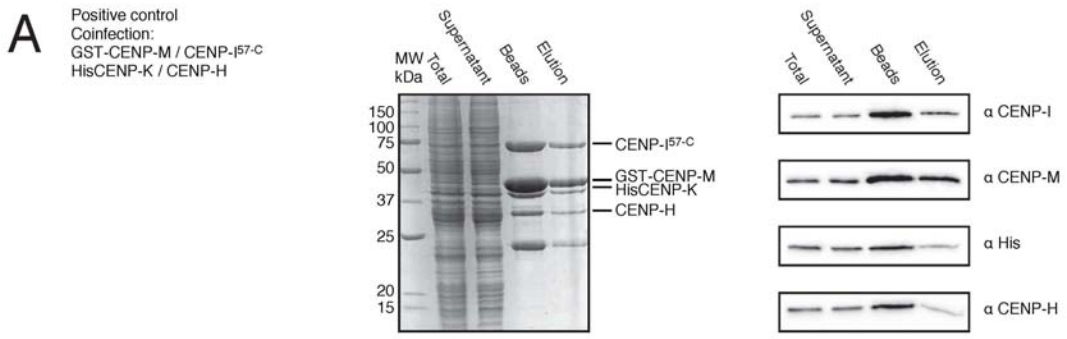
SEC elution profiles from Superdex 200 5/150 column and SDS-PAGE analyses of purified recombinant CENP-M wt and mutated constructs. All CENP-M mutants behave as the wt protein, thus indicating that the protein fold is preserved.

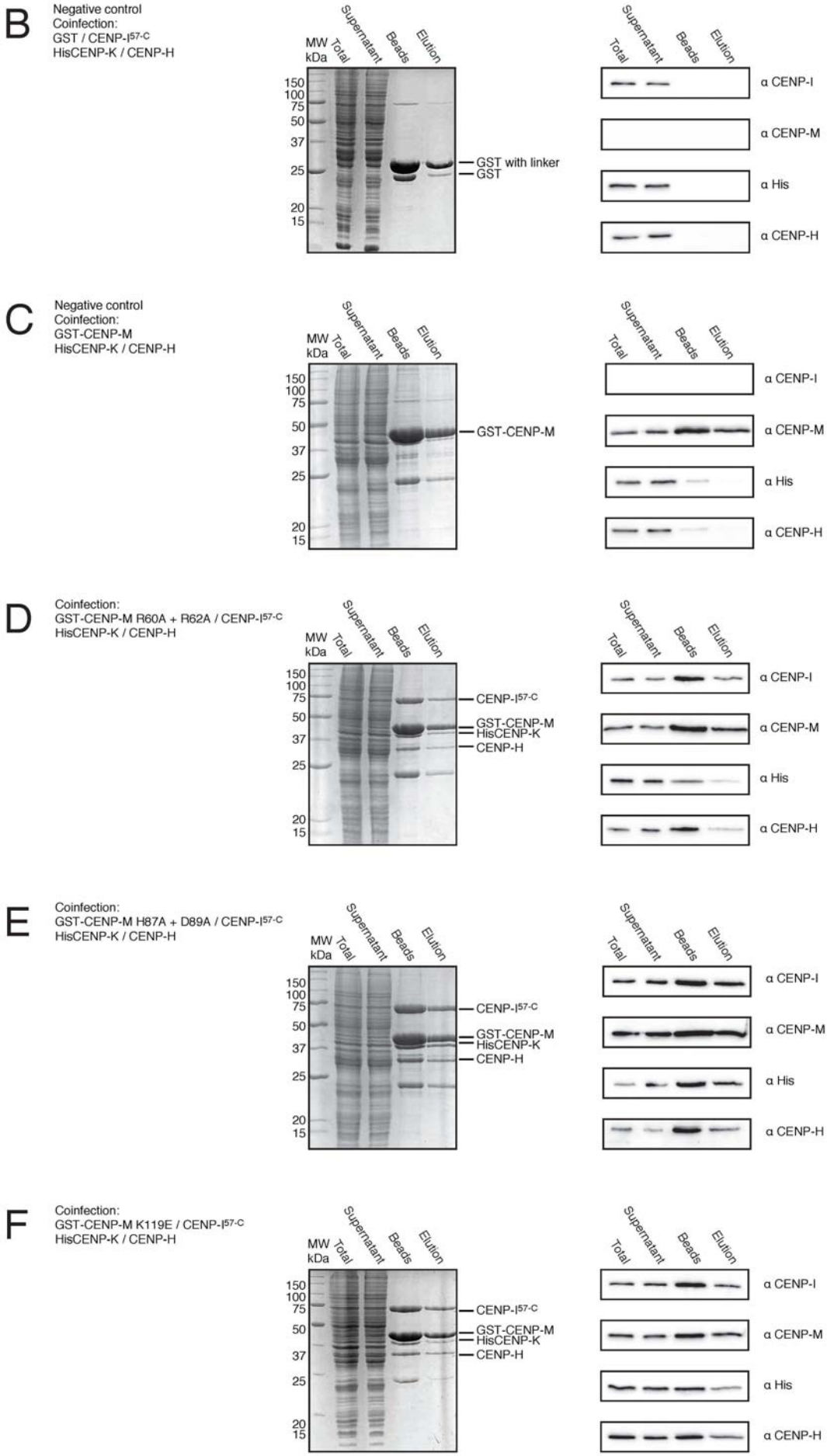
I then sought to verify the effects of these mutations on the interaction between CENP-M and

CENP-I. With this aim, I coexpressed in insect cells GST-CENP-M wt or mutants with CENP-I$^{57\text{-}C}$ and performed a coinfection with a virus expressing His-tagged CENP-K and CENP-H. The experimental setting then included a GST-pull-down, where samples of total lysate, supernatant, beads before elution, and elution were analysed by SDS-PAGE and Coomassie blue staining and by Western blotting against the four proteins of interest.

As already mentioned, coexpression of GST-CENP-M wt with CENP-I$^{57\text{-}C}$, HisCENP-K and CENP-H results in the formation of a quaternary complex (Figure 16 A). As expected, when CENP-M is absent, CENP-I$^{57\text{-}C}$, HisCENP-K and CENP-H are expressed but are not found on the beads before elution and in the elution, indicating that they do not interact with GST and do not display unspecific binding to Glutathione Sepharose beads (Figure 16 B). Also, when CENP-I$^{57\text{-}C}$ is absent, GST-CENP-M, HisCENP-K and CENP-H are expressed, but HisCENP-K and CENP-H do not interact with GST-CENP-M, in line with analytical SEC results, and, therefore, are not present on the beads before elution and in the elution (Figure 16 C).

Coming to CENP-M mutants, some of them (in particular, R60A + R62A, H87A + D89A, K119E, K75E) do not impair CENP-M binding to CENP-I$^{57\text{-}C}$ and allow the formation of a quaternary complex with HisCENP-K and CENP-H (Figure 16 D - G). Conversely, two of the tested CENP-M mutants (in particular, L94A + L163E and R145E + R148A) completely abolish CENP-M binding to CENP-I$^{57\text{-}C}$. Thus, CENP-I$^{57\text{-}C}$, HisCENP-K and CENP-H, despite being expressed, are not present on the beads before elution and in the elution (Figure 16 H - I). L163, located in helix α5', and L94, situated in a nearby loop, are part of a hydrophobic patch; R145 and R148 stem in a parallel orientation from helix α5 and contribute to the formation of a positively-charged basic patch. These two patches on CENP-M surface are quite close in space, thus suggesting that this area of the protein is crucial for the binding to CENP-I.




B
Negative control
Coinfection:
GST / CENP-I57-C
HisCENP-K / CENP-H
MW kDa
Total
Supernatant
Beads
Elution
150
100
75
50
37
25
20
15
GST with linker
GST
α CENP-I
α CENP-M
α His
α CENP-H
C
Negative control
Coinfection:
GST-CENP-M
HisCENP-K / CENP-H
GST-CENP-M
D
Coinfection:
GST-CENP-M R60A + R62A / CENP-I57-C
HisCENP-K / CENP-H
CENP-I57-C
GST-CENP-M
HisCENP-K
CENP-H
E
Coinfection:
GST-CENP-M H87A + D89A / CENP-I57-C
HisCENP-K / CENP-H
F
Coinfection:
GST-CENP-M K119E / CENP-I57-C
HisCENP-K / CENP-H

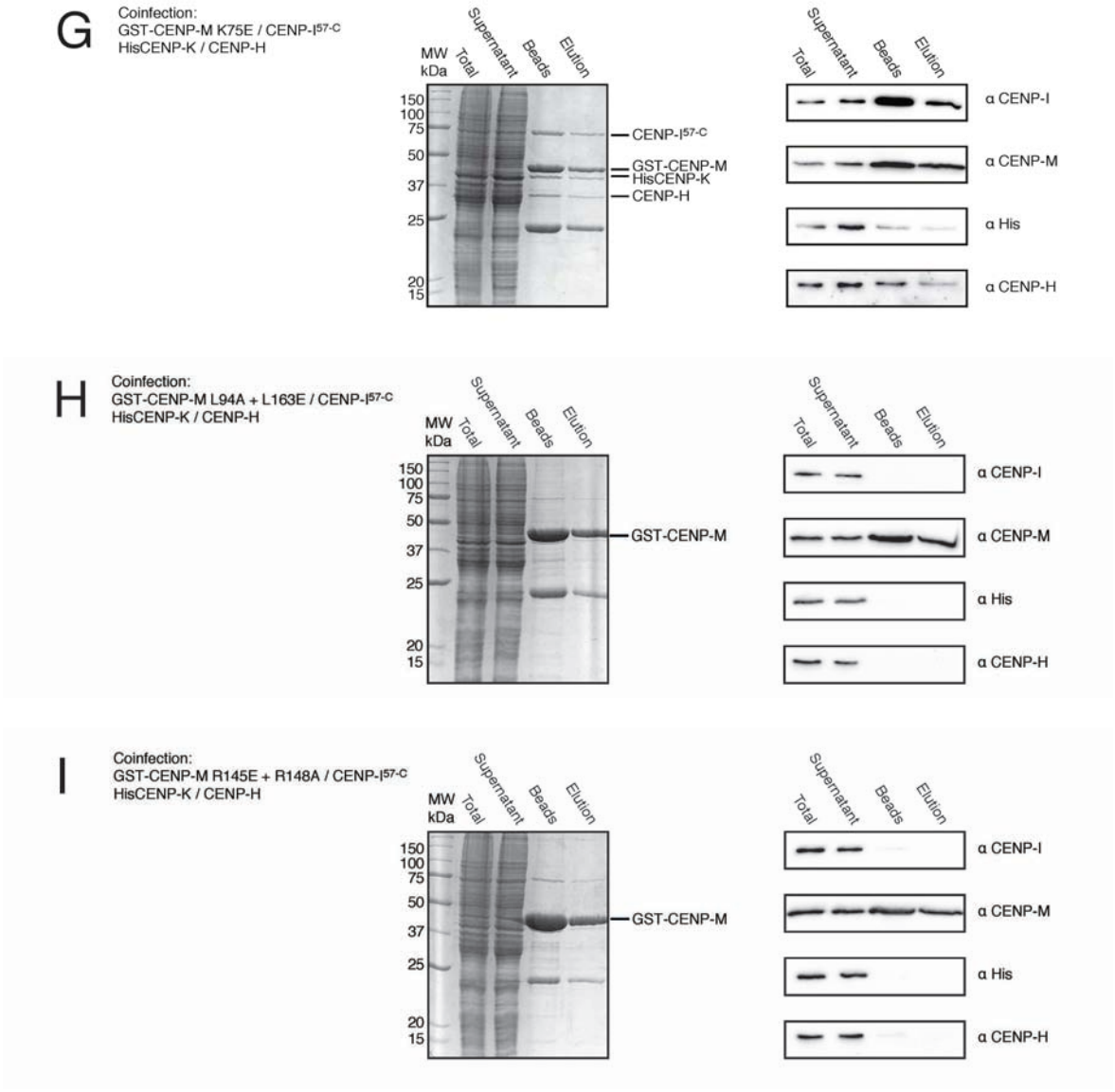


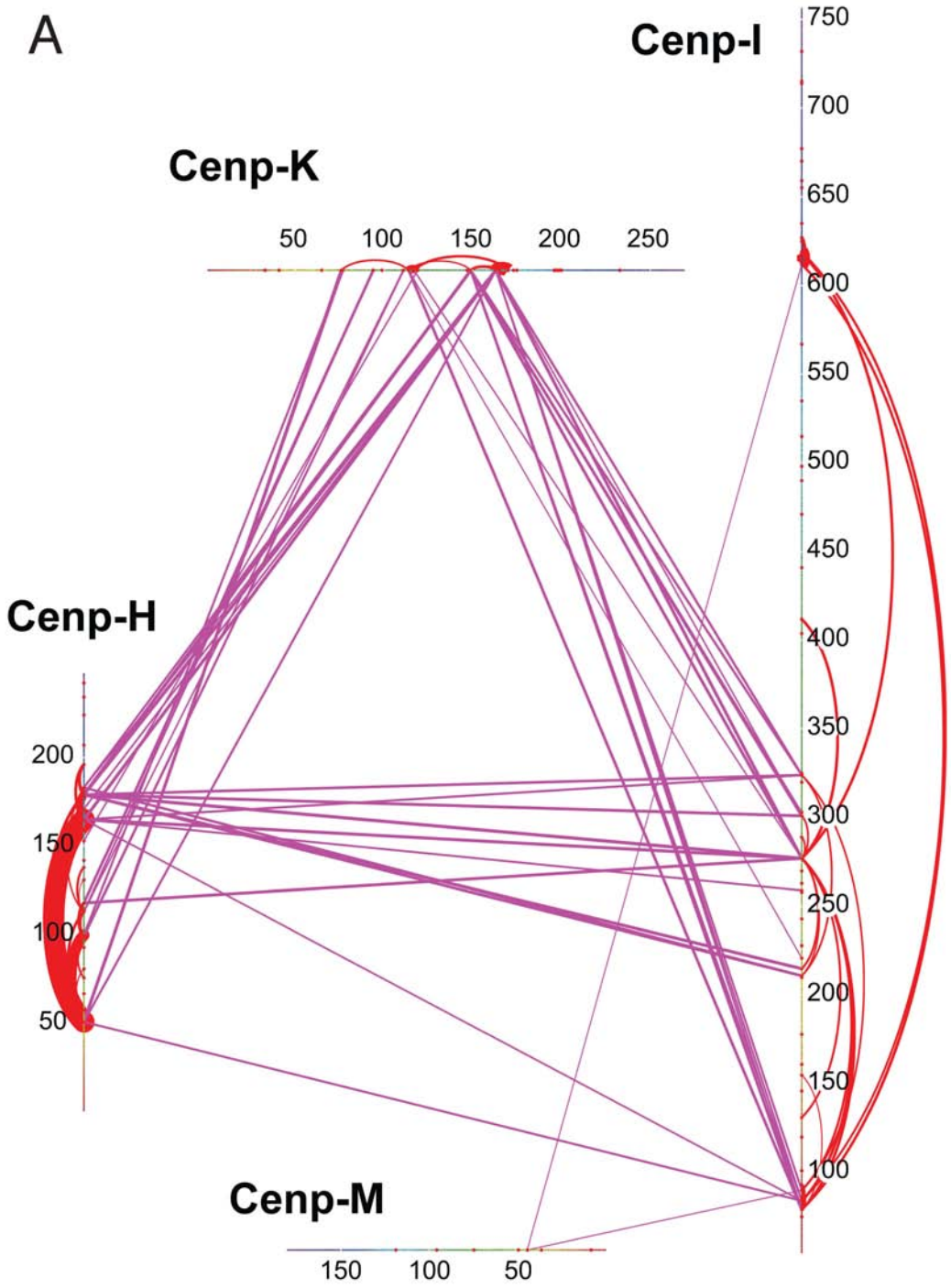
Figure 16 - Effects of point mutations in CENP-M sequence on its interaction with CENP-I.

GST-pull-down experiments on GST-CENP-M wt or mutants coexpressed in insect cells with CENP-I^{57-C}, HisCENP-K and CENP-H. Samples of total lysate, supernatant, beads before elution and elution were analysed by SDS-PAGE and Coomassie blue staining (left panels) and by Western blotting against the four proteins of interest (right panels). A) Formation of a quaternary complex upon coexpression of GST-CENP-M, CENP-I^{57-C}, HisCENP-K and CENP-H. B) When CENP-M is absent, CENP-I^{57-C}, HisCENP-K and CENP-H do not interact with GST and do not display unspecific binding to Glutathione Sepharose beads. C) When CENP-I^{57-C} is absent, HisCENP-K and CENP-H are not pulled-down by GST-CENP-M. D to G) Four of the tested CENP-M mutants do not impair its binding to CENP-I^{57-C} and, thus, allow the formation of the quaternary complex with HisCENP-K and CENP-H. H - I) Two of the tested CENP-M mutants abolish the binding of CENP-M to CENP-I.

2.1.9 Experiments of cross-linking coupled with mass spectrometry provide insights into the spatial organization of CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex

Chemical cross-linking coupled with mass spectrometry is a recent low-resolution structural technique for the characterization of the architecture of protein complexes[148]. Specifically, the method exploits the ability of the bifunctional reagent BS2G (bis[sulfosuccinimidyl]glutarate) to cross-link primary amines of lysines situated within a distance compatible with the length of the cross-linker (7.7 Å). Therefore, this approach provides information about the spatial proximity of lysine residues that become cross-linked in protein complexes in solution.

CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex was subjected to this analysis, in order to obtain information about its spatial organization (Figure 17 A). Also, samples of CENP-H / CENP-K / CENP-I$^{57-281}$ complex (Figure 17 B) and of CENP-H / CENP-K complex (Figure 17 C) were investigated for comparison purposes. Finally, CENP-M was analysed to validate the results of cross-linking coupled with mass spectrometry (Figure 17 D) by comparing them with the protein structure determined by X-ray crystallography. The analyses were performed thanks to a collaboration with Dr. Franz Herzog's laboratory at the Ludwig Maximilian University in Munich.


A
Cenp-I
Cenp-K
Cenp-H
Cenp-M
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
200
150
100
50
50
100
150
200
250
150
100
50

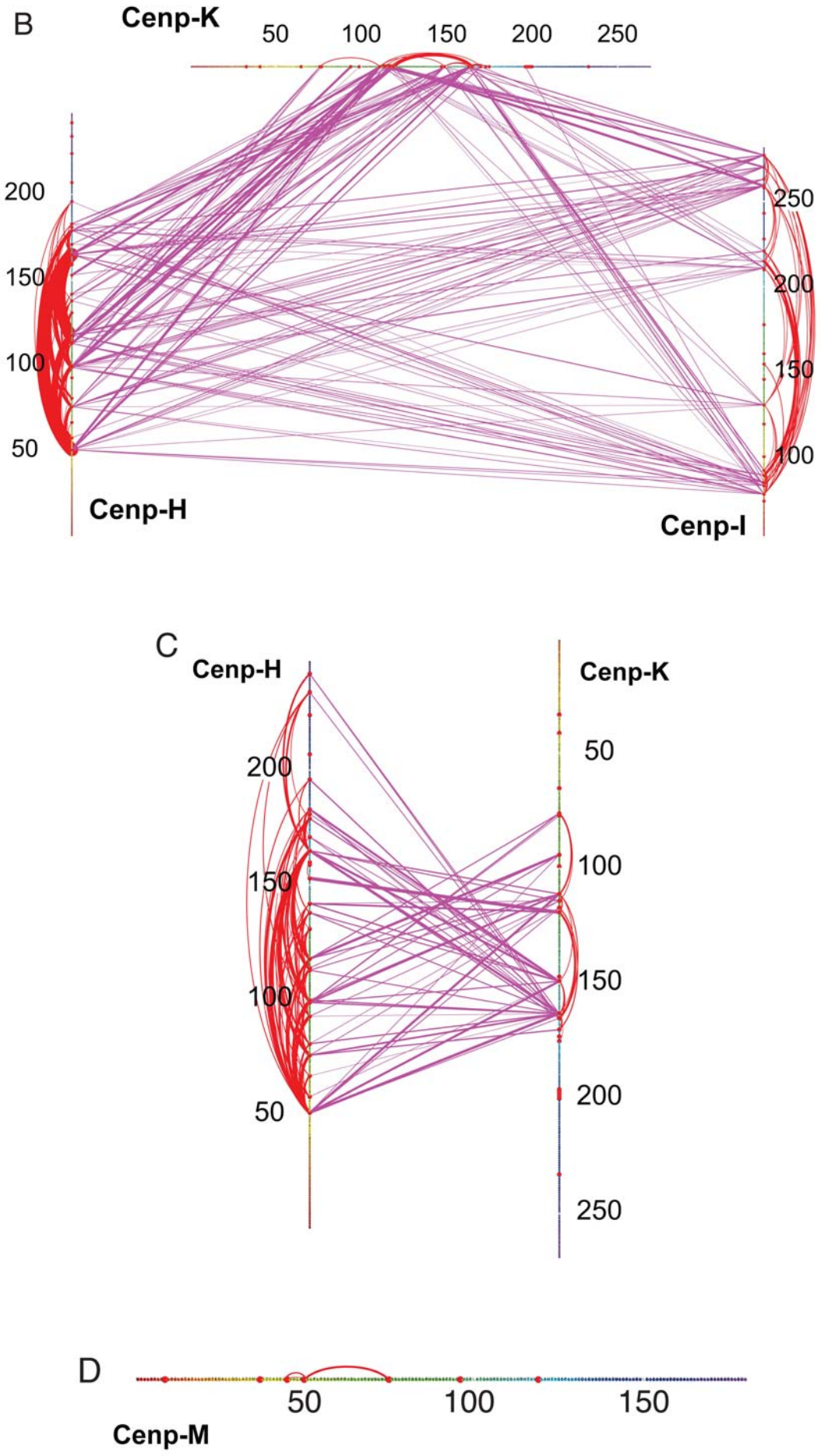


Figure 17 - Experiments of cross-linking coupled with mass spectrometry.

CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex (A), CENP-H / CENP-K / CENP-I$^{57-281}$ complex (B), CENP-H / CENP-K complex (C) and CENP-M (D) were analysed by cross-linking coupled with mass spectrometry in order to obtain information about their spatial organization. Protein sequences are displayed and residues are coloured with a rainbow gradient ranging from red at the N-terminus to violet at the C-terminus. Amino acid numbers are indicated every 50 residues. Lysines are highlighted with red dots. Cross-links between lysine pairs are represented as lines, whose thickness is proportional to the identification score of the corresponding peptide pair by mass spectrometry. Intra-molecular cross-links are coloured in red, while inter-molecular cross-links are coloured in violet.

Without dwelling upon details, a few considerations can be derived from these experiments. First, they clearly indicate that CENP-H and CENP-K are in close spatial proximity to the first ~350 amino acids of CENP-I, but not to the rest of this protein. In fact, not only in CENP-H / CENP-K / CENP-I$^{57-281}$ complex, but also in CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex all the inter-molecular cross-links between CENP-I and CENP-H and / or CENP-K occur within the first ~350 amino acids of CENP-I. This is in line with the fact that coexpression experiments in insect cells with CENP-H and CENP-K could not stabilize CENP-I constructs longer than the one comprising amino acids 57 - 281.

The coexpression with CENP-M was instead the key to solubilizing a CENP-I construct extending up to the C-terminus of the protein. Consistently, inter-molecular cross-links can be detected between CENP-M and two regions of CENP-I, one towards its N-terminus and one towards its C-terminus, which are distant in its sequence but might come close in its structure, as suggested by the presence of intra-molecular cross-links between them.

In addition, no inter-molecular cross-links are found between CENP-M and CENP-H and / or CENP-K in the context of CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex. This is in agreement with the fact that I could not detect any interaction between CENP-M and CENP-H / CENP-K complex both in analytical SEC migration shift assays and in GST-pull-downs.

Regarding CENP-M alone, the two observed intra-molecular cross-links are consistent with the distances between lysines that can be measured in the crystal structure of CENP-M.

2.1.10 Analysis of CENP-M interactions *in vitro* with other kinetochore components and nucleosomes through analytical SEC migration shift assays and with microtubules through a cosedimentation assay

As already introduced, CENP-M plays a critical function in the recruitment of a number of CCAN components to the centromere. In the previous sections I have disclosed its role as a direct interacting partner of CENP-I and in the context of a quaternary complex including also CENP-H and CENP-K. In the search for other direct interactions involving CENP-M, I tested its binding to other known kinetochore components and nucleosomes *in vitro* through analytical SEC migration shift assays.

CENP-M was reported to be a member of the so-called CENP-A nucleosome associated complex (NAC)[17], as it associated uniquely with CENP-A but not with H3 nucleosomes after multiple tandem affinity purifications from HeLa cell lines expressing either TAP-tagged CENP-A or H3. I therefore assessed CENP-M binding to *in vitro* reconstituted mononucleosomes containing either H3 or CENP-A. Given that nucleosome cores possess a two-fold symmetry axis, I decided to combine nucleosomes and CENP-M in the binding reaction using a 1 : 2 molarity ratio. In both cases, no interaction with CENP-M was observed (Figure 18 A and B). This result is in line with the fact that no binding of CENP-M to CENP-A nucleosomes *in vitro* was detected by electrophoretic mobility shift assays[56].

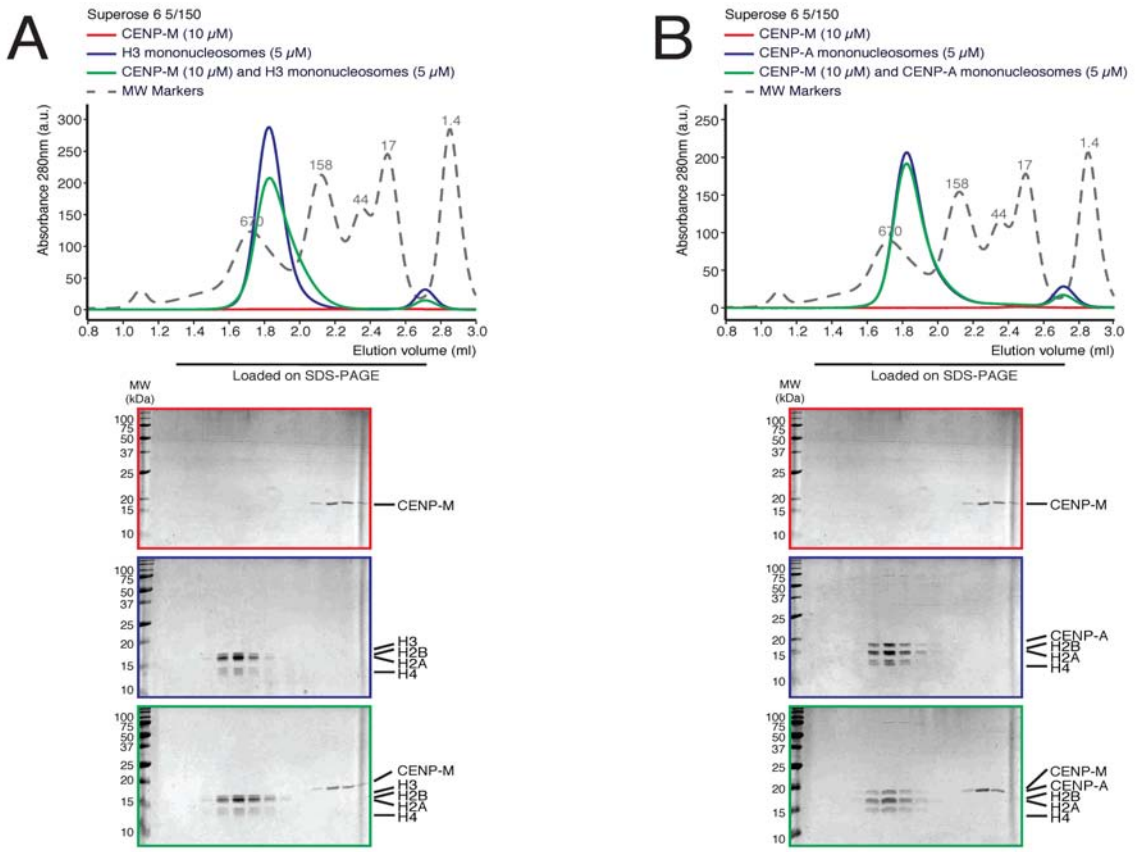


Figure 18 - CENP-M does not bind to H3 and CENP-A mononucleosomes *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of CENP-M (red), H3 mononucleosomes (blue) and their combination (green). B) Same experimental setting as in A, but using CENP-A mononucleosomes. In both cases, no direct interaction between nucleosomes and CENP-M is observed.

I then hypothesized that the association of CENP-M with CENP-A nucleosomes[17] might be indirect and mediated by proteins which in turn directly interact with CENP-A nucleosomes, such as CENP-C[57] and CENP-N[56].

So far, in the laboratory we have been unable to produce and purify recombinant full-length CENP-C in a soluble and well-behaved fashion. However, various CENP-C constructs that overall encompass the whole protein sequence could be obtained. CENP-N was reported to form a complex with CENP-L[56] and in the laboratory we indeed succeeded in producing the CENP-L / CENP-N complex, overcoming the difficulties that we were facing in obtaining these two proteins singularly.

Analytical SEC migration shift assays did not reveal any direct interaction of CENP-M with any of the various CENP-C constructs (Figure 19 A to C) or with CENP-L / CENP-N complex (Figure 19 D).

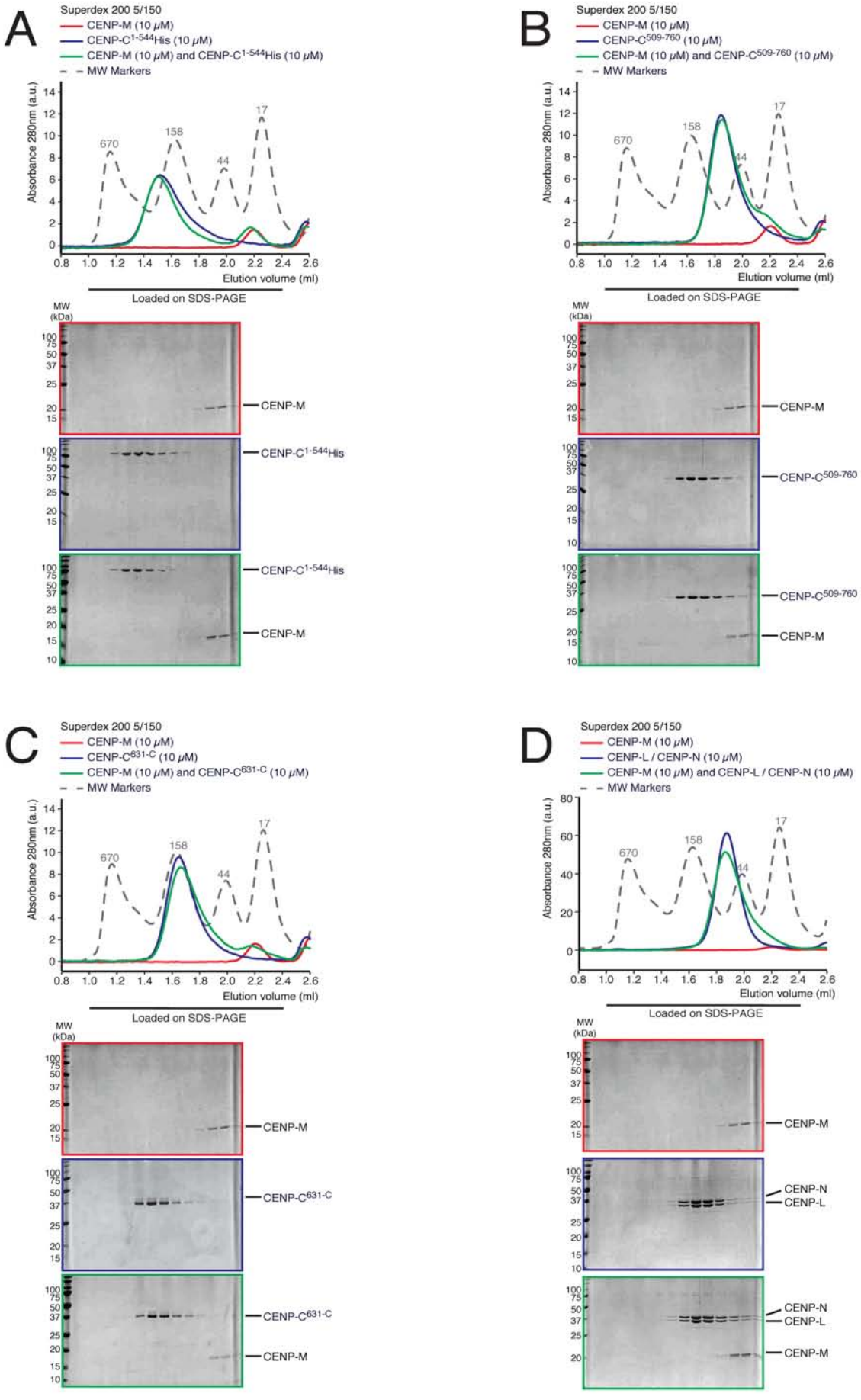


Figure 19 - CENP-M does not bind to various CENP-C constructs and CENP-L / CENP-N complex *in vitro*.

A to C) SEC elution profiles and SDS-PAGE analyses of CENP-M (red), various CENP-C constructs (blue) and their combination (green). No direct binding is observed. D) Same experimental setting as in A, but using CENP-L / CENP-N complex. Also in this case, no interaction with CENP-M occurs.

CENP-M has been shown to be required for the centromeric localization of the CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group of proteins, while in turn depending on them for its own recruitment[18].

In the laboratory we accomplished the production of a quaternary CENP-O / CENP-P / CENP-Q / CENP-U complex and of CENP-O / CENP-P and CENP-Q / CENP-U sub-complexes. Also, CENP-R can be obtained singularly, but it does not become incorporated into CENP-O / CENP-P / CENP-Q / CENP-U complex.

Therefore, I tested if any of these samples has the ability to bind to CENP-M *in vitro*, but no direct interaction could be detected (Figure 20 A to D).

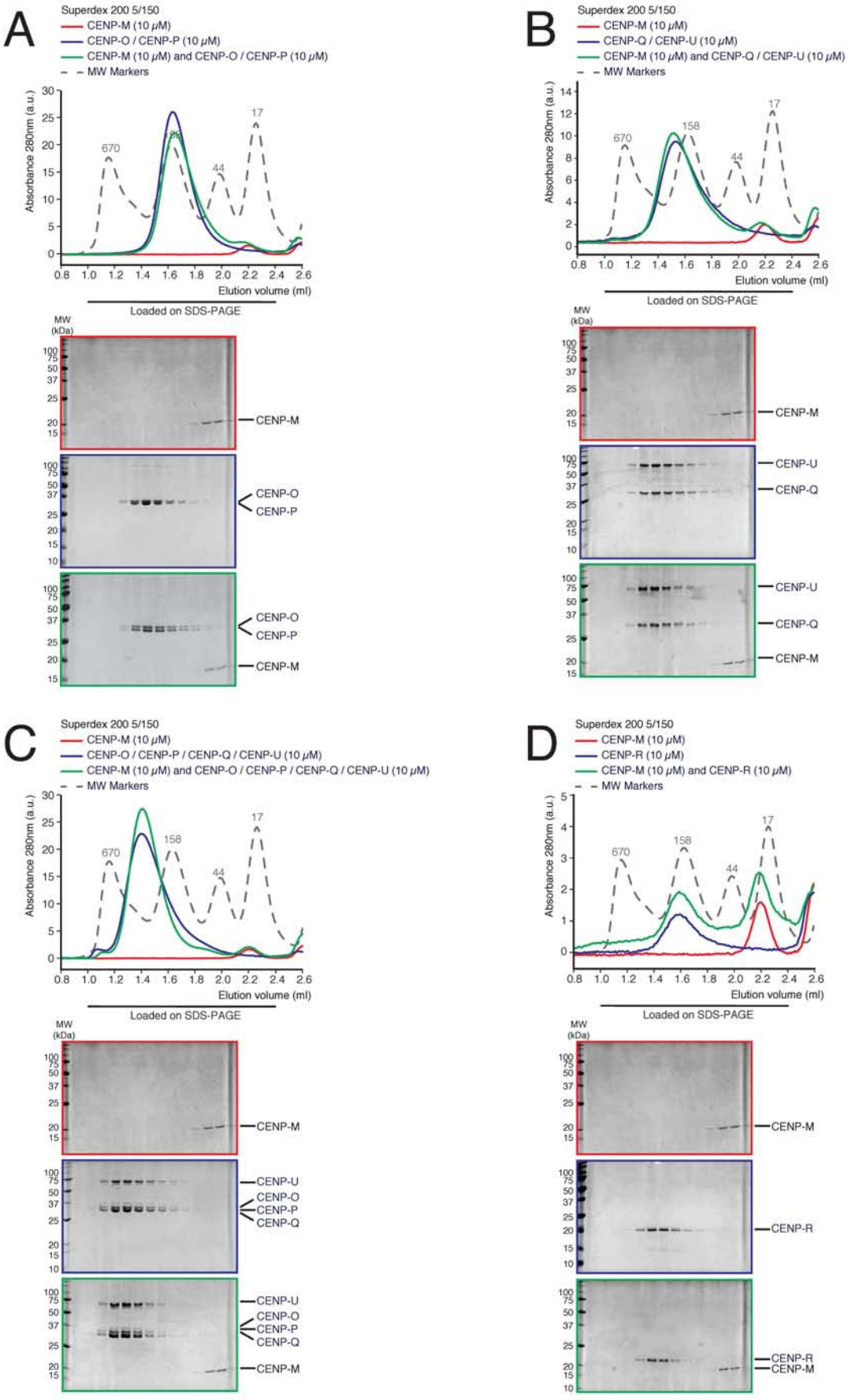


Figure 20 - CENP-M does not bind to the CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group of proteins *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of CENP-M (red), CENP-O / CENP-P complex (blue) and their

combination (green). B to D) Same experimental setting as in A, but using CENP-Q / CENP-U complex, CENP-O / CENP-P / CENP-Q / CENP-U complex or CENP-R, respectively. In all four cases, no direct binding to CENP-M occurs.

CENP-T was reported to be a member of the so-called CENP-A nucleosome associated complex (NAC)[17], together with CENP-C, CENP-H, CENP-M, CENP-N and CENP-U. Interestingly, suppression of CENP-T by siRNA resulted in the loss of CENP-M at centromeres. Also, CENP-S was initially identified in CENP-M and CENP-U affinity purifications and was thus considered part of the CENP-A-nucleosome distal (CAD) complex[17].

As already described, CENP-T forms a complex with CENP-W[77]. In addition, CENP-T / CENP-W complex associates with CENP-S / CENP-X complex[109] to form a stable CENP-T / CENP-W / CENP-S / CENP-X assembly[111]. I was able to produce and purify both recombinant CENP-T / CENP-W and CENP-S / CENP-X complexes. I then assessed their binding to CENP-M *in vitro*, but analytical SEC migration shift assays did not reveal any direct interaction with CENP-M (Figure 21 A and B). Also, when incubated all together, CENP-T / CENP-W and CENP-S / CENP-X associate in a quaternary complex, as expected, while CENP-M does not become incorporated (Figure 21 C). Cdk1 phosphorylation of CENP-T, particularly of its N-terminal tail, has been shown to be crucial for kinetochore assembly in human cells, at both endogenous and ectopic loci[98,114]. Therefore, I also tested if recombinant CENP-T / CENP-W complex phosphorylated by Cdk1 would bind to CENP-M *in vitro*, but this was not the case (Figure 21 D).

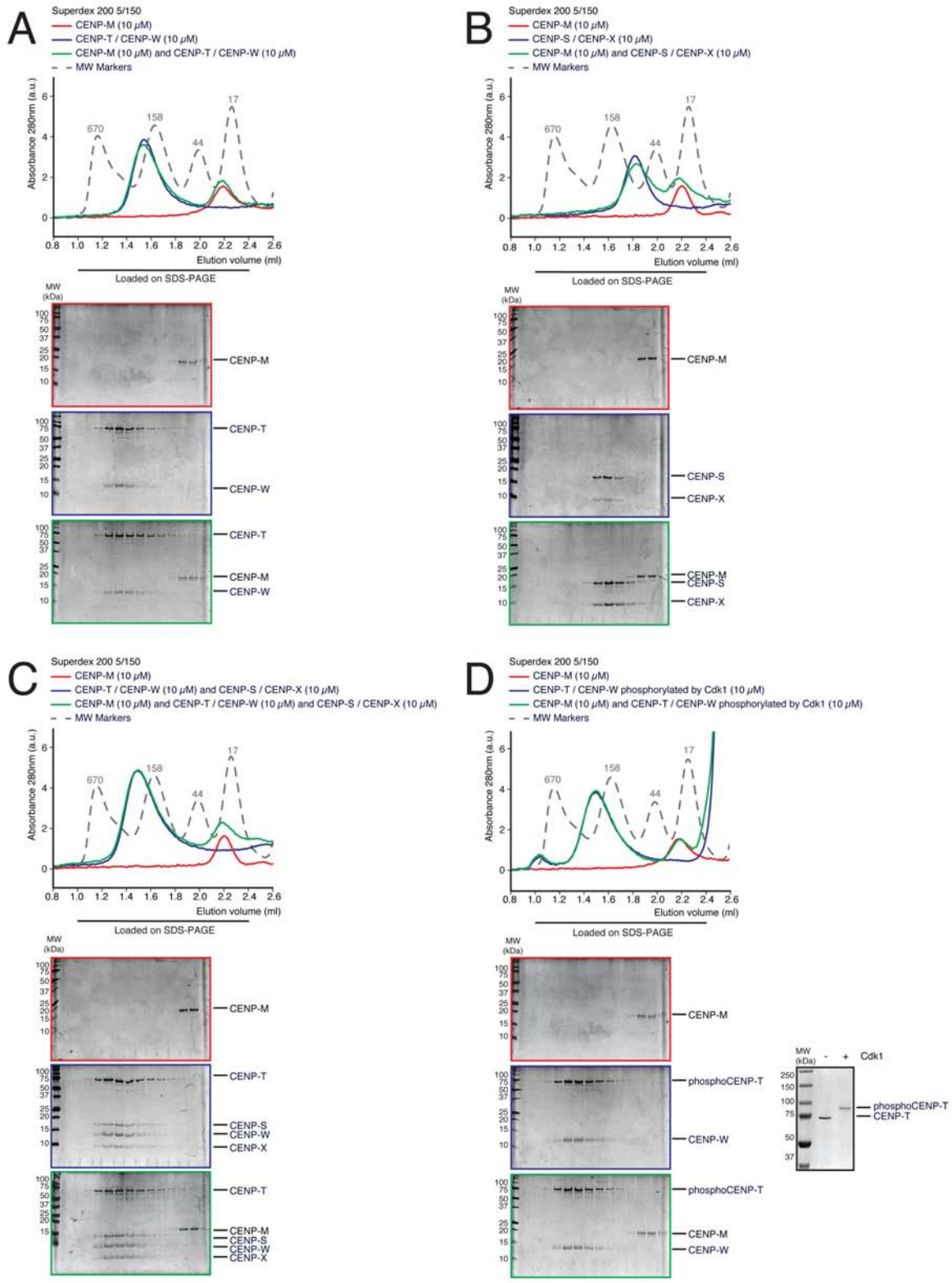


Figure 21 - CENP-M does not bind to the CENP-T / CENP-W / CENP-S / CENP-X group of proteins *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of CENP-M (red), CENP-T / CENP-W complex (blue) and their combination (green). B) Same experimental setting as in A, but using CENP-S / CENP-X complex. In both cases, no direct binding to CENP-M occurs. C) Upon incubation of CENP-M, CENP-T / CENP-W complex and CENP-S / CENP-X complex all together, CENP-T / CENP-W / CENP-S / CENP-X assemble into a complex, as reported in the literature[111], while CENP-M is not incorporated. D) Even when phosphorylated by Cdk1, CENP-T / CENP-W complex does not interact with CENP-M. On the right side of the panel, samples of CENP-T / CENP-W complex before and after phosphorylation by Cdk1 are

analysed by SDS-PAGE. The increase in molecular weight of CENP-T after the phosphorylation reaction indicates that this worked properly and CENP-T became phosphorylated, as expected.

Depletion of CENP-M by siRNA results in reduced levels of Ndc80 at kinetochores, suggesting that CENP-M also affects the assembly of the outer kinetochore[18]. This effect is probably indirect, as CENP-M does not exhibit any binding to purified recombinant Mis12 complex, Ndc80 complex, Knl1$^{2000-2311}$ (the longest Knl1 fragment available in the laboratory) and Zwint *in vitro* (Figure 22 A to D). Also, the incubation of all these proteins together yields the formation of a single complex comprising the so-called KMN network, while CENP-M is not incorporated (data not shown).

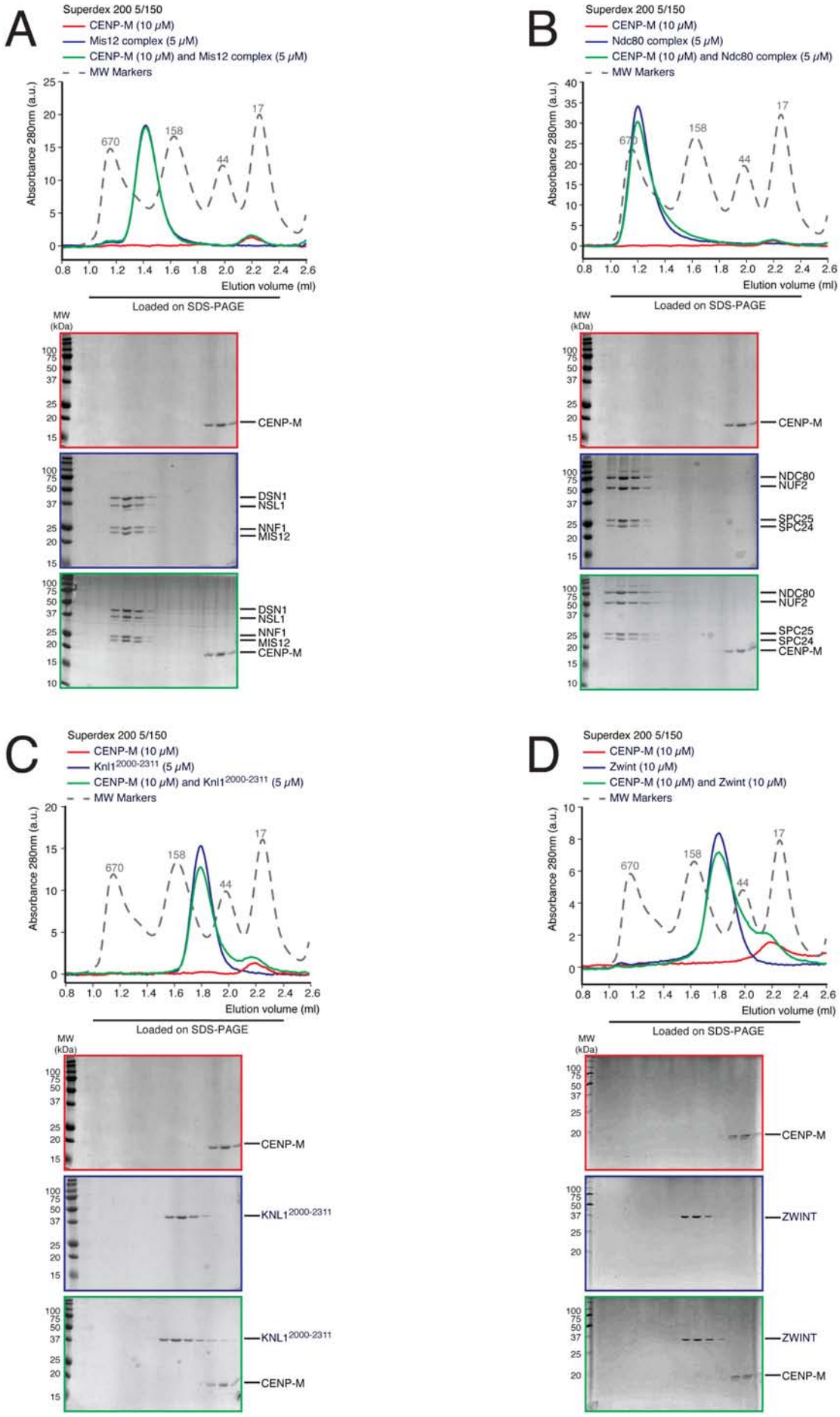


Figure 22 - CENP-M does not bind to Mis12 complex, Ndc80 complex, Knl1$^{2000-2311}$ and Zwint *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of CENP-M (red), Mis12 complex (blue) and their combination (green). B to D) Same experimental setting as in A, but using Ndc80 complex, Knl1$^{2000-2311}$ or Zwint. In all cases, no direct binding to CENP-M occurs.

For the sake of completeness, I also assessed the ability of CENP-M to interact with microtubules *in vitro* in a microtubules cosedimentation assay. As expected, no binding was detected (Figure 23).

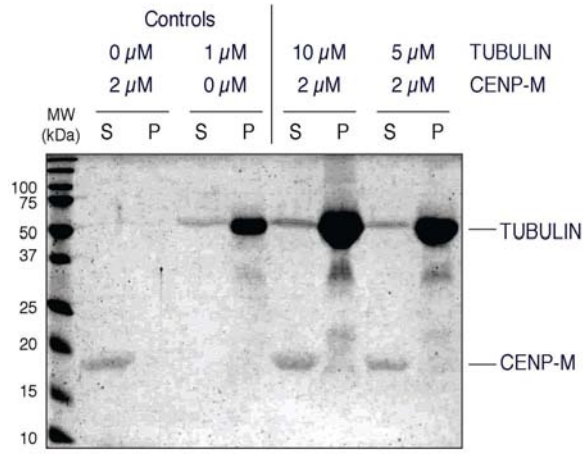


Figure 23 - CENP-M does not bind to microtubules *in vitro*.

Partition of CENP-M in supernatant (S) and pellet (P) fractions in a cosedimentation assay with polymeric tubulin. CENP-M is retained in the supernatant fraction both in the absence and in the presence of microtubules, indicating that CENP-M does not interact with microtubules *in vitro*.

2.1.11 First steps towards the validation and characterization of the interaction between CENP-M and CENP-I *in vivo*

I recently started, with the collaboration of expert cell biologists in the laboratory, to address the validation and characterization of the interaction between CENP-M and CENP-I *in vivo*. Even if only preliminary results have been obtained so far, I include them in this dissertation to provide a glimpse of what is being pursued in this direction.

The first step was the generation of an inducible GFP-CENP-M HeLa cell line. We initially characterised it by Western blotting using a commercially available monoclonal antibody against CENP-M and compared it with an inducible HeLa cell line expressing GFP only. We could verify

that endogenous CENP-M was expressed in both cell lines at comparable levels regardless of the addition of Doxycycline, the expression inducer. Conversely, the appearance of a band at the expected molecular weight of GFP-CENP-M only occurred in the GFP-CENP-M cell line upon induction with Doxycycline (Figure 24).

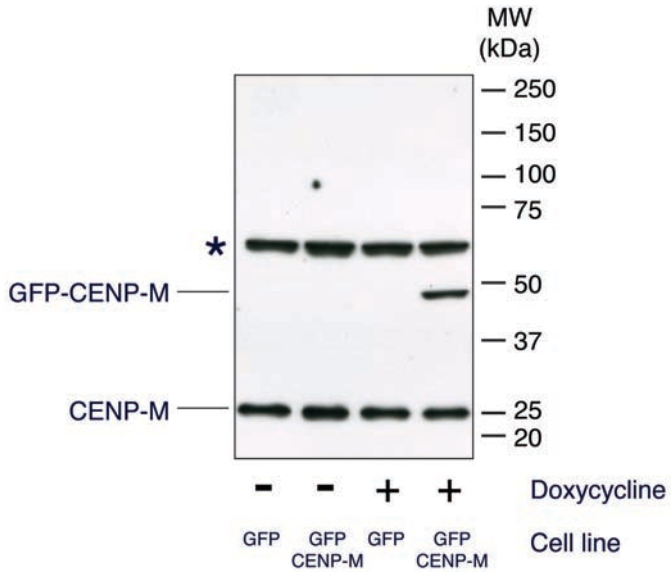


Figure 24 - Characterization of an inducible GFP-CENP-M HeLa cell line in comparison with an inducible GFP HeLa cell line by Western blotting against CENP-M.

Lower band: endogenous CENP-M. Middle band: GFP-CENP-M, only expressed in the GFP-CENP-M cell line upon induction with Doxycycline. Upper band, marked with an asterisk: unspecific band recognised by the anti CENP-M antibody.

We then verified that GFP-CENP-M localizes at centromeres in both mitosis (Figure 25 A) and interphase (Figure 25 B), as expected for a CCAN component. To observe GFP-CENP-M localization in interphase, we treated cells with RO-3306, a Cdk1 inhibitor, which arrests cells in a G2-like phase[149].

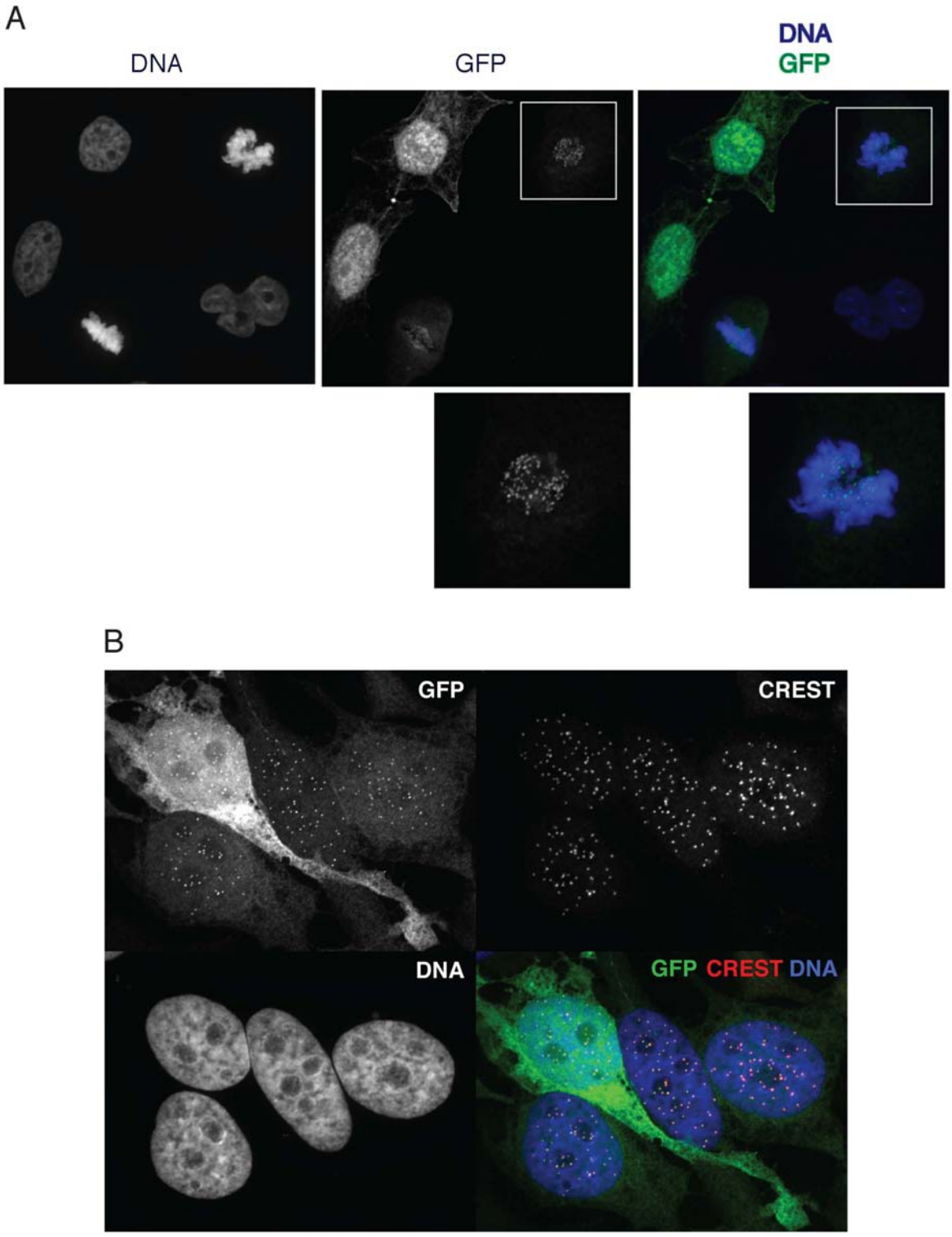


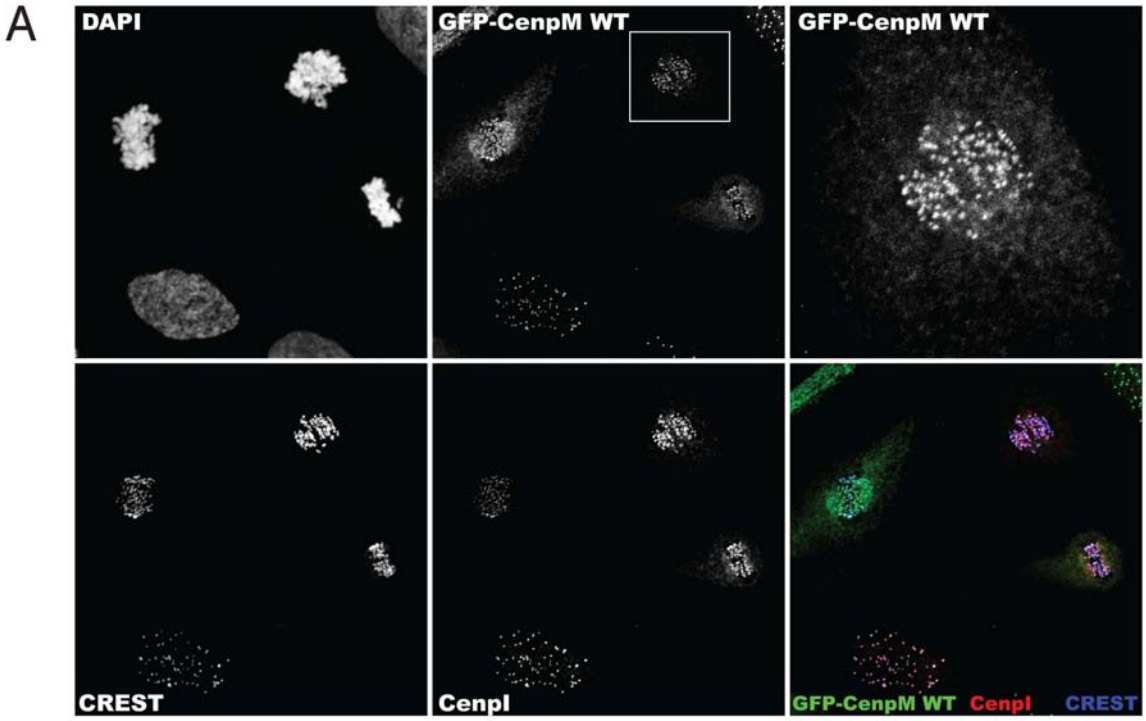
Figure 25 - GFP-CENP-M localizes at centromeres in both mitosis and interphase.

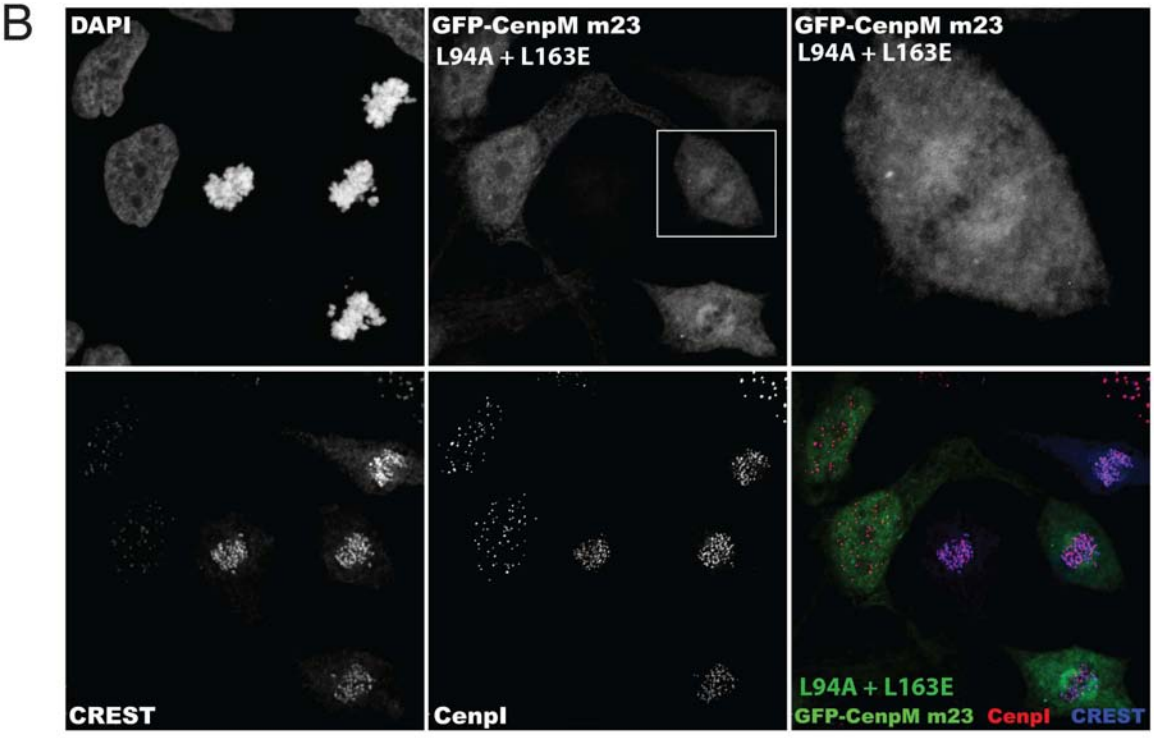
GFP-CENP-M expression was induced by the addition of Doxycycline to an inducible GFP-CENP-M HeLa cell line. Its localization was assessed in both mitotic cells (A) and cells arrested in a G2-like phase by the Cdk1 inhibitor RO-3306 (B). DAPI was employed to stain DNA and CREST was used as centromeric marker.

As a control, we also generated an inducible CENP-M-GFP HeLa cell line and verified that CENP-M displays a similar behaviour regardless of whether the GFP-tag is fused at the N- or C-terminus

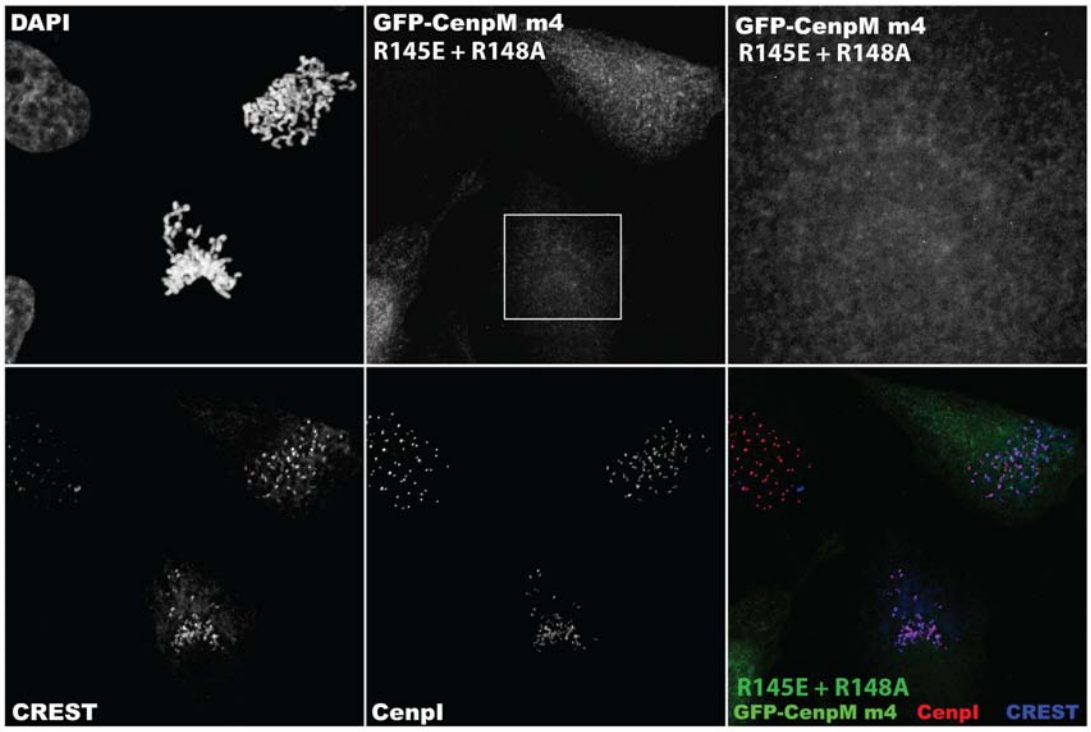
of the protein (data not shown).

We next produced inducible GFP-CENP-M HeLa cell lines for four of the CENP-M mutants that I had previously tested for their binding to CENP-I *in vitro*. In particular, we chose the two mutants that abolish CENP-M interaction with CENP-I (L94A + L163E and R145E + R148A) and two of the mutants that retain the ability to bind to CENP-I (R60A + R62A and H87A + D89A) *in vitro*. We first asked if these mutations have any effect on the centromeric localization of GFP-CENP-M. Mutants L94A + L163E (Figure 26 B) and R145E + R148A (Figure 26 C) lose the ability to localize at centromeres. We hypothesize that, as they are unable to interact with CENP-I *in vitro*, they are most likely incapable of doing so also *in vivo* and, consequently, are not recruited to the centromere, in agreement with the dependency of CENP-M on CENP-I for its kinetochore localization[18]. Interestingly, also mutant R60A + R62A hampers the localization of GFP-CENP-M at centromeres (Figure 26 D). This mutant retains the ability to interact with CENP-I *in vitro* and, therefore, we suppose that its inability to localize at centromeres *in vivo* is due to the loss of interaction with some other kinetochore component on which CENP-M depends for its recruitment. In particular, this mutant contains two amino acid substitutions in the region of CENP-M structure that corresponds to the switch II of G proteins, which is often involved in interactions with their partners[150]. Mutant H87A + D89A instead exhibits centromeric localization (Figure 26 E), as GFP-CENP-M wt (Figure 26 A). It is worth mentioning that all these experiments were carried out in the presence of endogenous CENP-M and we believe this is the reason why endogenous CENP-I could always be detected at centromeres.

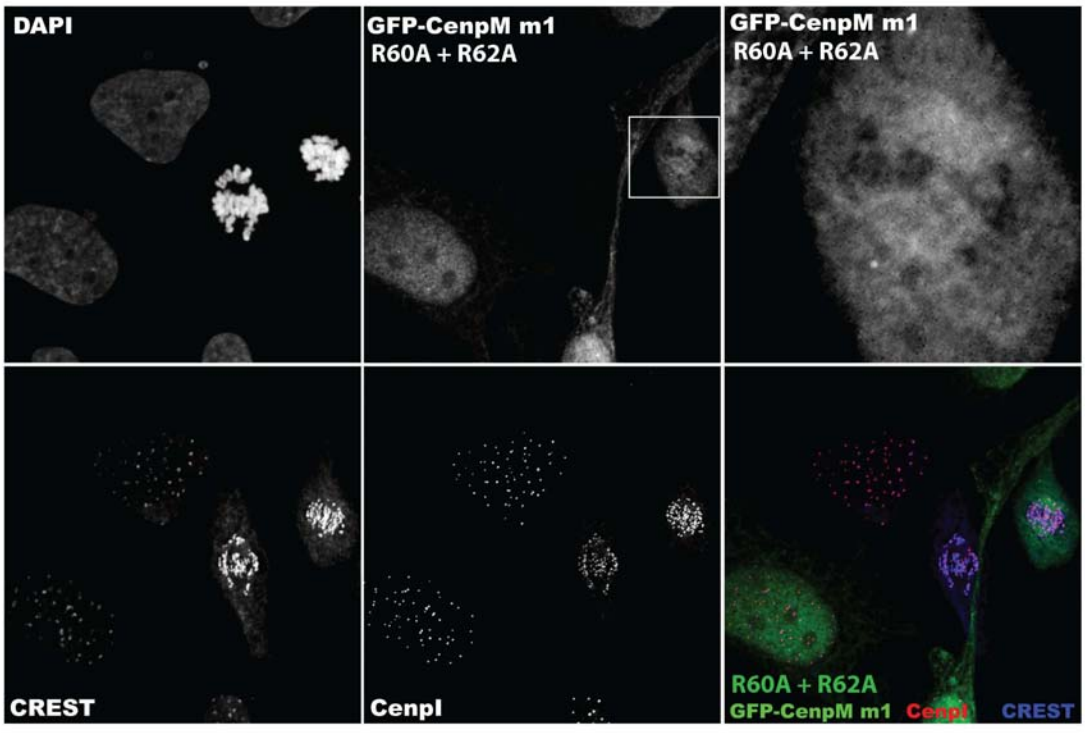

A
DAPI
GFP-CenpM WT
GFP-CenpM WT
CREST
CenpI
GFP-CenpM WT CenpI CREST


B
DAPI
GFP-CenpM m23
L94A + L163E
GFP-CenpM m23
L94A + L163E
CREST
CenpI
L94A + L163E
GFP-CenpM m23 CenpI CREST

C


DAPI
GFP-CenpM m4
R145E + R148A
GFP-CenpM m4
R145E + R148A
CREST
CenpI
R145E + R148A
GFP-CenpM m4 CenpI CREST

D


DAPI
GFP-CenpM m1
R60A + R62A
GFP-CenpM m1
R60A + R62A
CREST
CenpI
R60A + R62A
GFP-CenpM m1 CenpI CREST

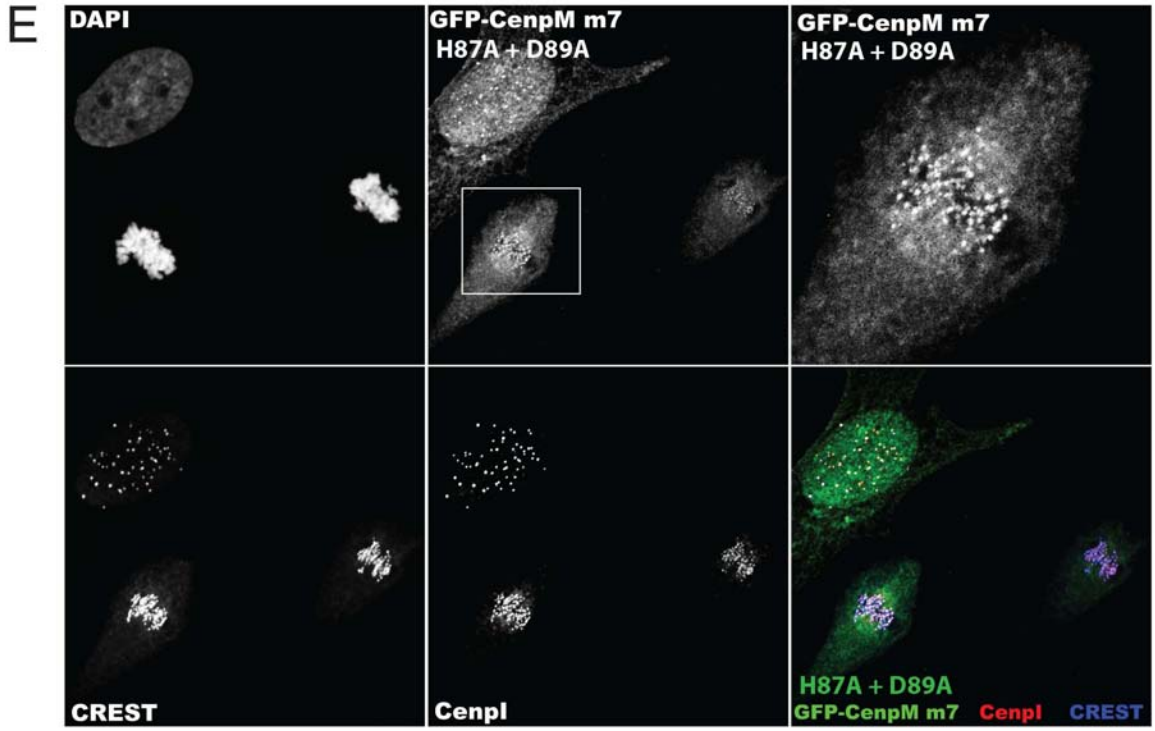


Figure 26 - Cellular localization of GFP-CENP-M wt and mutants.

The expression of GFP-CENP-M wt or mutants was induced by the addition of Doxycycline to the respective inducible HeLa cell lines; DAPI was employed to stain DNA; CREST was used as centromeric marker; endogenous CENP-I was detected using a polyclonal antibody against CENP-I. GFP-CENP-M wt (A) and mutant H87A + D89A (E) exhibit centromeric localization. On the contrary, GFP-CENP-M mutants L94A + L163E (B), R145E + R148A (C) and R60A + R62A (D) lose the ability to localize at centromeres. Endogenous CENP-I is recruited at centromeres in all these cases, most likely because of the presence of endogenous CENP-M.

As already pointed out, this section was simply meant to represent a brief description of some reagents that we have generated and of preliminary results that we have obtained to validate and characterize the interaction between CENP-M and CENP-I *in vivo*. A more detailed description of our future plans in this direction is provided in the “Discussion” chapter.

2.2 Section 2 – CENP-H / CENP-K complex

2.2.1 CENP-H / CENP-K complex interacts with CENP-C *in vitro*

As already described, during the first year of my PhD I was able to produce and purify recombinant human CENP-H / CENP-K complex. I thus sought to investigate interactions of this complex with other kinetochore components.

The formation of a quaternary complex with CENP-I and CENP-M has been discussed in detail in the previous “Results" section.

Contrasting data can be found in the literature regarding the role of CENP-T[77] and CENP-C[57,91] in the kinetochore recruitment of CENP-H and CENP-K. In order to tackle this issue, I performed analytical SEC migration shift assays with purified recombinant versions of these proteins, to assess direct interactions *in vitro*. The same experimental setting as described for CENP-M was applied.

No binding of CENP-H / CENP-K complex to CENP-T / CENP-W complex was observed (Figure 27 A). This was also the case with CENP-T / CENP-W complex phosphorylated *in vitro* by Cdk1 (Figure 27 B). The possible binding of CENP-H / CENP-K / CENP-I$^{57\text{-}C}$ / CENP-M quaternary complex to CENP-T / CENP-W complex still has to be addressed.

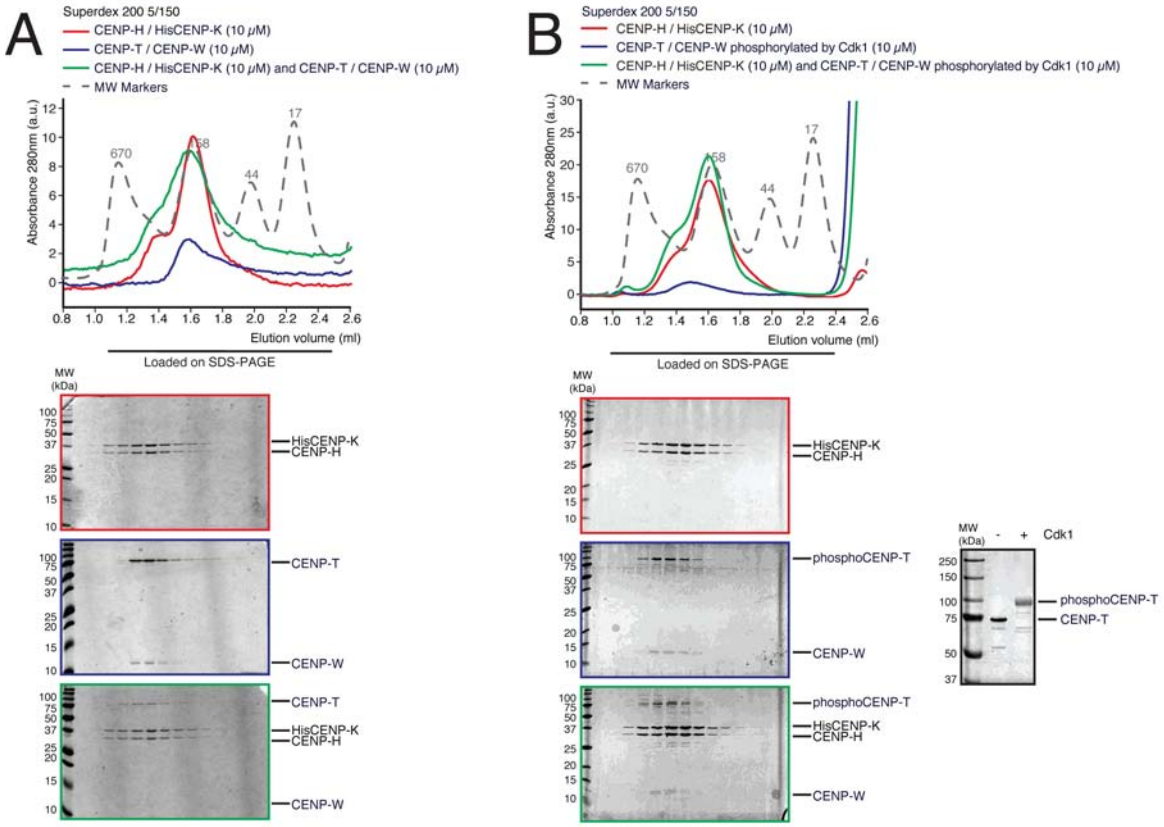


Figure 27 - CENP-H / CENP-K complex does not bind to CENP-T / CENP-W complex *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of CENP-H / HisCENP-K complex (red), CENP-T / CENP-W complex (blue) and their combination (green). B) Same experimental setting as in A, but using CENP-T / CENP-W complex phosphorylated by Cdk1. In both cases, no direct binding of CENP-H / CENP-K complex to CENP-T / CENP-W complex occurs. On the right side of the figure, samples of CENP-T / CENP-W complex before and after phosphorylation by Cdk1 are analysed by SDS-PAGE. The increase in molecular weight of CENP-T after the phosphorylation reaction indicates that this worked properly and CENP-T became phosphorylated, as expected.

I instead observed the formation of an apparently stoichiometric ternary complex upon incubation of CENP-H / CENP-K complex with CENP-C$^{1-544}$His (Figure 28).

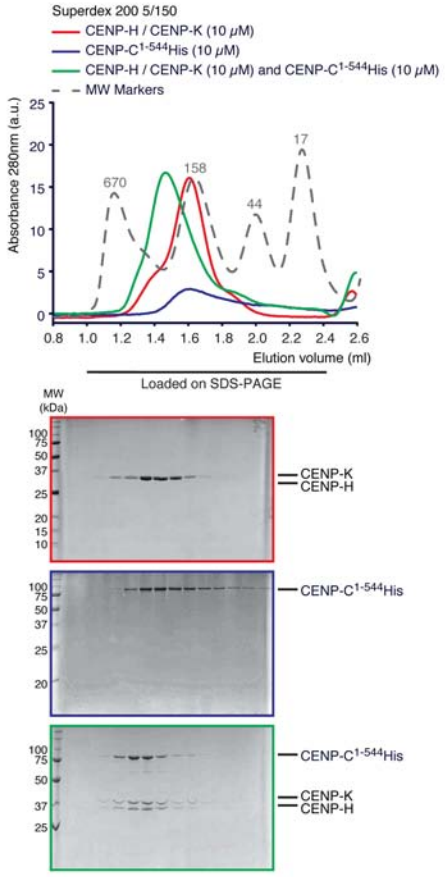


Figure 28 - CENP-H / CENP-K complex binds to CENP-C$^{1-544}$ *in vitro*.

SEC elution profiles and SDS-PAGE analyses of CENP-H / CENP-K complex (red), CENP-C$^{1-544}$His (blue) and their combination (green). The formation of an apparently stoichiometric ternary complex is observed.

After discovering that CENP-H and CENP-K are part of a quaternary complex with CENP-I^{57-C} and CENP-M, it was natural to ask if this is compatible with CENP-C binding. Indeed, the quaternary CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex binds to CENP-C$^{1-544}$His (Figure 29).

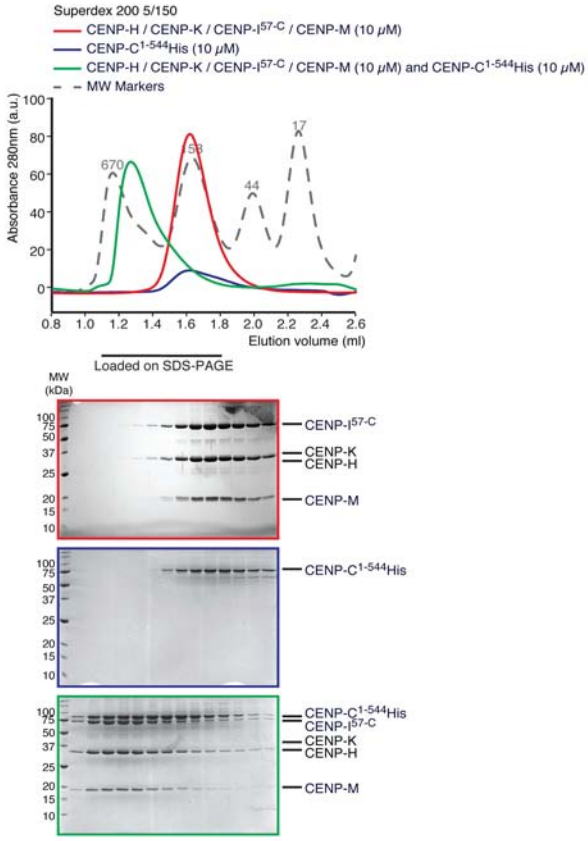


Figure 29 - CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex binds to CENP-C$^{1-544}$ *in vitro*.

SEC elution profiles and SDS-PAGE analyses of CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex (red), CENP-C$^{1-544}$His (blue) and their combination (green). A direct interaction between CENP-H / CENP-K / CENP-I^{57-C} / CENP-M complex and CENP-C$^{1-544}$His occurs.

2.2.2 Experiments of cross-linking coupled with mass spectrometry provide insights into the spatial organization of CENP-H / CENP-K / CENP-C$^{1-544}$ complex

It is worth remembering that CENP-H and CENP-K are predicted to possess an elongated structure enriched in coiled-coils, while CENP-C$^{1-544}$ is predicted to be mostly unstructured. In order to get an indication about the spatial arrangement of CENP-H / CENP-K / CENP-C$^{1-544}$ complex, an experiment of cross-linking coupled with mass spectrometry was carried out (Figure 30 A). CENP-H / CENP-K complex was also subjected to the same analysis for comparison purposes (Figure 30 B). As already mentioned, this approach provides information about the spatial proximity of lysine residues that become cross-linked in protein complexes in solution and could be performed thanks

to a collaboration with Dr. Franz Herzog's laboratory.

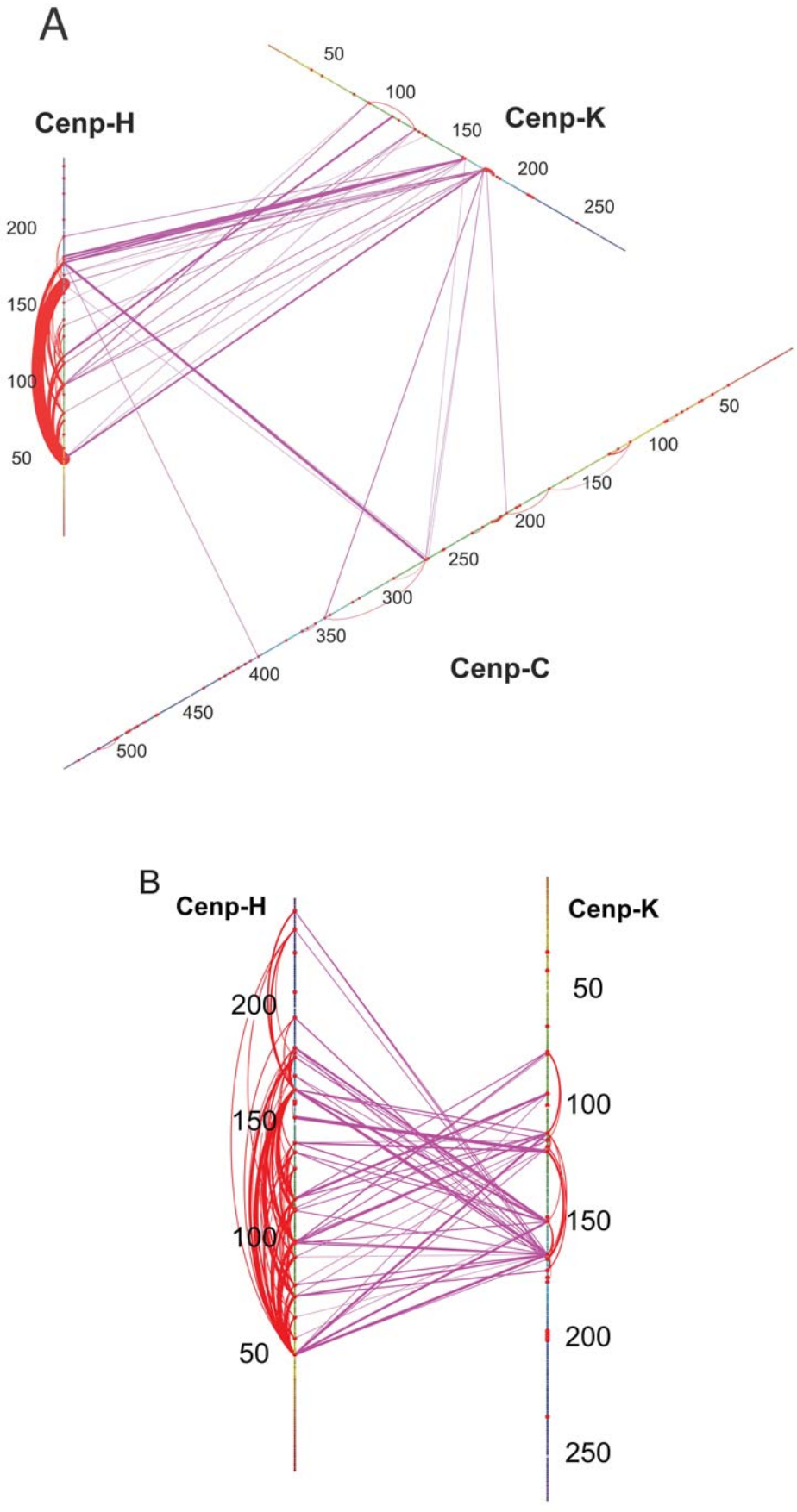


Figure 30 - Experiments of cross-linking coupled with mass spectrometry.

CENP-H / CENP-K / CENP-C^{1-544} complex (A) and CENP-H / CENP-K complex (B) were analysed by cross-linking coupled with mass spectrometry in order to obtain information about their spatial organization. Protein sequences are displayed and residues are coloured with a rainbow gradient ranging from red at the N-terminus to violet at the C-terminus. Amino acid numbers are indicated every 50 residues. Lysines are highlighted with red dots. Cross-links between lysine pairs are represented as lines, whose thickness is proportional to the identification score of the corresponding peptide pair by mass spectrometry. Intra-molecular cross-links are coloured in red, while inter-molecular cross-links are coloured in violet. The image of CENP-H / CENP-K complex is the same as shown in Figure 17 C.

Without dwelling upon details, it is apparent that CENP-H and CENP-K display a large number of inter-molecular cross-links, which involve most of their sequence length. Conversely, only few inter-molecular cross-links are observed between CENP-C^{1-544} and CENP-H and / or CENP-K. Specifically, they are all restricted to the region of CENP-C between residues ~200 - 400.

2.2.3 First steps towards the identification of CENP-C regions that are necessary and / or sufficient for binding to CENP-H / CENP-K complex

The cross-linking analysis paved the way for the design of shorter CENP-C constructs, in order to identify regions of this protein that are necessary and / or sufficient for binding to CENP-H / CENP-K complex. Specifically, a construct comprising CENP-C residues 1 - 400 proved sufficient for this interaction (Figure 31 A), while CENP-C residues 402 - 544 did not display any binding to CENP-H / CENP-K complex (Figure 31 B).

These only represent the first steps towards the characterization of the interaction between CENP-C and CENP-H / CENP-K complex. A more detailed description of our future plans in this direction is provided in the “Discussion” chapter.

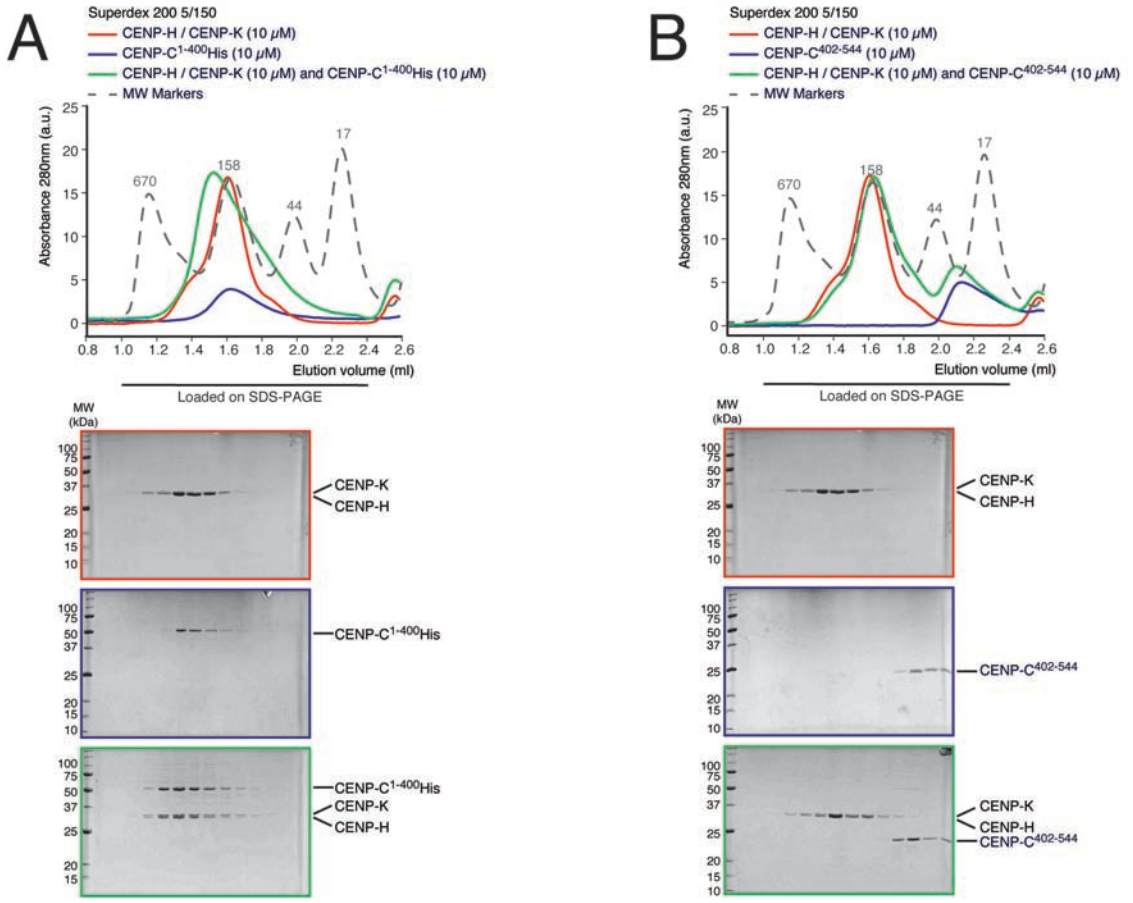


Figure 31 - CENP-H / CENP-K complex binds to CENP-C$^{1-400}$ but not to CENP-C$^{402-544}$ *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of CENP-H / CENP-K complex (red), CENP-C$^{1-400}$His (blue) and their combination (green). B) Same experimental setting as in A, but using CENP-C$^{402-544}$. CENP-C$^{1-400}$ is sufficient for binding to CENP-H / CENP-K complex, while CENP-C$^{402-544}$ is not.

2.2.4 Analysis of interactions of CENP-H / CENP-K complex *in vitro* with other kinetochore components through analytical SEC migration shift assays

The centromeric localization of members of the CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group of proteins was shown to be dependent on the CENP-H / CENP-I / CENP-K group, but not the opposite[18,100,104].

As previously mentioned, in the laboratory we accomplished the production of a quaternary CENP-O / CENP-P / CENP-Q / CENP-U complex and of CENP-O / CENP-P and CENP-Q / CENP-U sub-complexes. Also, CENP-R can be obtained singularly, but it does not become incorporated into

CENP-O / CENP-P / CENP-Q / CENP-U complex.

Therefore, I tested if CENP-H / CENP-K complex has the ability to interact with any of these samples *in vitro.* Specifically, CENP-H / CENP-K complex does not bind to CENP-O / CENP-P / CENP-Q / CENP-U complex (Figure 32 A) or to CENP-R (Figure 32 B), not even when all proteins are incubated together (data not shown). Likewise, no binding is observed when assessing interactions of CENP-H / CENP-K complex with CENP-O / CENP-P and CENP-Q / CENP-U sub-complexes separately (data not shown). Moreover, the quaternary CENP-H / CENP-K / CENP-I$^{57\text{-}C}$ / CENP-M complex also does not bind to any of these samples (data not shown).

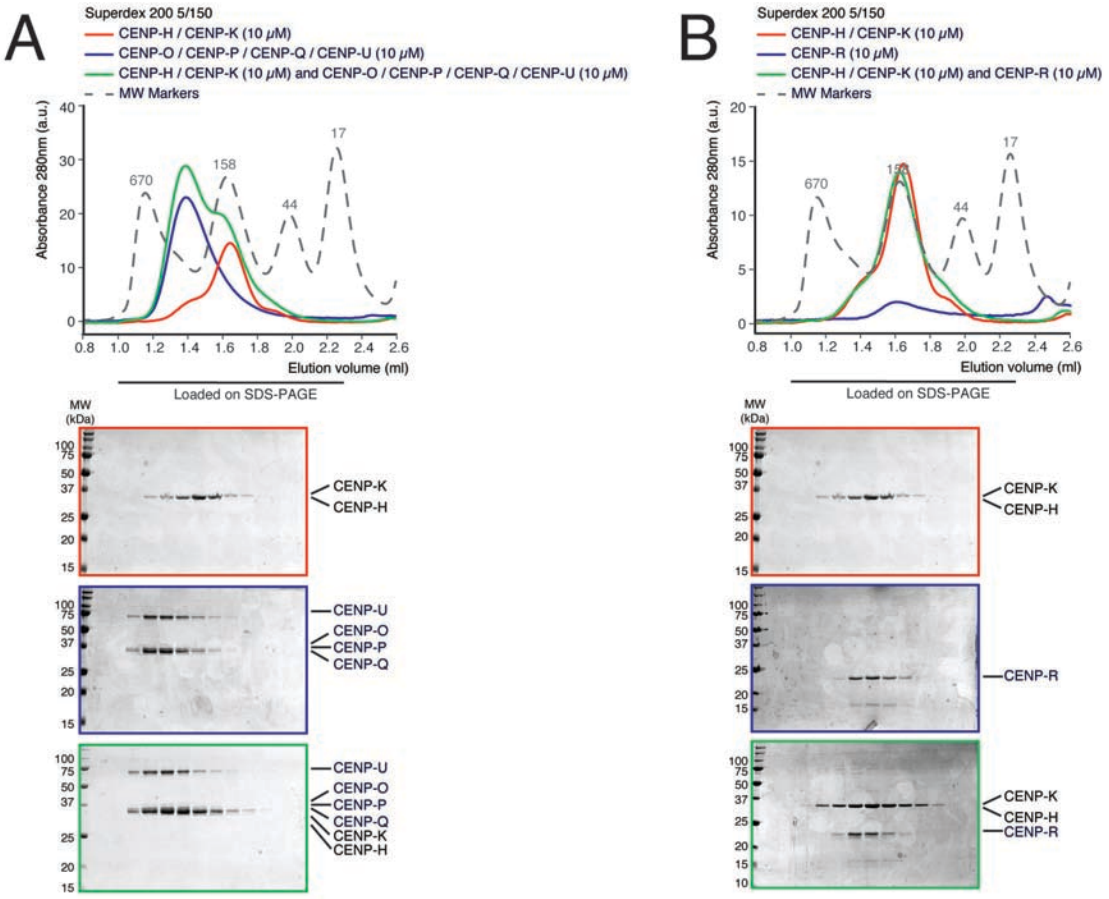


Figure 32 - CENP-H / CENP-K complex does not bind to the CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group of proteins *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of CENP-H / CENP-K complex (red), CENP-O / CENP-P / CENP-Q / CENP-U complex (blue) and their combination (green). B) Same experimental setting as in A, but using CENP-R. In both cases, no binding to CENP-H / CENP-K complex occurs.

In addition, the CENP-H / CENP-I / CENP-K group of proteins contributes to the recruitment of outer kinetochore components[18,93,104,105]. In particular, an interaction between CENP-H and Ndc80

has been suggested by yeast two-hybrid analyses and by coimmunoprecipitation experiments in chicken DT40 cells[105]. However, I did not detect any direct binding of CENP-H / CENP-K complex to Ndc80 complex *in vitro* (Figure 33). This was also the case when incubating CENP-H / CENP-K / CENP-I$^{57\text{-}C}$ / CENP-M complex with Ndc80 complex (data not shown). These observations do not exclude an indirect role of CENP-H / CENP-K / CENP-I / CENP-M complex in the kinetochore recruitment of Ndc80 complex *in vivo*, but suggest that there is no direct physical interaction between these proteins. Also, a thorough investigation of possible interactions of CENP-H / CENP-K / CENP-I$^{57\text{-}C}$ / CENP-M complex with the whole KMN network is foreseen.

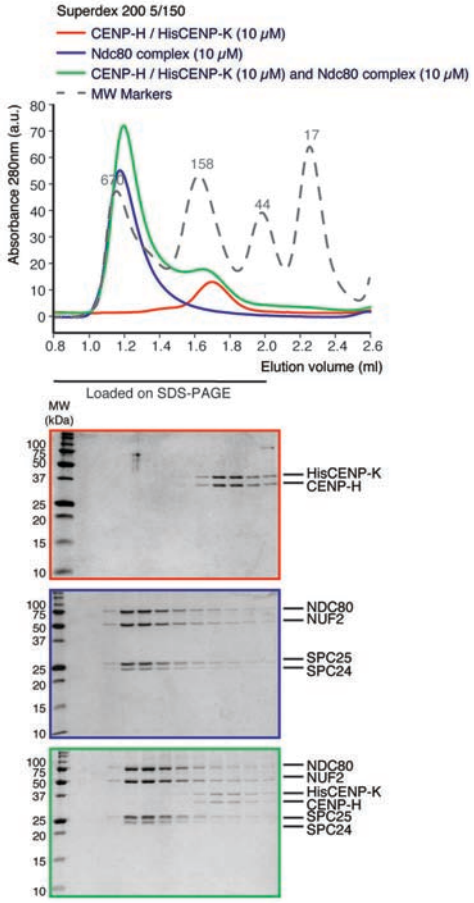


Figure 33 - CENP-H / CENP-K complex does not bind to Ndc80 complex *in vitro*.

SEC elution profiles and SDS-PAGE analyses of CENP-H / HisCENP-K complex (red), Ndc80 complex (blue) and their combination (green). No binding of CENP-H / CENP-K complex to Ndc80 complex occurs.

2.3 Section 3 – Histone H3- and CENP-A-containing nucleosomes

2.3.1 *In vitro* reconstitution of histone H3- and CENP-A-containing mononucleosomes

As already introduced, kinetochores are multiprotein complexes built on centromeres. No specific DNA sequence is strictly required for centromere specification and kinetochore assembly in most organisms. Instead, a hallmark of functional centromeres is a specialized nucleosome containing the histone H3 variant CENP-A, hinting at an epigenetic mechanism for centromere specification[30]. In particular, centromeric chromatin is characterized by the presence of CENP-A nucleosomes interspersed among H3 nucleosomes[34].

In the context of the collaborative effort of the laboratory towards the *in vitro* reconstitution of the human kinetochore from purified recombinant components, a crucial goal in the last few years has been the establishment of techniques for the recombinant expression and purification of histones and for the *in vitro* reconstitution of both H3-containing mononucleosomes (H3-MN)[151] and CENP-A-containing mononucleosomes (CA-MN)[152]. I was involved in pursuing this task together with a technical assistant and the precious collaboration of Prof. Daniela Rhodes' laboratory at the Medical Research Council in Cambridge, who helped us to get started with the reconstitution of canonical H3-MN, and Prof. Aaron F. Straight's laboratory at Stanford University, who assisted us on issues more specifically related to CA-MN.

We have recently achieved the production of both H3-MN (Figure 34 A) and CA-MN (Figure 34 B) of sufficient quality and quantity to begin to address their *in vitro* interactions with kinetochore components. To provide a glimpse of what is being pursued in this direction, I will illustrate some preliminary observations concerning a direct interaction of Mis12 complex with nucleosomes.

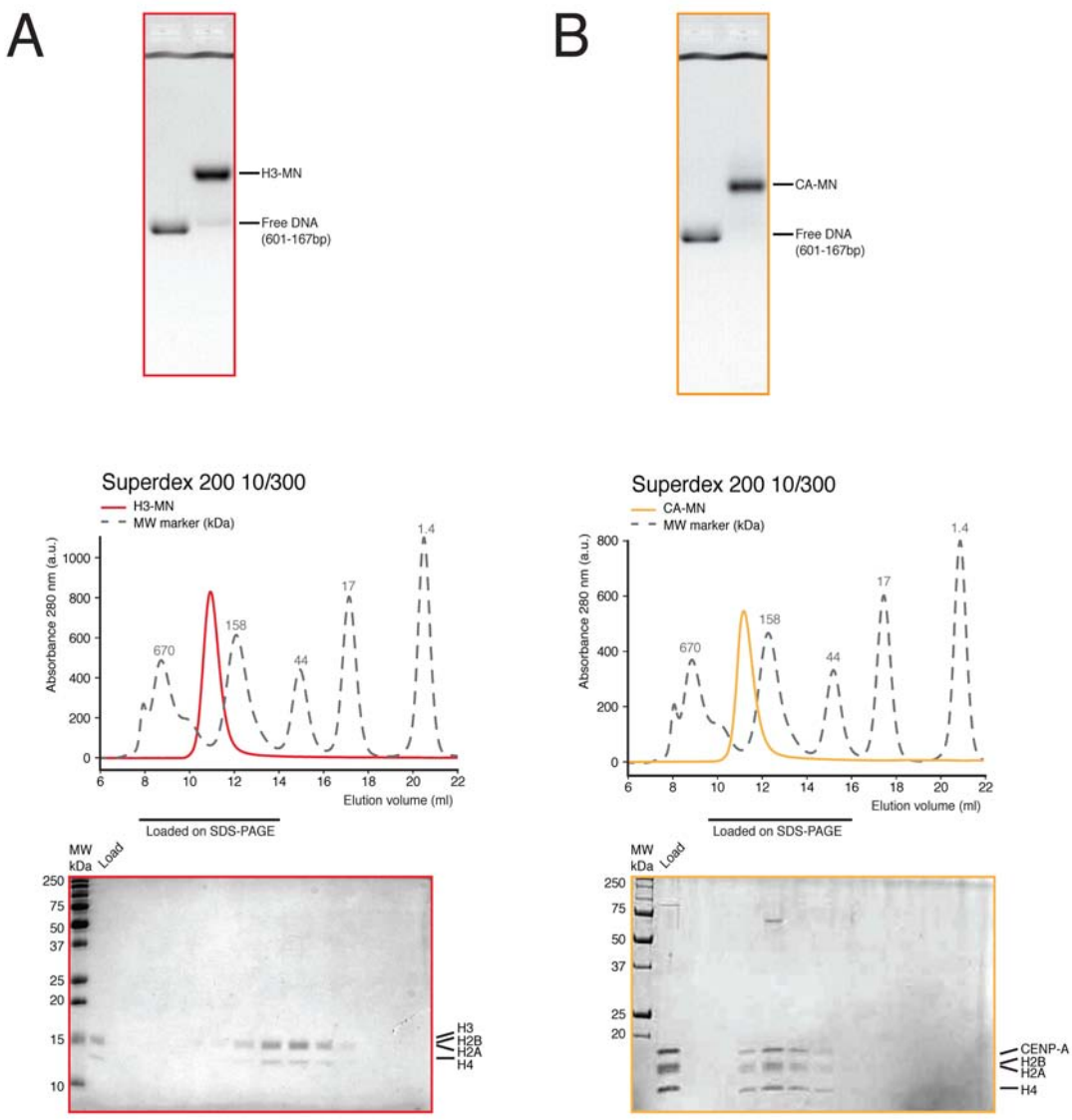


Figure 34 - Quality assessments of *in vitro* reconstituted H3-MN (A) and CA-MN (B).

Upper panels: native agarose gels with Midori green staining of free 601-167bp DNA (left lanes) and reconstituted H3-MN or CA-MN (right lanes). Both H3-MN and CA-MN migrate as a sharp band, indicating that the samples are homogeneous, and only a minor amount of free DNA is left. Lower panels: SEC elution profiles from a Superdex 200 column and SDS-PAGE analyses of reconstituted H3-MN or CA-MN. In both cases, mononucleosomes elute as a symmetric peak, again denoting the homogeneity of the samples.

2.3.2 H3 and CENP-A mononucleosomes interact with Mis12 complex *in vitro*

With the aim of identifying direct interacting partners of H3-MN and CA-MN, I recently started performing analytical SEC migration shift assays with various purified recombinant human kinetochore proteins available in the laboratory.

The formation of a larger assembly could be observed after incubation of H3-MN with Mis12

complex (Mis12 C), which comprises the Nnf1, Mis12, Dsn1 and Nsl1 subunits. In particular, Mis12 C seems to bind to H3-MN in a 2 : 1 ratio (Figure 35), in agreement with the fact that nucleosome cores possess a two-fold symmetry axis.

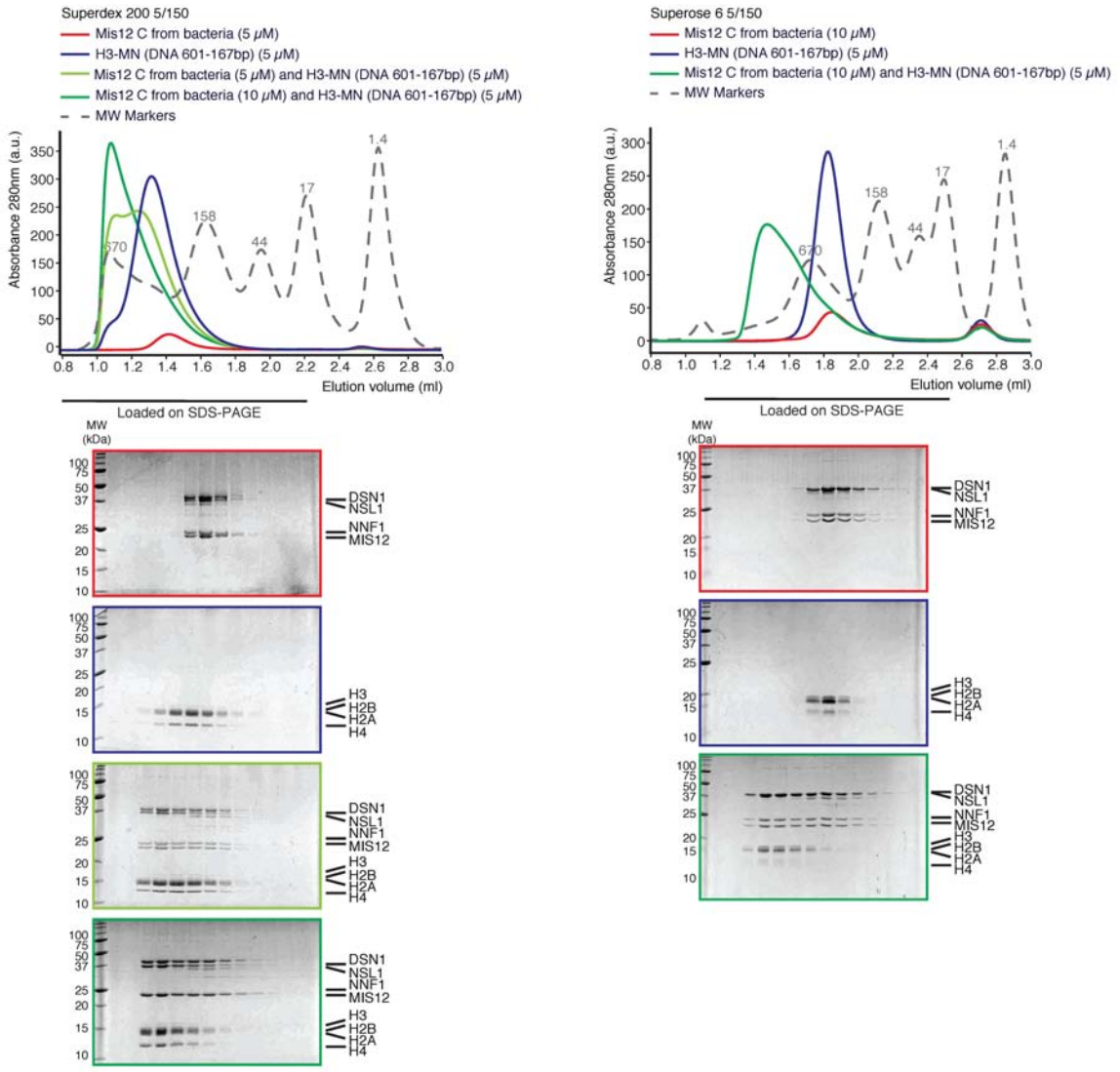


Figure 35 - Mis12 C binds to H3-MN *in vitro*.

Left panel: SEC elution profiles and SDS-PAGE analyses of Mis12 C (red), H3-MN (blue) and their combination in a 1 : 1 (light green) or 2 : 1 (dark green) ratio. The shift in the elution profile observed upon combination of Mis12 C and H3-MN indicates the formation of a larger protein assembly. In particular, a double peak is observed upon combination of Mis12 C and H3-MN in a 1 : 1 ratio (light green), while a complete shift occurs upon their combination in a 2 : 1 ratio (dark green), suggesting that two Mis12 C can simultaneously bind to one H3-MN, in agreement with the two-fold symmetry of nucleosome cores. This analysis was performed on a Superdex 200 5/150 column. Right panel: the same experiment using Mis12 C (red), H3-MN (blue) and their combination in a 2 : 1 ratio (dark green) was also run on a Superose 6 5/150 column, which then became the column of choice for analytical SEC migration shift assays involving mononucleosomes, given that it displays a better resolution at high molecular weights with respect to a Superdex 200 5/150 column.

Likewise, incubation of Mis12 C with CA-MN in a 2 : 1 ratio resulted in the formation of a complex (Figure 36).

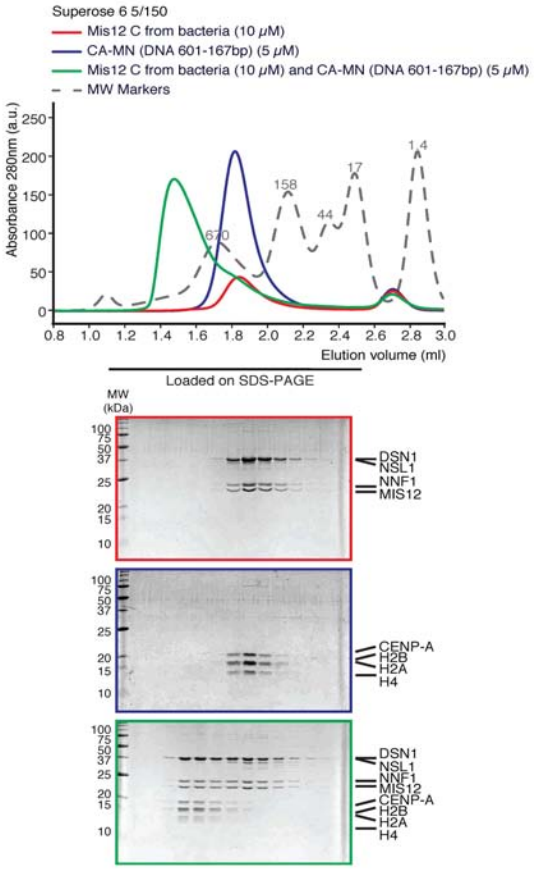


Figure 36 - Mis12 C binds to CA-MN *in vitro*.

SEC elution profiles and SDS-PAGE analyses of Mis12 C (red), CA-MN (blue) and their combination in a 2 : 1 ratio (green). The shift in the elution profile observed upon combination of Mis12 C and CA-MN indicates the formation of a larger protein assembly.

Recombinant Mis12 C was purified from both bacteria and insect cells. I carried out the above experiments using Mis12 C produced in bacteria. I then tested Mis12 C produced in insect cells, to verify if the post-translational modifications added by a eukaryotic system such as insect cells influence the ability of Mis12 C to interact with nucleosomes *in vitro*. I obtained comparable results, suggesting that the observed binding is not appreciably regulated by post-translational modifications of Mis12 C, or at least by those introduced in insect cells (Figure 37 A and B).

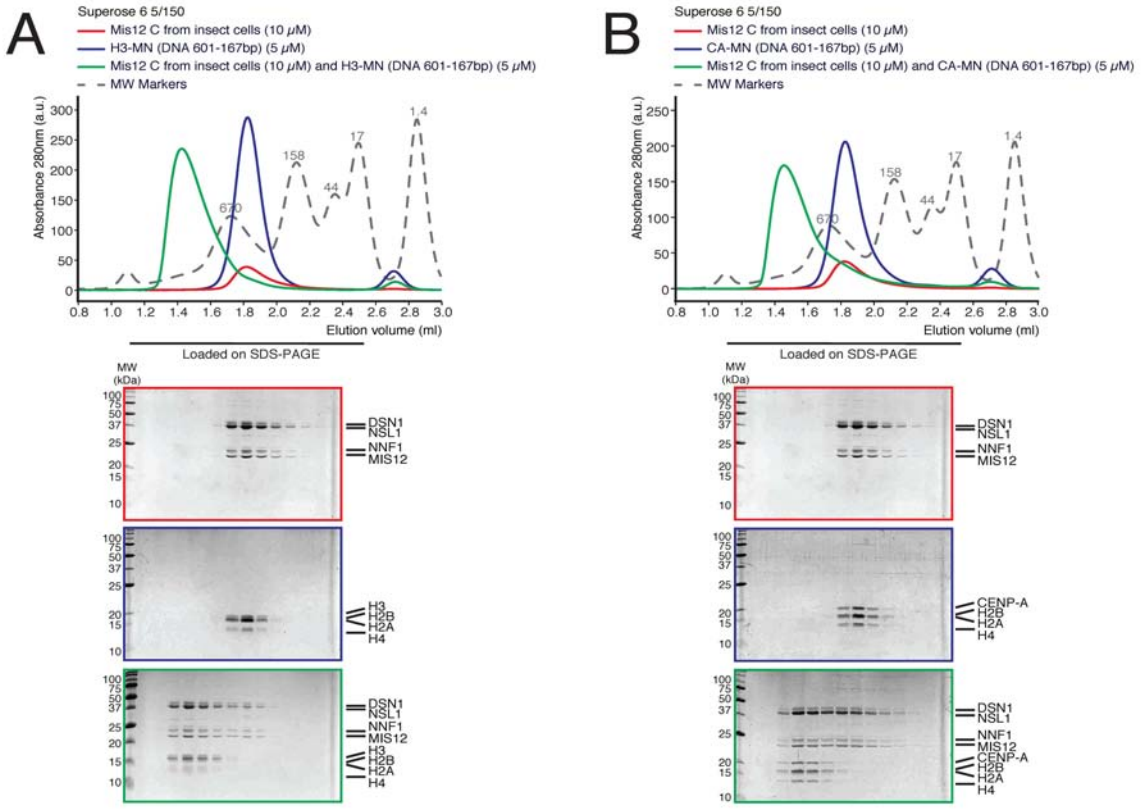


Figure 37 - Mis12 C produced in insect cells binds to H3-MN and CA-MN *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of Mis12 C produced in insect cells (red), H3-MN (blue) and their combination in a 2 : 1 ratio (green). B) Same experimental setting as in A, but using CA-MN. In both cases, a direct interaction between Mis12 C and mononucleosomes is observed. Samples of purified recombinant Mis12 C produced either in bacteria or in insect cells display a similar behaviour in terms of nucleosome binding.

2.3.3 The C-terminal region of Nsl1 is necessary for nucleosome binding *in vitro*

It is known from the literature that the C-terminal region of the Nsl1 subunit of Mis12 C is unstable in the absence of binding partners and undergoes spontaneous proteolysis during purification and subsequent storage[20]. I noticed that, in the context of my SEC migration shift assays, the portion of Mis12 C displaying Nsl1 degradation was not shifting to a lower retention volume, indicating that it was not forming a complex with mononucleosomes; on the contrary, Mis12 C engaged in interactions with mononucleosomes showed a stabilized Nsl1 C-terminal tail. Thus, I asked if this region of Nsl1 is involved in nucleosome binding. Indeed, a version of recombinant Mis12 C containing a proteolitically stable segment of Nsl1, consisting of residues 1 - 258 ($Nsl1^{1-258}$) and

lacking 23 residues at its C-terminus, loses the ability to bind to nucleosomes, revealing that the C-terminal tail of Nsl1 is necessary for nucleosome binding (Figure 38 A and B).

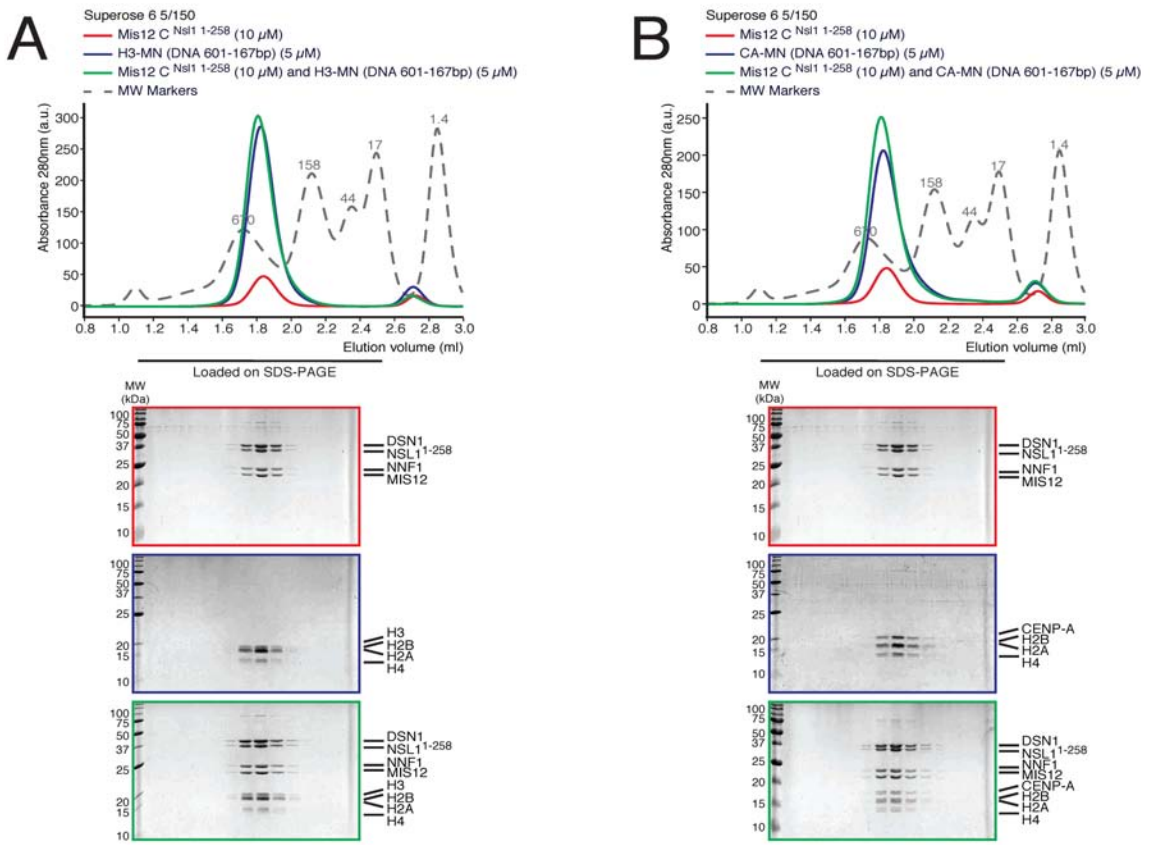


Figure 38 - A Mis12 C construct lacking the C-terminal region of Nsl1 does not bind to H3-MN and CA-MN *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of Mis12 C containing Nsl1$^{1-258}$ (red), H3-MN (blue) and their combination in a 2 : 1 ratio (green). B) Same experimental setting as in A, but using CA-MN. In both cases, this truncated Mis12 C construct loses the ability to bind to nucleosomes, revealing that the C-terminal tail of Nsl1 is necessary for nucleosome binding.

2.3.4 The C-terminal region of Nsl1 is sufficient for nucleosome binding *in vitro*

I then sought to verify if this region of Nsl1 is also sufficient for nucleosome binding. For this purpose, I employed a construct encompassing Nsl1 residues 250 - 281 as a C-terminal fusion to GST. This proved sufficient for nucleosome binding (Figure 39 A and B). In order to check that the observed binding was mediated by the Nsl1 tail and not, unspecifically, by GST, I also tested a synthetic peptide comprising Nsl1 residues 256 - 281 and obtained comparable results (data not

shown).

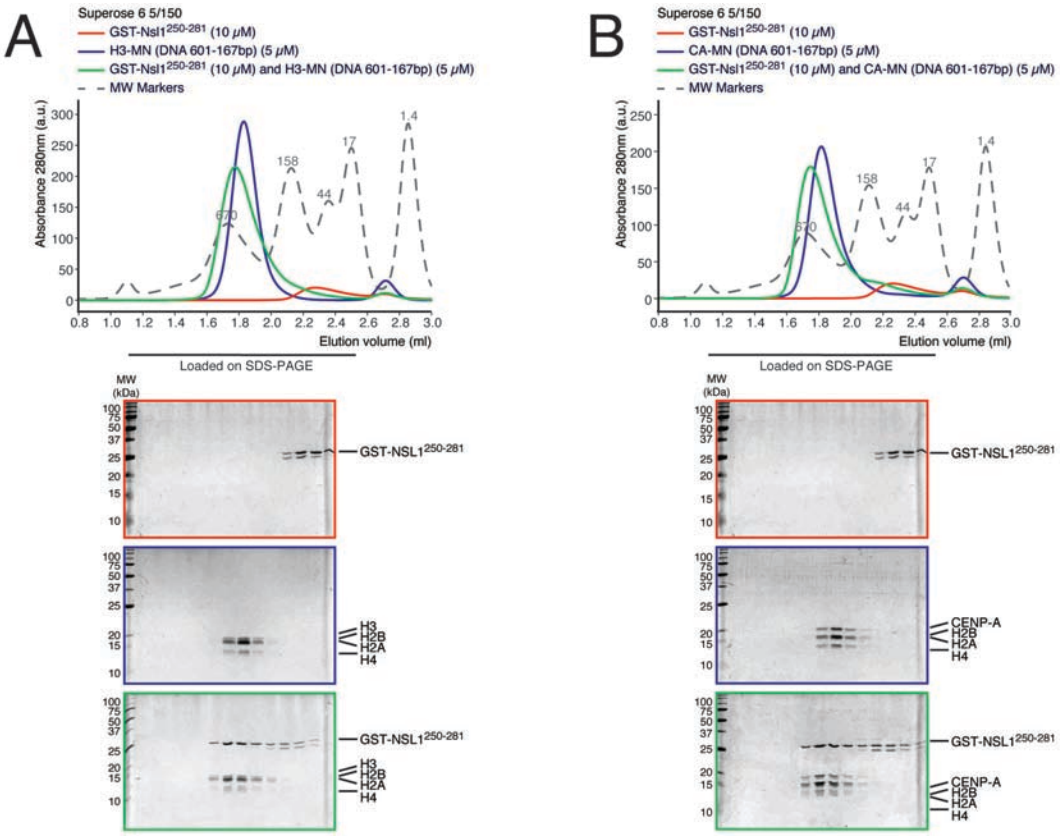


Figure 39 - The C-terminal region of Nsl1 binds to H3-MN and CA-MN *in vitro*.

A) SEC elution profiles and SDS-PAGE analyses of GST-Nsl1$^{250-281}$ (red), H3-MN (blue) and their combination in a 2 : 1 ratio (green). B) Same experimental setting as in A, but using CA-MN. In both cases, the C-terminal tail of Nsl1 retains the ability to bind to nucleosomes, revealing that this region of Nsl1 is sufficient for nucleosome binding.

2.3.5 The binding of Mis12 complex to nucleosomes is sensitive to salt

The analytical SEC migration shift assays presented above were carried out in a buffer containing 10 mM Hepes pH 7.5, 50 mM NaCl and 1 mM TCEP. Such a low ionic strength is the characteristic end point of salt gradient dialysis methods for *in vitro* reconstitution of nucleosomes.

In order to assess the salt sensitivity of the interaction between Mis12 C and mononucleosomes, I repeated the assay at progressively increasing NaCl concentrations. Specifically, the binding of Mis12 C to mononucleosomes still takes place in 100 mM NaCl (Figure 40 A), while it is not observed in 200 mM NaCl (Figure 40 B), suggesting that electrostatic interactions might be important. Interestingly, the SEC elution profile of mononucleosomes varies in different NaCl

concentrations, in agreement with previous reports of salt-dependent changes in nucleosome conformation[153].

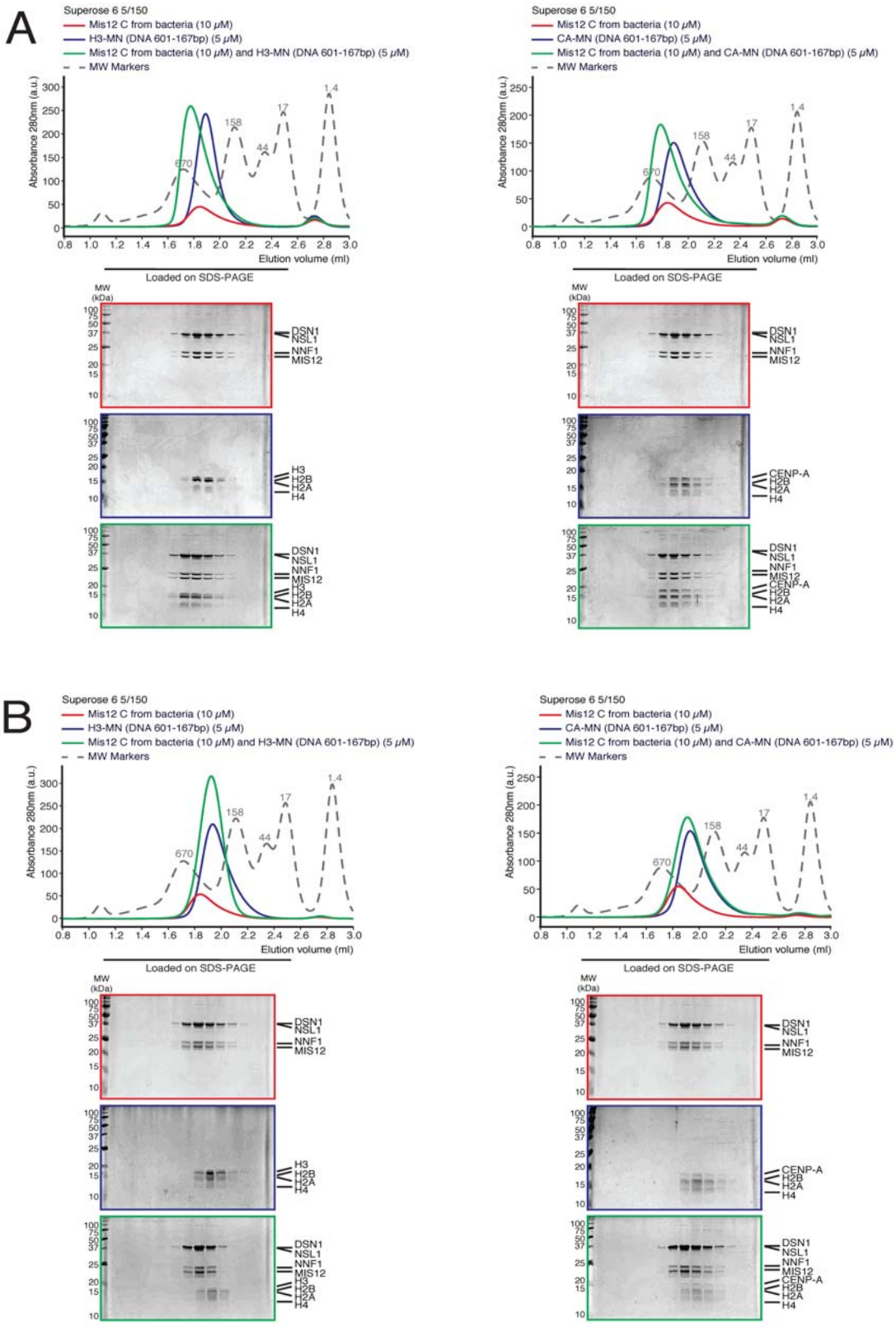


Figure 40 - The interaction of Mis12 C with H3-MN and CA-MN *in vitro* is sensitive to salt.

A) SEC elution profiles and SDS-PAGE analyses of Mis12 C (red), mononucleosomes (blue) and their combination in a 2 : 1

ratio (green). H3-MN are displayed in the left panel and CA-MN in the right panel. A buffer containing 10 mM Hepes pH 7.5, 100 mM NaCl and 1 mM TCEP was employed. B) Same experimental setting as in A, but with a buffer containing 10 mM Hepes pH 7.5, 200 mM NaCl and 1 mM TCEP. The binding of Mis12 C to mononucleosomes still occurs in 100 mM NaCl, while it is not observed in 200 mM NaCl.

2.3.6 The binding of Mis12 complex to nucleosomes is mediated by contacts with the nucleosome core particle

As already detailed in the "Introduction" chapter, 146 bp of DNA are visible in the crystal structure of the canonical H3 nucleosome, while only the central 121 bp are visible in the CENP-A nucleosome structure and 13 bp from both ends of the DNA are disordered. This indicates that the DNA regions at the entrance and exit of the CENP-A nucleosome lack a fixed conformation.

Unlike sequence-specific DNA-binding proteins, histones seem to have evolved to largely minimize DNA sequence discrimination, a property that allows them to be distributed all over the genome. Nevertheless, sequence-dependent attributes influence the interaction of DNA with the histone octamer in terms of curvature, flexibility and affinity[154]. Together with other determinants, like for example chromatin composition and chromatin remodelling factors, DNA sequence regulates nucleosome dynamics and positioning *in vivo,* imparting sophisticated regulatory features that are critical for genomic functions.

Various DNA constructs, differing with respect to length and sequence composition, have classically been employed for *in vitro* reconstitution of nucleosomes. We decided to reconstitute nucleosomes with the widely used 601 DNA sequence. This was identified in a SELEX experiment designed to find a DNA sequence able to bind to the histone octamer with high affinity and with a unique position[155], thus yielding a homogeneous population of reconstituted nucleosomes. In particular, all the experiments that I presented above were carried out with mononucleosomes reconstituted *in vitro* using the 601 - 167 bp DNA, thus possessing some "linker" DNA extending from both ends of the minimal positioning sequence. In addition, I repeated the SEC migration shift assays described above employing mononucleosomes reconstituted *in vitro* using the 601 - 145 bp DNA sequence[154]. I obtained comparable results (data not shown), suggesting that the observed binding of Mis12 C to nucleosomes is mediated by contacts with the nucleosome core particle.

3 Discussion

In this study, I have presented a biochemical and structural characterization of the CCAN protein CENP-M and I have disclosed its unprecedented role in the context of a quaternary complex that also includes CENP-H, CENP-K and CENP-I. Specifically, the crystal structure of CENP-M has revealed that it possesses a G protein-like fold. However, CENP-M does not display the enzymatic activity of a G protein. This can already be inferred from CENP-M sequence[22], which lacks all of the five characteristic motifs of the G domain. Indeed, I could also prove with biochemical experiments that CENP-M is unable to bind to guanine and adenine nucleotides *in vitro.*

In this respect, after performing a multiple sequence alignment of CENP-M orthologs (Figure 41), it was interesting to notice that the most distant CENP-M orthologs that could be found (namely, in *Nematostella vectensis*, *Capsaspora owczarzaki* and *Mucor circinelloides*) seem to retain the P-loop motif, while lacking the other four functionally relevant motifs of the G domain. This observation raises the interesting hypothesis that CENP-M sequence might have progressively lost the determinants of the G domain during evolution. Reasoning along these lines and considering that no CENP-M ortholog has been so far identified in yeast through sequence homology searches[25], it is intriguing to postulate that an actual G protein might fulfil CENP-M function in yeast, thus representing a case of structural and functional conservation, rather than of sequence conservation. Further investigations are clearly needed to address this possibility. In particular, in order to narrow down the list of potential candidates, I envision that bioinformatic analyses will prove valuable to understand which are the closest G protein relatives of CENP-M. Complementary information would be possibly provided by genetic, proteomic and biochemical approaches aiming at highlighting any yeast G protein that appears to be associated with kinetochore biology.

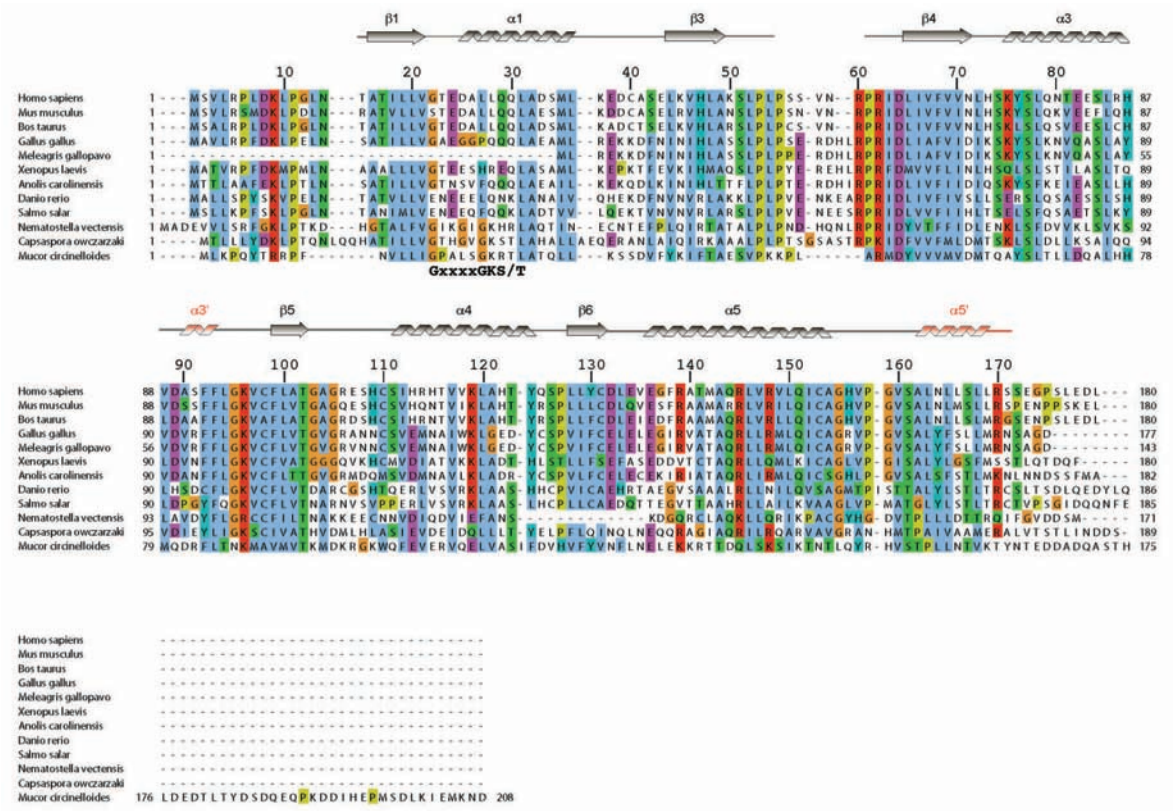


Figure 41 - Multiple sequence alignment of CENP-M orthologs.

The MUSCLE server[146] was employed to perform the alignment. Default Clustal colour scheme is applied. Residue numbers above the alignment refer to the human CENP-M sequence. Secondary structure elements are illustrated. The P-loop motif of the G domain is indicated below the alignment.

This study also represents a successful case of “from structure to function” workflow, where the determination of CENP-M structure, which possesses a G protein-like fold, together with the prediction of CENP-I structure, which displays an Importin-β-like fold, paved the way for understanding CENP-M role as a direct interacting partner of CENP-I. In particular, my results speak in favour of a model where CENP-M is embraced by CENP-I, similarly to how Ran is encircled by Importin-β, and acts as a chaperone-like protein for CENP-I, stabilizing its conformation. In addition, I speculate that CENP-H and CENP-K, both predicted to have an elongated structure enriched in coiled-coils, might lie on a region comprising about 350 amino acids at the N-terminus of CENP-I (Figure 42).

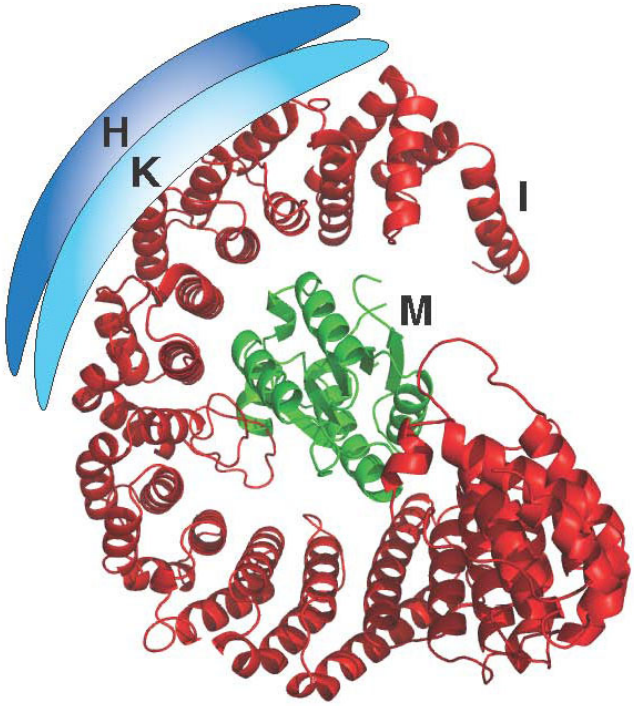


Figure 42 - Model of the structural organization of CENP-H / CENP-K / CENP-I / CENP-M complex.

The I-TASSER model of CENP-I structure (red) and the crystal structure of CENP-M (green) are displayed as cartoon representations. The position of CENP-I and CENP-M with respect to each other was modelled by superposition on the structure of Importin-β in complex with Ran (PDB ID: 1IBR). CENP-H (blue) and CENP-K (light blue) are depicted as having an elongated structure and lying on the N-terminal region of CENP-I.

This tentative model of the structural organization of CENP-H / CENP-K / CENP-I / CENP-M complex takes into account the information that I could derive from the behaviour of different protein constructs, the results of analytical SEC migration shift assays and GST-pull-downs, and the experiments of cross-linking coupled with mass spectrometry.

Indeed, CENP-I$^{57-281}$ is the only CENP-I construct that could be obtained in a soluble and well-behaved fashion in the absence of any binding partners, in agreement with the fact that it is predicted to comprise a first domain of HEAT repeats in the model of CENP-I structure. The coexpression with CENP-H and CENP-K was insufficient to stabilize longer CENP-I constructs. Accordingly, cross-linking coupled with mass spectrometry analyses confirmed that CENP-H and CENP-K are in close spatial proximity to the first ~350 amino acids of CENP-I, but not with the rest of this protein.

The coexpression with CENP-M was instead the key to solubilizing a longer CENP-I construct extending up to the C-terminus of the protein. The view that CENP-M might literally be embraced by CENP-I is in agreement with the fact that inter-molecular cross-links could be detected between

CENP-M and two regions of CENP-I, one towards its N-terminus and one towards its C-terminus, which are distant in its sequence but might come close in its structure, as suggested by the presence of intra-molecular cross-links between them.

In addition, it appears that CENP-M, which I expect to be located in the concave surface of CENP-I, does not establish any direct contact with CENP-H and CENP-K, which I imagine to be laid on the convex surface of CENP-I. This is indicated by the results of my analytical SEC migration shift assays and GST-pull-down experiments, in which CENP-M and CENP-H / CENP-K complex do not display any binding, and is confirmed by the fact that no inter-molecular cross-links could be detected between CENP-M and CENP-H and / or CENP-K in the context of the quaternary complex.

Another interesting observation is that CENP-H / CENP-K complex displays a SEC elution profile suggestive of an elongated structure, as its theoretical molecular weight is about 60 kDa, but it migrates in close proximity to the 158 kDa marker protein. CENP-H / CENP-K / CENP-$I^{57\text{-}C}$ / CENP-M complex instead elutes from SEC slightly before the 158 kDa marker protein, in accordance with its theoretical molecular weight of about 160 kDa, thus denoting a globular assembly.

It emerges from these considerations that significant information can already be inferred about the structural organization of CENP-H / CENP-K / CENP-$I^{57\text{-}C}$ / CENP-M complex. Nevertheless, medium- or high-resolution structural analyses on this complex are obviously foreseen. These will possibly include electron microscopy, small-angle X-ray scattering (SAXS) and X-ray crystallography studies.

As previously mentioned, the localization dependencies of CCAN proteins have been extensively investigated. Such dependencies were identified mainly through protein deletion and / or depletion experiments and are consistent with an extremely complicated network of relationships. It has recently become evident that a thorough mapping of the direct interactions that take place among these proteins is an essential condition to deconvolute such an intricate picture. In light of these considerations, the discovery of a quaternary complex comprising CENP-H / CENP-K / CENP-I / CENP-M and the elucidation of the direct interactions occurring within this complex represent an important step forward in this direction, as they substantiate the observation that these proteins display mutual dependencies for their kinetochore recruitment[18].

Obviously, a number of additional experiments are needed to validate and characterize the interaction between CENP-M and CENP-I *in vivo*. First of all, given that the commercially available

monoclonal antibody against CENP-M is only suitable for Western blotting, but not for immunofluorescence, I recently generated a polyclonal antibody raised against recombinant purified full-length human CENP-M that performs well in both applications. This will allow us to characterize not only the levels but also the localization of endogenous CENP-M in HeLa cells throughout the cell cycle and to verify that the GFP-CENP-M inducible HeLa cell line displays a similar behaviour. Also, we are currently developing a protocol for the effective depletion of endogenous CENP-M by RNAi. The next step will be to compare GFP-CENP-M wt and mutants, in a background of endogenous CENP-M depletion by RNAi, in terms of ability to localize at centromeres and to recruit other kinetochore components, such as CENP-I, as direct interacting partner of CENP-M, and other CCAN members whose centromeric localization has been reported to be dependent on CENP-M. Specifically, we plan to address this both via immunofluorescence experiments, which allow to obtain information about colocalization, and via immunoprecipitation experiments against GFP, in order to assess differences in the interactome of GFP-CENP-M wt in comparison with the mutants.

A few preliminary considerations can already be drawn from our initial experiments regarding the localization of GFP-CENP-M wt and mutants, although they were carried out in the presence of endogenous CENP-M. Specifically, mutants L94A + L163E and R145E + R148A lose the ability to localize at centromeres. We hypothesize that, as they are unable to interact with CENP-I *in vitro*, they are most likely incapable of doing so also *in vivo* and, consequently, are not recruited to the centromere, in agreement with the dependency of CENP-M on CENP-I for its kinetochore localization[18]. Interestingly, also mutant R60A + R62A hampers the localization of GFP-CENP-M at centromeres. This mutant retains the ability to interact with CENP-I *in vitro* and, therefore, we suppose that its inability to localize at centromeres *in vivo* is due to the loss of interaction with some other kinetochore component on which CENP-M depends for its recruitment. In particular, this mutant contains two amino acid substitutions in the region of CENP-M structure that corresponds to the switch II of G proteins, which is often involved in interactions with their partners[150]. My analysis of CENP-M binding to putative interacting partners among known centromere and kinetochore components *in vitro* through analytical SEC migration shift assays did not reveal any further direct interaction other than the one with CENP-I. A possible explanation is that CENP-M is involved in low affinity interactions with other kinetochore proteins, for which SEC is not a suitable detection method. It is plausible that, by reconstituting more complex protein

assemblies, CENP-M might be incorporated, maybe even in the absence of CENP-I, thanks to the contributions of multiple low affinity interactions to an overall higher binding affinity. Another option is that CENP-M, upon binding to CENP-I, undergoes conformational changes that are required for other interactions.

Moreover, the novel finding of a direct interaction of CENP-H / CENP-K complex with the N-terminal half of CENP-C provides a molecular corroboration to previous studies reporting a dependency of CENP-H and CENP-K on CENP-C for their centromeric localization[57,91]. A number of experiments are in the pipeline to better characterize this interaction. Specifically, the design of shorter CENP-C constructs will be instrumental in identifying regions of this protein that are necessary and / or sufficient for binding to CENP-H / CENP-K complex. A more detailed point mutation analysis of conserved residues is then foreseen. Moreover, given that the quaternary CENP-H / CENP-K / CENP-I$^{57\text{-}C}$ / CENP-M complex also binds to CENP-C$^{1\text{-}544}$His, a possible contribution of CENP-I to CENP-C binding will also be assessed. This will include a comparison of the behaviour of CENP-H / CENP-K complex and CENP-H / CENP-K / CENP-I$^{57\text{-}C}$ / CENP-M complex in analytical SEC migration shift assays with various CENP-C constructs and the measurement of binding affinities, for example through microscale thermophoresis (MST) or isothermal titration calorimetry (ITC). Given the results of my SEC migration shift assays, a contribution of CENP-M to CENP-C binding seems instead unlikely, although it cannot be completely excluded at this stage. Also, high-resolution structural studies are being pursued. In particular, limited proteolysis experiments on single proteins and complexes are underway, in order to identify minimal interacting regions and take advantage of this information to design constructs possibly amenable to X-ray crystallography. Structural studies by small-angle X-ray scattering (SAXS) and single particle electron microscopy are also envisaged. In addition, the validation and characterization of CENP-H / CENP-K binding to CENP-C *in vivo* is foreseen. A possible way of addressing this point would be, similarly to what is being performed for CENP-M, the generation of inducible GFP-CENP-C HeLa cell lines for wt and mutated constructs, designed according to results obtained *in vitro*. This would allow us to test the ability of relevant CENP-C constructs to sustain or impair the centromeric recruitment of CENP-H and CENP-K, and eventually of other CCAN components that depend on them for their kinetochore localization.

In fact, I believe that a valuable aspect of our approach resides in the possibility to rely on molecular details about protein structures and direct interactions obtained through *in vitro*

investigations to then design tailored *in vivo* experiments. This provides a means to dissect the role of specific protein regions or even residues and, eventually, to substantiate and deconvolute the usually broader and more complicated effects observed when deleting and / or depleting entire proteins.

In Figure 43 I propose a model of the molecular architecture of human CCAN, which takes into account the discoveries presented in this study and positions them in the context of prior knowledge reported in the literature.

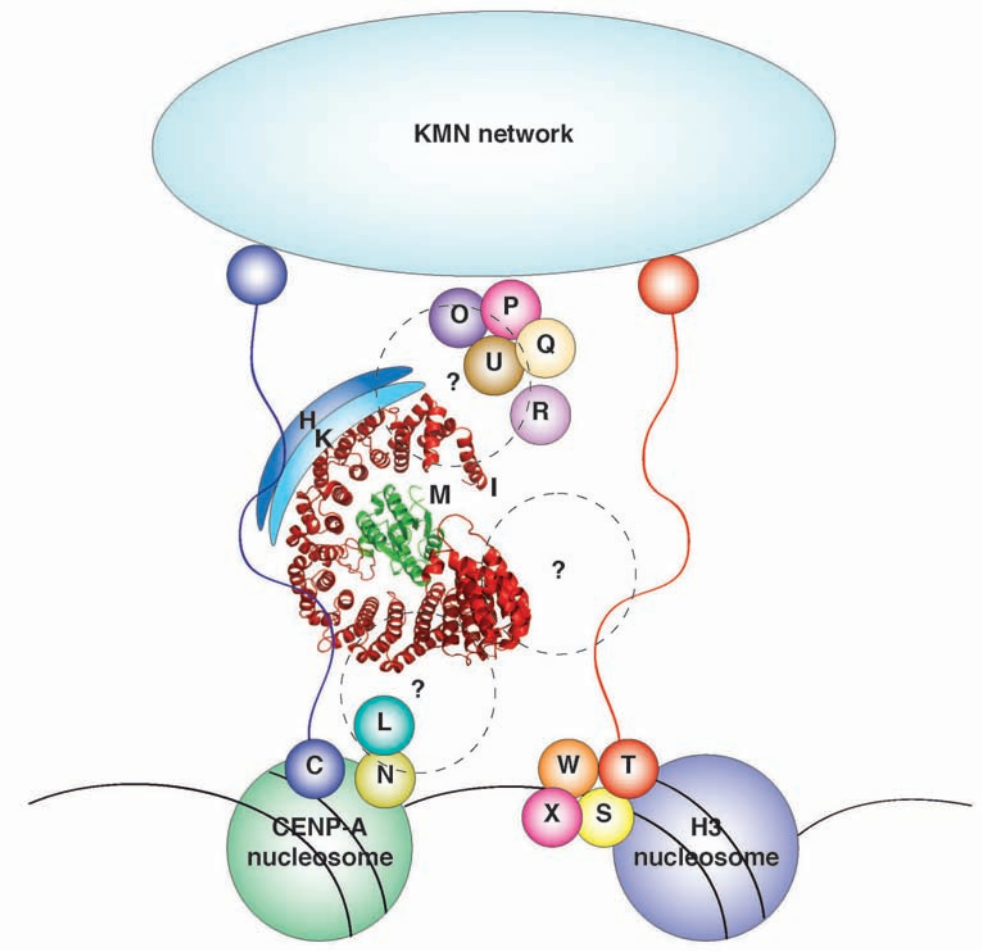


Figure 43 - Model of the molecular architecture of human CCAN.

Centromeric chromatin, including both canonical H3 and centromere-specific CENP-A nucleosomes[34], lies at the basis of kinetochore. CENP-C directly binds to CENP-A nucleosomes[43,57] and represents a bridge with the outer kinetochore[94,95]. The evidence that CENP-C possesses two nucleosome binding regions (the CENP-C central domain and the CENP-C motif) and the fact that it dimerizes through its C-terminal region have been omitted in this model for simplicity. CENP-N also directly interacts with CENP-A nucleosomes as well as with CENP-L[56]. CENP-T / CENP-W / CENP-S / CENP-X associate through their histone fold domains to form a quaternary complex[111], whose relationship with centromeric chromatin is currently a matter of debate. The N-terminal region of CENP-T reaches the outer kinetochore[25,114,115]. My study has revealed the existence of a quaternary CENP-H / CENP-K / CENP-I / CENP-M complex, which, through its CENP-H, CENP-K and

possibly CENP-I subunits, directly binds to the N-terminal half of CENP-C. Further investigations are required to better elucidate the relationships among this complex, the CENP-L / CENP-N group, the CENP-T / CENP-W / CENP-S / CENP-X group and the CENP-O / CENP-P / CENP-Q / CENP-U / CENP-R group of proteins.

As already mentioned, a collaborative effort is in progress in our group with the aim to reconstitute the human kinetochore *in vitro* from purified recombinant components. Within this framework, in the last few years I have been committed to the establishment in the laboratory of techniques for the recombinant expression and purification of histones and for the *in vitro* reconstitution of both H3-containing mononucleosomes (H3-MN) and CENP-A-containing mononucleosomes (CA-MN). The production of material of good quality and quantity has recently been achieved, allowing us to start analysing their *in vitro* interactions with kinetochore components. Obviously, one of our priorities is to be able to reproduce the direct interactions of CENP-C[43,57] and CENP-N[56] with CA-MN, which have already been described in the literature, and to characterize them in the context of our reconstitution endeavours. Also, the nature of the proposed interaction of CENP-T / CENP-W / CENP-S / CENP-X complex with centromeric chromatin[111] is still a matter of debate and definitely requires further investigations.

During my initial experiments aimed at testing interactions of H3-MN and CA-MN with kinetochore components *in vitro* though analytical SEC migration shift assays, I observed an unprecedented direct binding of Mis12 complex (Mis12 C) to nucleosomes. Specifically, my preliminary results suggest that Mis12 C is capable of binding to both H3-MN and CA-MN *in vitro* with a 2 : 1 stoichiometry, which reflects the fact that nucleosomes possess a two-fold symmetry axis. The interaction seems to involve the nucleosome core particle, as it occurs also in the absence of DNA tails extending from the minimal DNA region that is wrapped around the histone octamer. Also, it appears that the C-terminal region (residues ~250 to 281) of the Nsl1 subunit of Mis12 C is both necessary and sufficient for the binding. This is in agreement with the fact that Nsl1 was shown to associate with centromeric chromatin independently of the other Mis12 C subunits and of CENP-C in *Xenopus laevis* sperm[91]. Moreover, the binding is sensitive to the concentration of salt used in the reaction buffer, suggesting that electrostatic interactions play a major role.

The significance of this observation and its relevance in the context of kinetochore biology remain to be understood. First of all, the interaction of Mis12 C with nucleosomes needs to be analysed in the context of known Mis12 C interactions. In particular, it is well established that Mis12 C is part of

the so-called KMN network[20], which is essential to form load-bearing end-on attachments to spindle microtubules. Thus, it is crucial to ask if the observed binding of Mis12 C to nucleosomes is compatible with its additional interactions within the KMN network. Specifically, as a starting point in this direction, I considered that Mis12 C directly binds to the C-terminal region of Knl1 and the C-terminal tail of Nsl1 is necessary and sufficient for this interaction. A preliminary experiment that I performed suggests that a pre-formed complex comprising the C-terminal tails of Nsl1 and Knl1, specifically Nsl1$^{256-281}$ bound to Knl1$^{2091-2311}$, is not able to simultaneously bind to H3-MN (data not shown), while I still have to repeat this experiment using CA-MN. Thus, it seems that the C-terminal tail of Nsl1 is not able to contemporarily bind to the C-terminal region of Knl1 and to H3-MN. However, the C-terminal tail of Nsl1 does not fully recapitulate the binding of full-length Mis12 C to the C-terminal region of Knl1, which displays a significantly higher overall binding affinity[20]. Therefore, it is important to verify if the use of Mis12 C full-length instead of the C-terminal tail of Nsl1 makes any difference in the context of my assay.

Moreover, it is known that Mis12 C directly binds, probably through its Mis12 subunit, to the N-terminal region of the CCAN component CENP-C[94]. Also, the central domain of CENP-C (residues 422 - 537) and the so-called CENP-C motif (residues 736 - 758), directly bind to CA-MN and, although with reduced affinity, to H3-MN[43,57]. Thus, an investigation on the relationships between Mis12 C binding to CENP-C and the nucleosome binding properties of these proteins is also foreseen.

As previously mentioned, Nsl1 was shown to associate with centromeric chromatin independently of the other Mis12 C subunits and of CENP-C in *Xenopus laevis* sperm[91]. In light of this consideration, I can hypothesize that Nsl1 is recruited to the kinetochore through a direct binding to centromeric chromatin, while CENP-C, through the interaction of its N-terminal region with the Mis12 subunit, could contribute additional binding affinity for the kinetochore recruitment of the whole Mis12 C. This might determine the formation of a stable chromatin-bound load-bearing basis on which, during mitosis, the whole KMN network could be reconstituted. I envision that, during mitosis, the kinetochore localization of Mis12 C could be maintained through its binding to CENP-C and thus to CA-MN. Therefore, the Nsl1 C-terminal tail might become dispensable for the kinetochore localization of Mis12 C and be available for the interaction with the C-terminal region of Knl1.

In addition, one of our priorities is certainly the confirmation of the stoichiometry of the interaction

between Mis12 C and nucleosomes and the measurement of binding affinities, for example through isothermal titration calorimetry (ITC) or microscale thermophoresis (MST). In particular, a crucial point is the comparison of the binding affinities of Mis12 C for H3-MN and CA-MN. This will allow us to understand if there is a preference for one of these two nucleosome types, as is the case for CENP-C, and, on this basis, speculate about possible mechanisms involved in determining binding specificity. Also, it will be important to clarify if the C-terminal tail of Nsl1 is sufficient to completely recapitulate the binding of Mis12 C full-length to nucleosomes or if other Mis12 C regions contribute to the overall binding affinity.

To gain further understanding of the regions involved in the interaction, experiments of cross-linking coupled with mass spectrometry will be valuable.

Furthermore, high-resolution structural studies can be envisioned. In particular, the C-terminal tail of Nsl1 (residues ~250 to 281) bound to mononucleosomes may represent an appropriate crystallographic target. An interesting feature of this region of Nsl1 is the presence of an extremely conserved arginine in position 258 and of a conserved arginine in position 267[20]. This consideration raises the intriguing hypothesis that the C-terminal tail of Nsl1 may bind to the acidic patch of histones H2A-H2B with a similar arginine-dependent mechanism to the one observed for Kaposi's sarcoma-associated Herpesvirus LANA (Latency-Associated Nuclear Antigen) peptide[156], RCC1 (Regulator of Chromosome Condensation 1) chromatin factor[157], Sir3 BAH domain[158], and the aforementioned CENP-C motif[43], which are among the few published crystallographic structures of nucleosome-bound proteins. As a complementary approach, structural studies by single particle electron microscopy on the whole Mis12 C bound to mononucleosomes are foreseen.

To confirm Mis12 C interaction with mononucleosomes *in vivo*, we are also planning FRET assays with H3-MN in comparison to CA-MN, which will be carried out in collaboration with Prof. Stephan Diekmann's laboratory at the Fritz Lipmann Institute in Jena.

4 Materials and methods

4.1 Plasmids, protein expression and purification

4.1.1 CENP-M, CENP-M$^{1-171}$ and CENP-M mutants

A cDNA segment encoding human CENP-M isoform 1, the canonical one, was subcloned in pGEX-6P-2rbs vector, a dicistronic derivative of pGEX-6P vector generated in house, as a C-terminal fusion to GST, with an intervening 3C protease site. The construct CENP-M$^{1-171}$ was created by insertion of a stop codon using the QuikChange Mutagenesis kit (Stratagene). Constructs were sequence verified. The expression and purification procedure was the same for both CENP-M constructs. *Escherichia coli* C41 (DE3) cells harbouring vectors expressing GST-CENP-M or GST-CENP-M$^{1-171}$ were grown in Terrific Broth at 37°C to an OD_{600} of 0.6 - 0.8, then 0.2 mM IPTG was added and the culture was grown at 18°C overnight. Cell pellets were resuspended in lysis buffer (50 mM Tris/HCl pH 7.4, 300 mM NaCl, 5 % Glycerol, 1 mM DTT) supplemented with protease inhibitor cocktail (Serva), lysed by sonication and cleared by centrifugation at 48000 g at 4°C for 1 h. The cleared lysate was applied to Glutathione Sepharose 4 Fast Flow beads (GE Healthcare) pre-equilibrated in lysis buffer, incubated at 4°C for 2 h, washed with 70 volumes of lysis buffer and subjected to an overnight cleavage reaction with 3C protease. Resource S cation exchange chromatography column (GE Healthcare) was pre-equilibrated in a mixture of 95 % buffer A (20 mM MES pH 6.0, 5 % Glycerol, 1 mM DTT) and 5 % buffer B (20 mM MES pH 6.0, 1 M NaCl, 5 % Glycerol, 1 mM DTT). The eluate from Glutathione beads was diluted with buffer A to reach a final concentration of 50 mM NaCl, loaded onto the Resource S column and eluted with a linear gradient of buffer B from 50 to 500 mM NaCl in 10 bed column volumes. Fractions containing CENP-M were concentrated in 10 kDa molecular weight cut-off Vivaspin concentrators (Sartorius) and loaded onto a Superdex 75 size-exclusion chromatography (SEC) column (GE Healthcare) pre-equilibrated in SEC buffer (10 mM MES pH 6.0, 150 mM NaCl, 1 mM TCEP). SEC was performed under isocratic conditions at a flow rate of 0.5 ml/min. Fractions containing CENP-M were concentrated, flash-frozen in liquid nitrogen and stored at -80°C. Mutant CENP-M constructs were created by site-directed mutagenesis using the QuikChange Mutagenesis kit (Stratagene). Constructs were

sequence verified. The expression and purification procedure was the same as already described, except for the omission of the cation exchange chromatography step.

4.1.2 CENP-H / CENP-K complex

A cDNA segment encoding human CENP-K was subcloned in pFH vector as a C-terminal fusion to His-tag, with an intervening TEV protease site, while a cDNA segment encoding human CENP-H was subcloned in pUCDM vector, without any tag. Constructs were sequence verified. The two vectors were then fused via *in vitro* Cre-loxP recombination. Viral production and protein expression were then performed as in[159]. Cell pellets were resuspended in lysis buffer (50 mM Tris/HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole, 5 % Glycerol, 2 mM β-mercaptoethanol) supplemented with protease inhibitor cocktail (Serva), lysed by sonication and cleared by centrifugation at 48,000 g at 4°C for 1 h. The cleared lysate was applied to Ni-NTA Agarose beads (QIAGEN) pre-equilibrated in lysis buffer, incubated at 4°C for 2 h and washed with 70 volumes of lysis buffer. Bound proteins were eluted with lysis buffer supplemented with 200 mM imidazole and then dialysed against 50 mM Tris/HCl pH 8.0, 150 mM NaCl, 5 % Glycerol, 0.5 mM EDTA, 1 mM DTT at 4°C overnight. During this dialysis step, tag cleavage with TEV protease was also performed. Resource Q anion exchange chromatography column (GE Healthcare) was pre-equilibrated in a mixture of 92.5 % buffer A (50 mM Tris/HCl pH 8.0, 5 % Glycerol, 0.5 mM EDTA, 1 mM DTT) and 7.5 % buffer B (50 mM Tris/HCl pH 8.0, 1 M NaCl, 5 % Glycerol, 0.5 mM EDTA, 1 mM DTT). The dialysed sample was diluted with buffer A to reach a final concentration of 75 mM NaCl, loaded onto the Resource Q column and eluted with a linear gradient of buffer B from 75 to 500 mM NaCl in 10 bed column volumes. Fractions containing CENP-H / CENP-K complex were concentrated in 10 kDa molecular weight cut-off Vivaspin concentrators (Sartorius) and loaded onto a Superdex 200 SEC column (GE Healthcare) pre-equilibrated in SEC buffer (10 mM Hepes pH 7.5, 150 mM NaCl, 1 mM TCEP). SEC was performed under isocratic conditions at a flow rate of 0.5 ml/min. Fractions containing CENP-H / CENP-K complex were concentrated up to 10 mg/ml, flash-frozen in liquid nitrogen and stored at -80°C.

4.1.3 CENP-T / CENP-W complex and CENP-S / CENP-X complex

A cDNA segment encoding human CENP-T isoform 1, the canonical one, was subcloned in pGEX-6P-2rbs vector as a C-terminal fusion to GST, with an intervening 3C protease site. A cDNA

segment encoding human CENP-W was subcloned in the second cassette of the same vector. Similarly, a synthetic cDNA segment encoding human CENP-X isoform 1, the canonical one, codon-optimised for expression in bacteria as well as insect cells, was subcloned in pGEX-6P-2rbs vector as a C-terminal fusion to GST, with an intervening 3C protease site. Also, a cDNA segment encoding human CENP-S isoform 1, the canonical one, was subcloned in the second cassette of the same vector. Constructs were sequence verified. The expression and purification procedure was the same for CENP-T / CENP-W and CENP-S / CENP-X complexes. *Escherichia coli* BL21 Rosetta cells harbouring vectors expressing GST-CENP-T / CENP-W or GST-CENP-X / CENP-S were grown in Terrific Broth at 37°C to an OD_{600} of 0.6 - 0.8, then 0.3 mM IPTG was added and the culture was grown at 20°C overnight. Cell pellets were resuspended in lysis buffer (25 mM Tris/HCl pH 7.5, 300 mM NaCl, 10 % Glycerol, 1 mM DTT) supplemented with protease inhibitor cocktail (Serva), lysed by sonication and cleared by centrifugation at 48000 g at 4°C for 1 h. The cleared lysate was applied to Glutathione Sepharose 4 Fast Flow beads (GE Healthcare) pre-equilibrated in lysis buffer, incubated at 4°C for 2 h, washed with 70 volumes of lysis buffer and subjected to an overnight cleavage reaction with 3C protease. Heparin column (GE Healthcare) was pre-equilibrated in a mixture of 85 % buffer A (20 mM Tris/HCl pH 7.5, 5 % Glycerol, 1 mM DTT) and 15 % buffer B (20 mM Tris/HCl pH 7.5, 2 M NaCl, 5 % Glycerol, 1 mM DTT). The eluate from Glutathione beads was directly loaded onto the heparin column and eluted with a linear gradient of buffer B from 300 to 1200 mM NaCl in 10 bed column volumes. Fractions containing CENP-T / CENP-W or CENP-S / CENP-X complex were concentrated in 10 kDa molecular weight cut-off Vivaspin concentrators (Sartorius) and loaded onto a Superdex 200 size-exclusion chromatography (SEC) column (GE Healthcare) pre-equilibrated in SEC buffer (20 mM Hepes pH 7.5, 300 mM NaCl, 5 % Glycerol, 1 mM TCEP). SEC was performed under isocratic conditions at a flow rate of 0.5 ml/min. Fractions containing CENP-T / CENP-W or CENP-S / CENP-X complex were concentrated, flash-frozen in liquid nitrogen and stored at -80°C.

4.1.4 Other purified recombinant proteins

Mis12 complex (full-length and deletion mutants) and C-terminal region of Knl1 - Expression and purification were performed according to the protocol described in reference[20].

Ndc80 complex - Expression and purification were performed according to the protocol described in reference[19].

H3-containing mononucleosomes and CENP-A-containing mononucleosomes - Plasmids for the production of *Xenopus laevis* H2A, H2B, H3 and H4 histones were a kind gift of Prof. Daniela Rhodes' laboratory at the Medical Research Council in Cambridge. Plasmids for the production of human CENP-A / H4 histone tetramer and *Xenopus laevis* H2A / H2B histone dimer were a kind gift of Prof. Aaron F. Straight's laboratory at Stanford University. Plasmids for the production of 601 - 167 bp and 601 - 145 bp DNA were a kind gift of Prof. Daniela Rhodes' laboratory and Prof. Curt A. Davey's laboratory at the Nanyang Technological University in Singapore, respectively. *Xenopus laevis* histone expression and purification, histone octamer refolding, DNA production and reconstitution of H3-containing mononucleosomes were performed according to the protocol described in reference[151]. The production of CENP-A-containing mononucleosomes was instead carried out according to the protocol described in reference[152].

Other purified recombinant proteins were a kind gift of other members of the laboratory.

4.2 Antibodies

Mouse monoclonal antibody against CENP-M (H00079019-M01) was purchased from Abnova Corporation. Rabbit polyclonal antibody against recombinant purified full-length human CENP-M was produced in house. Rabbit polyclonal antibody against CENP-I was a kind gift of Prof. Song-Tao Liu's laboratory at the University of Toledo. Goat polyclonal antibody against CENP-H (sc-11297) was purchased from Santa Cruz Biotechnology. Mouse monoclonal antibody anti penta-His was from QIAGEN. Human anti-centromere antibodies serum was from Antibodies Inc.

4.3 CENP-M$^{1-171}$ crystallization and structure determination

CENP-M$^{1-171}$ (~10 mg/ml) was crystallized by sitting drop vapour diffusion using a Honeybee Cartesian robot and 96-well plates. Diffraction-quality crystals were obtained by optimizing the initial condition in hanging drops. The optimal reservoir buffer contained 100 mM Bicine pH 8.5, 11 % MPD and 8 mM Spermidine. Crystals were transferred to a cryobuffer containing the reservoir

liquor supplemented with 15 % Glycerol and flash-frozen in liquid nitrogen. A selenomethionine (SeMet)-derivative of the protein was crystallized under similar conditions. X-ray diffraction data were collected with synchrotron radiation at beamline ID14-4, European Synchrotron Radiation Facility (Grenoble, France) for the native crystal, and beamline X06DA (PXIII), Swiss Light Source (Villigen-PSI, Switzerland) for the SeMet crystal. X-ray diffraction data were processed with xia2 (version 0.3.3.1)[160]. Analysis of data quality and crystal defects was performed using phenix.xtriage[129]. SAD phases obtained using Phenix AutoSol[129] yielded an interpretable 2 Å electron density map. Model building was carried out in Coot[161], with the help of fragments built automatically by Phenix AutoBuild[129], ARP/wARP[162] and Buccaneer[163]. The model was then used for molecular replacement into the native dataset using Phenix AutoMR[129]. Iterative model building with Coot and refinement with phenix.refine[129] yielded a final model covering the full asymmetric unit. The Collaborative Computational Project 4 (CCP4) suite[164] was also used at several stages. The structure was illustrated with PyMOL (DeLano Scientific LLC).

4.4 *In vitro* protein binding to adenine and guanine nucleotides

N-methylanthraniloyl (MANT)-labelled nucleotides (ADP, ATP, GDP, GTP) (Pharma Waldhof) were employed and protein-nucleotide interactions were detected exploiting the environmental sensitivity of MANT, as its fluorescence quantum yield increases in nonpolar solvents and upon binding to proteins. Fluorescence data were recorded with a Fluoromax-4 spectrophotometer (Jobin Yvon), with excitation and emission wavelengths of MANT-nucleotides at 366 and 450 nm, respectively. Arl2, a member of the Ras superfamily of small GTPases, was used as control (kind gift of Dr. Mandy Miertzschke). 500 μl of 1 μM MANT-labelled nucleotides in CENP-M SEC buffer were used. After 7 min, when the fluorescence baseline signal was stabilized, 10 μM recombinant purified CENP-M or Arl2 was added and the fluorescence signal was monitored for 1 h.

4.5 GST-CENP-M (wt and mutants), CENP-I^{57-C}, HisCENP-K and CENP-H coexpression in insect cells and GST-pull-downs

The MultiBac baculovirus-based system was employed for multiprotein overexpression in insect cells[159].

A cDNA segment encoding human CENP-M isoform 1, the canonical one, was subcloned in pFG vector as a C-terminal fusion to GST, with an intervening 3C protease site. A synthetic cDNA (GENEART) encoding human CENP-I and codon optimized for expression in bacteria as well as insect cells was used. In particular, the segment encoding amino acids 57 - 756 (C-terminus) was subcloned in the second cassette of the same pFG vector. Mutant CENP-M constructs were created by site-directed mutagenesis using the QuikChange Mutagenesis kit (Stratagene). Constructs were sequence verified.

A cDNA segment encoding human CENP-K was subcloned in pFH vector as a C-terminal fusion to His-tag, with an intervening TEV protease site, while a cDNA segment encoding human CENP-H was subcloned in pUCDM vector, without any tag. Constructs were sequence verified. The two vectors were then fused via *in vitro* Cre-loxP recombination. This same vector was used for the production of CENP-H / CENP-K complex, as detailed in a previous section.

Viral production was then performed as in[159].

For each GST-pull-down experiment, 25 ml of freshly diluted Tna38 cells at a density of 1 * 10^6 cells / ml in serum-free medium (Sf-900 II SFM, Life Technologies) were coinfected with 1 : 10 V2 GST-CENP-M / CENP-I^{57-C} and V2 HisCENP-K / CENP-H for 72 h at 27°C with shaking. Cell pellets were resuspended in lysis buffer (20 mM Hepes pH 7.5, 300 mM NaCl, 1 mM TCEP) supplemented with protease inhibitor cocktail (Serva), lysed by sonication and cleared by centrifugation at 20000 g at 4°C for 30 min. The cleared lysate was applied to Glutathione Sepharose 4 Fast Flow beads (GE Healthcare) pre-equilibrated in lysis buffer, incubated at 4°C for 2 h, washed with 60 volumes of lysis buffer and eluted with lysis buffer supplemented with 30 mM reduced Glutathione. Samples of total lysate, supernatant, beads before elution and elution were analysed by SDS-PAGE and Coomassie blue staining and by Western blotting.

4.6 Analytical size-exclusion chromatography (SEC) migration shift assays

Analytical SEC experiments were performed on calibrated Superdex 200 5/150 or Superose 6 5/150 columns (GE Healthcare). All samples were eluted under isocratic conditions at 4°C in SEC buffer at a flow rate of 0.2 ml/min for Superdex 200 5/150 or 0.1 ml/min for Superose 6 5/150. Elution of proteins was monitored at 280 nm. 100 µl fractions were collected and analysed by SDS-PAGE and Coomassie blue staining. To detect the formation of a complex, proteins were mixed at the indicated concentrations in 50 µl, incubated for at least 2 h on ice and then subjected to SEC.

For binding assays with nucleosomes, a SEC buffer containing 10 mM Hepes pH 7.5, 50 mM NaCl, 1 mM TCEP was used. Also, buffers containing 100 or 200 mM NaCl were employed for the salt screening experiments, as indicated.

For the other binding assays, a SEC buffer containing 10 mM Hepes pH 7.5, 150 mM NaCl, 1 mM TCEP was used when possible (namely, with CENP-M, CENP-H / CENP-K, HisCENP-I$^{57-281}$, CENP-H / CENP-K / CENP-I$^{57-281}$, CENP-H / CENP-K / CENP-I^{57-C} / CENP-M, Mis12 complex, Ndc80 complex, Knl1$^{2000-2311}$, Zwint). A SEC buffer containing 300 mM NaCl was instead employed with proteins that were not stable in lower NaCl concentrations (specifically, CENP-C constructs, CENP-T / CENP-W, CENP-S / CENP-X, CENP-L / CENP-N, CENP-O / CENP-P / CENP-Q / CENP-U, CENP-R).

4.7 CENP-M and microtubules cosedimentation assay

Tubulin was purchased from Cytoskeleton Inc. and microtubules were polymerized according to the producer's instructions. Briefly, 50 µM tubulin in tubulin buffer (80 mM PIPES pH 6.8, 1 mM $MgCl_2$, 1 mM EGTA) was incubated in the presence of 1 mM GTP for 30 min at 34°C, then supplemented with 0.5 mM Taxol and further incubated for 30 min at 34°C.

For binding assays, microtubules and CENP-M were diluted in tubulin buffer supplemented with 10 µM Taxol in a final volume of 25 µl. 2 µM CENP-M was employed in all binding reactions, while different microtubules concentrations were tested, as indicated. Reactions were incubated at room temperature for 10 min, transferred onto 100 µl of cushion buffer (tubulin buffer supplemented with

10 μM Taxol and 50 % Glycerol) and ultracentrifuged for 10 min at 90000 rpm in a Beckman TLA120.1 rotor at 25°C. Supernatant and pellet fractions were analysed by SDS-PAGE.

4.8 Experiments of cross-linking coupled with mass spectrometry

Experiments of cross-linking coupled with mass spectrometry were carried out essentially as in reference[148], thanks to a collaboration with Dr. Franz Herzog's laboratory at the Ludwig Maximilian University in Munich.

4.9 Plasmids for mammalian protein expression, cell culture and immunofluorescence

Plasmids - A cDNA segment encoding human CENP-M isoform 1, the canonical one, was subcloned in pcDNA5/FRT/TO-GFP-ires vector, a derivative of pcDNA5/FRT/TO vector generated in house. Mutant CENP-M constructs were created by site-directed mutagenesis using the QuikChange Mutagenesis kit (Stratagene). Constructs were sequence verified.

Cell culture - Flp-In T-REx HeLa cells were a kind gift of Prof. Stephen Taylor's laboratory at the University of Manchester. Stable Flp-In T-REx HeLa cell lines expressing GFP-CENP-M wt or mutant constructs were generated according to Invitrogen protocol and selected in DMEM (Euroclone) supplemented with 10 % TET-free Fetal Bovine Serum (Invitrogen), 250 μg/ml Hygromycin (Roche) and 5 μg/ml Blasticidin (ICN chemicals). HeLa cells were grown in DMEM supplemented with 10 % TET-free Fetal Bovine Serum and 2 mM L-glutamine. Protein expression was induced with 1 - 50 nM Doxycycline (Sigma) for 24 h. A treatment with 9 μM RO-3306 for 18 h was employed to arrest cells in a G2-like phase[149].

Immunofluorescence - HeLa cells were grown on coverslips pre-treated with 15 μg/ml poly(D)lysine (Sigma). 24 h after induction of protein expression, cells were fixed with 4 % paraformaldehyde in PBS for 10 min. Before incubations with antibodies, cells were permeabilized using 0.1 % Triton X-100 in PBS for 10 min and treated with 2.5 % BSA in PBS as blocking agent. The primary

antibodies that were used are listed in the “Antibodies” section. Cy3- and Cy5-labelled secondary antibodies for immunofluorescence were purchased from Jackson Immuno Research Laboratories. Cells were imaged on a 3i Mariana^{TM} system (Intelligent Imaging Innovation Inc.) equipped with a spinning-disk confocal Axio Z1 microscope (Zeiss) provided with a 63X/NA1.4 objective (Zeiss) and an ORCA-Flash 4.0 CMOS camera (Hamamatsu) and controlled by Slidebook 5.0 software (Intelligent Imaging Innovation Inc.). Images were acquired as 0.27 μm Z-sections and displayed as maximal intensity projections.

5 Acknowledgements and contributions

I have been enrolled as a Ph.D. student at the European School of Molecular Medicine (SEMM) under the supervision of Prof. Andrea Musacchio. The work presented in this dissertation was carried out between January 2010 and October 2013. In particular, I spent the first two years at the IFOM-IEO Campus in Milan (Italy) and the last two years at the Max-Planck Institute of Molecular Physiology in Dortmund (Germany).

First of all, I am grateful to my supervisor Prof. Andrea Musacchio for his outstanding scientific guidance and constant support throughout these years as well as for giving me the freedom to pursue independent work.

My internal advisor Dr. Marina Mapelli and my external advisor Prof. Michel O. Steinmetz are also highly acknowledged for fruitful scientific discussions.

I would definitely like to thank all the past and present laboratory members for their contributions to my project in terms of both ideas and reagents and for creating a pleasant working atmosphere.

More specifically, Lucia Massimiliano and Dr. Emanuela Screpanti taught me the basis of recombinant protein expression and purification. Dr. Sebastiano Pasqualato introduced me to the world of crystallography and shared with me the tenacity in pursuing the determination of CENP-M crystal structure, while Valentina Cecatiello conveyed to me her great expertise in crystallogenesis. Dr. John Weir has been a brilliant collaborator on the biochemical work on CENP-H / CENP-K / CENP-I / CENP-M complex, while Dr. Stefano Maffini, Beate Voss, Dr. Veronica Krenn and Suzan van-Gerwen have provided their precious competence in cell biology. Dr. Ingrid Vetter, with her multifaceted proficiency, is highly acknowledged for profitable crystallographic discussions and for developing a script that allows the graphical visualization of cross-linking data. Dr. Mandy Miertzschke kindly showed me how to test CENP-M binding to nucleotides *in vitro,* Dr. Alex Faesen taught me how to perform microtubules cosedimentation assays and Dr. Radovan Dvorsky provided suggestions about protein structure predictions. Kerstin Klare has been a great coworker on the interaction between CENP-H / CENP-K complex and CENP-C and has now taken this project over. Warm thanks also go to Dorothee Vogt, the technical assistant who has closely collaborated with me on the endeavour to establish in the laboratory the techniques for nucleosomes reconstitution. In this respect, we could count on the precious collaborations with

Prof. Daniela Rhodes' laboratory, and specifically with Dr. Fabrizio Martino, at the Medical Research Council in Cambridge and with Prof. Aaron F. Straight's laboratory at the Stanford University. Another fruitful collaboration has been the one with Dr. Franz Herzog's laboratory at the Ludwig Maximilian University in Munich, regarding the experiments of cross-linking coupled with mass spectrometry.

6 References

1 Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643-649, doi:10.1038/25292 (1998).

2 Weaver, B. A. & Cleveland, D. W. Does aneuploidy cause cancer? *Current opinion in cell biology* **18**, 658-667, doi:10.1016/j.ceb.2006.10.002 (2006).

3 Santaguida, S. & Musacchio, A. The life and miracles of kinetochores. *The EMBO journal* **28**, 2511-2531, doi:10.1038/emboj.2009.173 (2009).

4 McEwen, B. F., Dong, Y. & VandenBeldt, K. J. Using electron microscopy to understand functional mechanisms of chromosome alignment on the mitotic spindle. *Methods in cell biology* **79**, 259-293, doi:10.1016/S0091-679X(06)79011-2 (2007).

5 McEwen, B. F. & Dong, Y. Contrasting models for kinetochore microtubule attachment in mammalian cells. *Cellular and molecular life sciences : CMLS* **67**, 2163-2172, doi:10.1007/s00018-010-0322-x (2010).

6 Westhorpe, F. G. & Straight, A. F. Functions of the centromere and kinetochore in chromosome segregation. *Current opinion in cell biology* **25**, 334-340, doi:10.1016/j.ceb.2013.02.001 (2013).

7 DeLuca, J. G. & Musacchio, A. Structural organization of the kinetochore-microtubule interface. *Current opinion in cell biology* **24**, 48-56, doi:10.1016/j.ceb.2011.11.003 (2012).

8 Lara-Gonzalez, P., Westhorpe, F. G. & Taylor, S. S. The spindle assembly checkpoint. *Current biology : CB* **22**, R966-980, doi:10.1016/j.cub.2012.10.006 (2012).

9 Tanaka, T. U. Kinetochore-microtubule interactions: steps towards bi-orientation. *The EMBO journal* **29**, 4070-4082, doi:10.1038/emboj.2010.294 (2010).

10 Santaguida, S., Vernieri, C., Villa, F., Ciliberto, A. & Musacchio, A. Evidence that Aurora B is implicated in spindle checkpoint signalling independently of error correction. *The EMBO journal* **30**, 1508-1519, doi:10.1038/emboj.2011.70 (2011).

11 Cheeseman, I. M. & Desai, A. Molecular architecture of the kinetochore-microtubule interface. *Nature reviews. Molecular cell biology* **9**, 33-46, doi:10.1038/nrm2310 (2008).

12 Wei, R. R., Al-Bassam, J. & Harrison, S. C. The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. *Nature structural & molecular biology* **14**, 54-59, doi:10.1038/nsmb1186 (2007).

13 Zheng, L., Chen, Y. & Lee, W. H. Hec1p, an evolutionarily conserved coiled-coil protein, modulates chromosome segregation through interaction with SMC proteins. *Molecular and cellular biology* **19**, 5417-5428 (1999).

14 Ciferri, C. *et al.* Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell* **133**, 427-439, doi:10.1016/j.cell.2008.03.020 (2008).

15 De Wulf, P., McAinsh, A. D. & Sorger, P. K. Hierarchical assembly of the budding yeast kinetochore from multiple subcomplexes. *Genes & development* **17**, 2902-2921, doi:10.1101/gad.1144403 (2003).

16 Obuse, C. *et al.* Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. *Genes to cells : devoted to molecular & cellular mechanisms* **9**, 105-120 (2004).

17 Foltz, D. R. *et al.* The human CENP-A centromeric nucleosome-associated complex. *Nature cell biology* **8**, 458-469, doi:10.1038/ncb1397 (2006).

18 Okada, M. *et al.* The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nature cell biology* **8**, 446-457, doi:10.1038/ncb1396 (2006).

19 Ciferri, C. *et al.* Architecture of the human ndc80-hec1 complex, a critical constituent of the outer kinetochore. *The Journal of biological chemistry* **280**, 29088-29095, doi:10.1074/jbc.M504070200 (2005).

20 Petrovic, A. *et al.* The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *The Journal of cell biology* **190**, 835-852, doi:10.1083/jcb.201002070 (2010).

21 Westermann, S., Drubin, D. G. & Barnes, G. Structures and functions of yeast kinetochore complexes. *Annual review of biochemistry* **76**, 563-591, doi:10.1146/annurev.biochem.76.052705.160607 (2007).

22 Westermann, S. & Schleiffer, A. Family matters: structural and functional conservation of centromere-associated proteins from yeast to humans. *Trends in cell biology* **23**, 260-269, doi:10.1016/j.tcb.2013.01.010 (2013).

23 Joglekar, A. P. *et al.* Molecular architecture of the kinetochore-microtubule attachment site is conserved between point and regional centromeres. *The Journal of cell biology* **181**, 587-594, doi:10.1083/jcb.200803027 (2008).
24 Meraldi, P., McAinsh, A. D., Rheinbay, E. & Sorger, P. K. Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome biology* **7**, R23, doi:10.1186/gb-2006-7-3-r23 (2006).
25 Schleiffer, A. *et al.* CENP-T proteins are conserved centromere receptors of the Ndc80 complex. *Nature cell biology* **14**, 604-613, doi:10.1038/ncb2493 (2012).
26 Welburn, J. P. & Cheeseman, I. M. Toward a molecular structure of the eukaryotic kinetochore. *Developmental cell* **15**, 645-655, doi:10.1016/j.devcel.2008.10.011 (2008).
27 Henikoff, S., Ahmad, K. & Malik, H. S. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098-1102, doi:10.1126/science.1062939 (2001).
28 Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407-421 (2003).
29 Maddox, P. S., Oegema, K., Desai, A. & Cheeseman, I. M. "Holo"er than thou: chromosome segregation and kinetochore function in C. elegans. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **12**, 641-653, doi:10.1023/B:CHRO.0000036588.42225.2f (2004).
30 Allshire, R. C. & Karpen, G. H. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nature reviews. Genetics* **9**, 923-937, doi:10.1038/nrg2466 (2008).
31 Marshall, O. J., Chueh, A. C., Wong, L. H. & Choo, K. H. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *American journal of human genetics* **82**, 261-282, doi:10.1016/j.ajhg.2007.11.009 (2008).
32 Earnshaw, W. C. & Migeon, B. R. Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma* **92**, 290-296 (1985).
33 Meluh, P. B., Yang, P., Glowczewski, L., Koshland, D. & Smith, M. M. Cse4p is a component of the core centromere of Saccharomyces cerevisiae. *Cell* **94**, 607-613 (1998).
34 Blower, M. D., Sullivan, B. A. & Karpen, G. H. Conserved organization of centromeric chromatin in flies and humans. *Developmental cell* **2**, 319-330 (2002).
35 Warburton, P. E. *et al.* Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Current biology : CB* **7**, 901-904 (1997).
36 Stoler, S., Keith, K. C., Curnick, K. E. & Fitzgerald-Hayes, M. A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes & development* **9**, 573-586 (1995).
37 Howman, E. V. *et al.* Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 1148-1153 (2000).
38 Oegema, K., Desai, A., Rybina, S., Kirkham, M. & Hyman, A. A. Functional analysis of kinetochore assembly in Caenorhabditis elegans. *The Journal of cell biology* **153**, 1209-1226 (2001).
39 Regnier, V. *et al.* CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. *Molecular and cellular biology* **25**, 3967-3981, doi:10.1128/MCB.25.10.3967-3981.2005 (2005).
40 Van Hooser, A. A. *et al.* Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *Journal of cell science* **114**, 3529-3542 (2001).
41 Torras-Llort, M., Moreno-Moreno, O. & Azorin, F. Focus on the centre: the role of chromatin on the regulation of centromere identity and function. *The EMBO journal* **28**, 2337-2348, doi:10.1038/emboj.2009.174 (2009).
42 Arents, G. & Moudrianakis, E. N. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 11170-11174 (1995).
43 Kato, H. *et al.* A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340**, 1110-1113, doi:10.1126/science.1235532 (2013).
44 Tachiwana, H. *et al.* Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **476**, 232-235, doi:10.1038/nature10258 (2011).
45 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

46 Conde e Silva, N. *et al.* CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. *Journal of molecular biology* **370**, 555-573, doi:10.1016/j.jmb.2007.04.064 (2007).
47 Kingston, I. J., Yung, J. S. & Singleton, M. R. Biophysical characterization of the centromere-specific nucleosome from budding yeast. *The Journal of biological chemistry* **286**, 4021-4026, doi:10.1074/jbc.M110.189340 (2011).
48 Keith, K. C. *et al.* Analysis of primary structural determinants that distinguish the centromere-specific function of histone variant Cse4p from histone H3. *Molecular and cellular biology* **19**, 6130-6139 (1999).
49 Vermaak, D., Hayden, H. S. & Henikoff, S. Centromere targeting element within the histone fold domain of Cid. *Molecular and cellular biology* **22**, 7553-7561 (2002).
50 Black, B. E. *et al.* Structural determinants for generating centromeric chromatin. *Nature* **430**, 578-582, doi:10.1038/nature02766 (2004).
51 Black, B. E., Brock, M. A., Bedard, S., Woods, V. L., Jr. & Cleveland, D. W. An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5008-5013, doi:10.1073/pnas.0700390104 (2007).
52 Fachinetti, D. *et al.* A two-step mechanism for epigenetic specification of centromere identity and function. *Nature cell biology* **15**, 1056-1066, doi:10.1038/ncb2805 (2013).
53 Dunleavy, E. M. *et al.* HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485-497, doi:10.1016/j.cell.2009.02.040 (2009).
54 Hu, H. *et al.* Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes & development* **25**, 901-906, doi:10.1101/gad.2045111 (2011).
55 Cho, U. S. & Harrison, S. C. Recognition of the centromere-specific histone Cse4 by the chaperone Scm3. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 9367-9371, doi:10.1073/pnas.1106389108 (2011).
56 Carroll, C. W., Silva, M. C., Godek, K. M., Jansen, L. E. & Straight, A. F. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nature cell biology* **11**, 896-902, doi:10.1038/ncb1899 (2009).
57 Carroll, C. W., Milks, K. J. & Straight, A. F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *The Journal of cell biology* **189**, 1143-1155, doi:10.1083/jcb.201001013 (2010).
58 Chen, Y. *et al.* The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. *Molecular and cellular biology* **20**, 7037-7048 (2000).
59 Samel, A., Cuomo, A., Bonaldi, T. & Ehrenhofer-Murray, A. E. Methylation of CenH3 arginine 37 regulates kinetochore integrity and chromosome segregation. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 9029-9034, doi:10.1073/pnas.1120968109 (2012).
60 Sugasawa, K. *et al.* Nonconservative segregation of parental nucleosomes during simian virus 40 chromosome replication in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 1055-1059 (1992).
61 Shelby, R. D., Monier, K. & Sullivan, K. F. Chromatin assembly at kinetochores is uncoupled from DNA replication. *The Journal of cell biology* **151**, 1113-1118 (2000).
62 Jansen, L. E., Black, B. E., Foltz, D. R. & Cleveland, D. W. Propagation of centromeric chromatin requires exit from mitosis. *The Journal of cell biology* **176**, 795-805, doi:10.1083/jcb.200701066 (2007).
63 Schuh, M., Lehner, C. F. & Heidmann, S. Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. *Current biology : CB* **17**, 237-243, doi:10.1016/j.cub.2006.11.051 (2007).
64 Camahort, R. *et al.* Cse4 is part of an octameric nucleosome in budding yeast. *Molecular cell* **35**, 794-805, doi:10.1016/j.molcel.2009.07.022 (2009).
65 Zhang, W., Colmenares, S. U. & Karpen, G. H. Assembly of Drosophila centromeric nucleosomes requires CID dimerization. *Molecular cell* **45**, 263-269, doi:10.1016/j.molcel.2011.12.010 (2012).
66 Hasson, D. *et al.* The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nature structural & molecular biology* **20**, 687-695, doi:10.1038/nsmb.2562 (2013).
67 Padeganeh, A. *et al.* Octameric CENP-A nucleosomes are present at human centromeres throughout the cell cycle. *Current biology : CB* **23**, 764-769, doi:10.1016/j.cub.2013.03.037 (2013).
68 Dalal, Y., Wang, H., Lindsay, S. & Henikoff, S. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS biology* **5**, e218, doi:10.1371/journal.pbio.0050218 (2007).

69 Dimitriadis, E. K., Weber, C., Gill, R. K., Diekmann, S. & Dalal, Y. Tetrameric organization of vertebrate centromeric nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20317-20322, doi:10.1073/pnas.1009563107 (2010).
70 Black, B. E. & Bassett, E. A. The histone variant CENP-A and centromere specification. *Current opinion in cell biology* **20**, 91-100, doi:10.1016/j.ceb.2007.11.007 (2008).
71 Miell, M. D. *et al.* CENP-A confers a reduction in height on octameric nucleosomes. *Nature structural & molecular biology* **20**, 763-765, doi:10.1038/nsmb.2574 (2013).
72 Bui, M. *et al.* Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. *Cell* **150**, 317-326, doi:10.1016/j.cell.2012.05.035 (2012).
73 Shivaraju, M. *et al.* Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast. *Cell* **150**, 304-316, doi:10.1016/j.cell.2012.05.034 (2012).
74 Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M. M. & Wu, C. Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell* **129**, 1153-1164, doi:10.1016/j.cell.2007.04.026 (2007).
75 Izuta, H. *et al.* Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. *Genes to cells : devoted to molecular & cellular mechanisms* **11**, 673-684, doi:10.1111/j.1365-2443.2006.00969.x (2006).
76 Liu, S. T., Rattner, J. B., Jablonski, S. A. & Yen, T. J. Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. *The Journal of cell biology* **175**, 41-53, doi:10.1083/jcb.200606020 (2006).
77 Hori, T. *et al.* CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* **135**, 1039-1052, doi:10.1016/j.cell.2008.10.019 (2008).
78 Takahashi, K., Chen, E. S. & Yanagida, M. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science* **288**, 2215-2219 (2000).
79 Hori, T., Shang, W. H., Takeuchi, K. & Fukagawa, T. The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. *The Journal of cell biology* **200**, 45-60, doi:10.1083/jcb.201210106 (2013).
80 Hori, T. & Fukagawa, T. Establishment of the vertebrate kinetochores. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **20**, 547-561, doi:10.1007/s10577-012-9289-9 (2012).
81 Hemmerich, P. *et al.* Dynamics of inner kinetochore assembly and maintenance in living cells. *The Journal of cell biology* **180**, 1101-1114, doi:10.1083/jcb.200710052 (2008).
82 Kang, Y. H. *et al.* Self-regulated Plk1 recruitment to kinetochores by the Plk1-PBIP1 interaction is critical for proper chromosome segregation. *Molecular cell* **24**, 409-422, doi:10.1016/j.molcel.2006.10.016 (2006).
83 McAinsh, A. D., Meraldi, P., Draviam, V. M., Toso, A. & Sorger, P. K. The human kinetochore proteins Nnf1R and Mcm21R are required for accurate chromosome segregation. *The EMBO journal* **25**, 4033-4049, doi:10.1038/sj.emboj.7601293 (2006).
84 Hellwig, D. *et al.* Dynamics of CENP-N kinetochore binding during the cell cycle. *Journal of cell science* **124**, 3871-3883, doi:10.1242/jcs.088625 (2011).
85 Saitoh, H. *et al.* CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell* **70**, 115-125 (1992).
86 Tomkiel, J., Cooke, C. A., Saitoh, H., Bernat, R. L. & Earnshaw, W. C. CENP-C is required for maintaining proper kinetochore size and for a timely transition to anaphase. *The Journal of cell biology* **125**, 531-545 (1994).
87 Meluh, P. B. & Koshland, D. Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. *Molecular biology of the cell* **6**, 793-807 (1995).
88 Holland, S., Ioannou, D., Haines, S. & Brown, W. R. Comparison of Dam tagging and chromatin immunoprecipitation as tools for the identification of the binding sites for S. pombe CENP-C. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **13**, 73-83, doi:10.1007/s10577-005-7062-z (2005).
89 Heeger, S. *et al.* Genetic interactions of separase regulatory subunits reveal the diverged Drosophila Cenp-C homolog. *Genes & development* **19**, 2041-2053, doi:10.1101/gad.347805 (2005).
90 Moore, L. L. & Roth, M. B. HCP-4, a CENP-C-like protein in Caenorhabditis elegans, is required for resolution of sister centromeres. *The Journal of cell biology* **153**, 1199-1208 (2001).
91 Milks, K. J., Moree, B. & Straight, A. F. Dissection of CENP-C-directed centromere and kinetochore assembly. *Molecular biology of the cell* **20**, 4246-4255, doi:10.1091/mbc.E09-05-0378 (2009).

92 Tanaka, K., Chang, H. L., Kagami, A. & Watanabe, Y. CENP-C functions as a scaffold for effectors with essential kinetochore functions in mitosis and meiosis. *Developmental cell* **17**, 334-343, doi:10.1016/j.devcel.2009.08.004 (2009).
93 Kwon, M. S., Hori, T., Okada, M. & Fukagawa, T. CENP-C is involved in chromosome segregation, mitotic checkpoint function, and kinetochore assembly. *Molecular biology of the cell* **18**, 2155-2168, doi:10.1091/mbc.E07-01-0045 (2007).
94 Screpanti, E. *et al.* Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Current biology : CB* **21**, 391-398, doi:10.1016/j.cub.2010.12.039 (2011).
95 Przewloka, M. R. *et al.* CENP-C is a structural platform for kinetochore assembly. *Current biology : CB* **21**, 399-405, doi:10.1016/j.cub.2011.02.005 (2011).
96 Trazzi, S. *et al.* The C-terminal domain of CENP-C displays multiple and critical functions for mammalian centromere formation. *PloS one* **4**, e5832, doi:10.1371/journal.pone.0005832 (2009).
97 Cohen, R. L. *et al.* Structural and functional dissection of Mif2p, a conserved DNA-binding kinetochore protein. *Molecular biology of the cell* **19**, 4480-4491, doi:10.1091/mbc.E08-03-0297 (2008).
98 Gascoigne, K. E. *et al.* Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* **145**, 410-422, doi:10.1016/j.cell.2011.03.031 (2011).
99 Hinshaw, S. M. & Harrison, S. C. An Iml3-Chl4 Heterodimer Links the Core Centromere to Factors Required for Accurate Chromosome Segregation. *Cell reports*, doi:10.1016/j.celrep.2013.08.036 (2013).
100 McClelland, S. E. *et al.* The CENP-A NAC/CAD kinetochore complex controls chromosome congression and spindle bipolarity. *The EMBO journal* **26**, 5033-5047, doi:10.1038/sj.emboj.7601927 (2007).
101 Sugata, N., Munekata, E. & Todokoro, K. Characterization of a novel kinetochore protein, CENP-H. *Journal of Biological Chemistry* **274**, 27343-27346, doi:Doi 10.1074/Jbc.274.39.27343 (1999).
102 Sugata, N. *et al.* Human CENP-H multimers colocalize with CENP-A and CENP-C at active centromere-kinetochore complexes. *Hum Mol Genet* **9**, 2919-2926, doi:Doi 10.1093/Hmg/9.19.2919 (2000).
103 Nishihashi, A. *et al.* CENP-I is essential for centromere function in vertebrate cells. *Developmental cell* **2**, 463-476, doi:Doi 10.1016/S1534-5807(02)00144-2 (2002).
104 Cheeseman, I. M., Hori, T., Fukagawa, T. & Desai, A. KNL1 and the CENP-H/I/K complex coordinately direct kinetochore assembly in vertebrates. *Molecular biology of the cell* **19**, 587-594, doi:10.1091/mbc.E07-10-1051 (2008).
105 Mikami, Y., Hori, T., Kimura, H. & Fukagawa, T. The functional region of CENP-H interacts with the Nuf2 complex that localizes to centromere during mitosis. *Molecular and cellular biology* **25**, 1958-1970, doi:10.1128/MCB.25.5.1958-1970.2005 (2005).
106 Qiu, S., Wang, J., Yu, C. & He, D. CENP-K and CENP-H may form coiled-coils in the kinetochores. *Science in China. Series C, Life sciences / Chinese Academy of Sciences* **52**, 352-359, doi:10.1007/s11427-009-0050-3 (2009).
107 Renou, J. P. *et al.* Identification of genes differentially expressed in mouse mammary epithelium transformed by an activated beta-catenin. *Oncogene* **22**, 4594-4610, doi:10.1038/sj.onc.1206596 (2003).
108 Bierie, B., Edwin, M., Melenhorst, J. J. & Hennighausen, L. The proliferation associated nuclear element (PANE1) is conserved between mammals and fish and preferentially expressed in activated lymphoid cells. *Gene expression patterns : GEP* **4**, 389-395, doi:10.1016/j.modgep.2004.01.008 (2004).
109 Amano, M. *et al.* The CENP-S complex is essential for the stable assembly of outer kinetochore structure. *The Journal of cell biology* **186**, 173-182, doi:10.1083/jcb.200903100 (2009).
110 Singh, T. R. *et al.* MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. *Molecular cell* **37**, 879-886, doi:10.1016/j.molcel.2010.01.036 (2010).
111 Nishino, T. *et al.* CENP-T-W-S-X forms a unique centromeric chromatin structure with a histone-like fold. *Cell* **148**, 487-501, doi:10.1016/j.cell.2011.11.061 (2012).
112 Bieniossek, C. *et al.* The architecture of human general transcription factor TFIID core complex. *Nature* **493**, 699-702, doi:10.1038/nature11791 (2013).
113 Kamada, K. *et al.* Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. *Cell* **106**, 71-81 (2001).
114 Nishino, T. *et al.* CENP-T provides a structural platform for outer kinetochore assembly. *The EMBO journal* **32**, 424-436, doi:10.1038/emboj.2012.348 (2013).

115 Malvezzi, F. *et al.* A structural basis for kinetochore recruitment of the Ndc80 complex via two distinct centromere receptors. *The EMBO journal* **32**, 409-423, doi:10.1038/emboj.2012.356 (2013).
116 Hori, T., Okada, M., Maenaka, K. & Fukagawa, T. CENP-O class proteins form a stable complex and are required for proper kinetochore function. *Molecular biology of the cell* **19**, 843-854, doi:10.1091/mbc.E07-06-0556 (2008).
117 Schmitzberger, F. & Harrison, S. C. RWD domain: a recurring module in kinetochore architecture shown by a Ctf19-Mcm21 complex structure. *EMBO reports* **13**, 216-222, doi:10.1038/embor.2012.1 (2012).
118 Nameki, N. *et al.* Solution structure of the RWD domain of the mouse GCN2 protein. *Protein science : a publication of the Protein Society* **13**, 2089-2100, doi:10.1110/ps.04751804 (2004).
119 Corbett, K. D. *et al.* The monopolin complex crosslinks kinetochore components to regulate chromosome-microtubule attachments. *Cell* **142**, 556-567, doi:10.1016/j.cell.2010.07.017 (2010).
120 Wei, R. R. *et al.* Structure of a central component of the yeast kinetochore: the Spc24p/Spc25p globular domain. *Structure* **14**, 1003-1009, doi:10.1016/j.str.2006.04.007 (2006).
121 Kim, S., Sun, H., Tomchick, D. R., Yu, H. & Luo, X. Structure of human Mad1 C-terminal domain reveals its involvement in kinetochore targeting. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6549-6554, doi:10.1073/pnas.1118210109 (2012).
122 Hua, S. *et al.* CENP-U cooperates with Hec1 to orchestrate kinetochore-microtubule attachment. *The Journal of biological chemistry* **286**, 1627-1638, doi:10.1074/jbc.M110.174946 (2011).
123 Amaro, A. C. *et al.* Molecular control of kinetochore-microtubule dynamics and chromosome oscillations. *Nature cell biology* **12**, 319-329, doi:10.1038/ncb2033 (2010).
124 Cheeseman, I. M., Chappie, J. S., Wilson-Kubalek, E. M. & Desai, A. The conserved KMN network constitutes the core microtubule-binding site of the kinetochore. *Cell* **127**, 983-997, doi:10.1016/j.cell.2006.09.039 (2006).
125 Kiyomitsu, T., Murakami, H. & Yanagida, M. Protein interaction domain mapping of human kinetochore protein Blinkin reveals a consensus motif for binding of spindle assembly checkpoint proteins Bub1 and BubR1. *Molecular and cellular biology* **31**, 998-1011, doi:10.1128/MCB.00815-10 (2011).
126 Pagliuca, C., Draviam, V. M., Marco, E., Sorger, P. K. & De Wulf, P. Roles for the conserved spc105p/kre28p complex in kinetochore-microtubule binding and the spindle assembly checkpoint. *PloS one* **4**, e7640, doi:10.1371/journal.pone.0007640 (2009).
127 Liu, D. *et al.* Regulated targeting of protein phosphatase 1 to the outer kinetochore by KNL1 opposes Aurora B kinase. *The Journal of cell biology* **188**, 809-820, doi:10.1083/jcb.201001006 (2010).
128 Krenn, V., Wehenkel, A., Li, X., Santaguida, S. & Musacchio, A. Structural analysis reveals features of the spindle checkpoint kinase Bub1-kinetochore subunit Knl1 interaction. *The Journal of cell biology* **196**, 451-467, doi:10.1083/jcb.201110013 (2012).
129 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
130 Parsons, S. Introduction to twinning. *Acta crystallographica. Section D, Biological crystallography* **59**, 1995-2003 (2003).
131 Dauter, Z. Twinned crystals and anomalous phasing. *Acta crystallographica. Section D, Biological crystallography* **59**, 2004-2016 (2003).
132 McPherson, A. & Cudney, B. Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *Journal of structural biology* **156**, 387-406, doi:10.1016/j.jsb.2006.09.006 (2006).
133 Walter, T. S. *et al.* Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617-1622, doi:10.1016/j.str.2006.09.005 (2006).
134 D'Arcy, A., Villard, F. & Marsh, M. An automated microseed matrix-screening method for protein crystallization. *Acta crystallographica. Section D, Biological crystallography* **63**, 550-554, doi:10.1107/S0907444907007652 (2007).
135 Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols* **4**, 363-371, doi:10.1038/nprot.2009.2 (2009).
136 Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic acids research* **38**, W545-549, doi:10.1093/nar/gkq366 (2010).

137 Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica. Section D, Biological crystallography* **60**, 2256-2268, doi:10.1107/S0907444904026460 (2004).
138 Wittinghofer, A. & Vetter, I. R. Structure-function relationships of the G domain, a canonical switch motif. *Annual review of biochemistry* **80**, 943-971, doi:10.1146/annurev-biochem-062708-134043 (2011).
139 Jameson, D. M. & Eccleston, J. F. Fluorescent nucleotide analogs: synthesis and applications. *Methods in enzymology* **278**, 363-390 (1997).
140 Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* **9**, 40, doi:10.1186/1471-2105-9-40 (2008).
141 Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164, doi:10.1126/science.252.5009.1162 (1991).
142 Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic acids research* **32**, W526-531, doi:10.1093/nar/gkh468 (2004).
143 Weis, K. Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* **112**, 441-451 (2003).
144 Vetter, I. R., Arndt, A., Kutay, U., Gorlich, D. & Wittinghofer, A. Structural view of the Ran-Importin beta interaction at 2.3 A resolution. *Cell* **97**, 635-646 (1999).
145 Conti, E., Muller, C. W. & Stewart, M. Karyopherin flexibility in nucleocytoplasmic transport. *Current opinion in structural biology* **16**, 237-244, doi:10.1016/j.sbi.2006.03.010 (2006).
146 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
147 Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research* **38**, W529-533, doi:10.1093/nar/gkq399 (2010).
148 Maiolica, A. *et al.* Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Molecular & cellular proteomics : MCP* **6**, 2200-2211, doi:10.1074/mcp.M700274-MCP200 (2007).
149 Vassilev, L. T. *et al.* Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10660-10665, doi:10.1073/pnas.0600447103 (2006).
150 Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299-1304, doi:10.1126/science.1062023 (2001).
151 Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods in enzymology* **375**, 23-44 (2004).
152 Guse, A., Fuller, C. J. & Straight, A. F. A cell-free system for functional centromere and kinetochore assembly. *Nature protocols* **7**, 1847-1869, doi:10.1038/nprot.2012.112 (2012).
153 Czarnota, G. J. & Ottensmeyer, F. P. Structural states of the nucleosome. *The Journal of biological chemistry* **271**, 3677-3683 (1996).
154 Chua, E. Y., Vasudevan, D., Davey, G. E., Wu, B. & Davey, C. A. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic acids research* **40**, 6338-6352, doi:10.1093/nar/gks261 (2012).
155 Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of molecular biology* **276**, 19-42, doi:10.1006/jmbi.1997.1494 (1998).
156 Barbera, A. J. *et al.* The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **311**, 856-861, doi:10.1126/science.1120541 (2006).
157 Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562-566, doi:10.1038/nature09321 (2010).
158 Armache, K. J., Garlick, J. D., Canzio, D., Narlikar, G. J. & Kingston, R. E. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 A resolution. *Science* **334**, 977-982, doi:10.1126/science.1210915 (2011).
159 Fitzgerald, D. J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nature methods* **3**, 1021-1032, doi:10.1038/nmeth983 (2006).
160 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186-190, doi:Doi 10.1107/S0021889809045701 (2010).
161 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
162 Morris, R. J. *et al.* Breaking good resolutions with ARP/wARP. *Journal of synchrotron radiation* **11**, 56-59 (2004).

163 Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica. Section D, Biological crystallography* **62**, 1002-1011, doi:10.1107/S0907444906022116 (2006).
164 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).