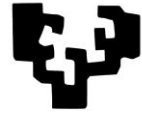


eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea



Structure-function relationships of tumor suppressor ING5

Georgina Ormaza Hernández



Tesis Doctoral, 2017

STRUCTURE-FUNCTION RELATIONSHIPS OF TUMOR SUPPRESSOR ING5

Departamento de Bioquímica y Biología Molecular
Facultad de Ciencia y Tecnología
Universidad del País Vasco/Euskal Herriko Unibertsitatea

Memoria presentada para optar al grado de Doctor por
Georgina Ormaza Hernández

Tesis dirigida por
Francisco J. Blanco

Realizada en
CIC bioGUNE

AGRADECIMIENTOS

Por fin ha llegado el momento...

Parece que fue ayer cuando empecé mi andadura por este camino lleno de duros esfuerzos y quebraderos de cabeza, pero también de alegrías, compañerismo y superación. Es con esto último con lo que me quedo y lo que me llevo y, es por ello, que no podía irme sin agradecerlo.

Quería agradecer a Paco la oportunidad de realizar mi tesis con él en el CICbioGUNE. Muchas gracias, Paco, por enseñarme a aprender, a desarrollar la visión crítica, por confiar en mí y en mi trabajo y, sobretodo, por esa cercanía y esa forma de hacer siempre bien las cosas.

Por supuesto, tengo que agradecer el no haber trabajado sola. Detrás de todo mi trabajo, se encuentran grandes compañeros y amigos con los que he tenido el placer de compartir poyata. Gracias a Neka y Mainer, por enseñarme todo lo que sé y ser mis “solucionadoras de vida” en momentos extremos, a Jhon Alex por su legado ING5, a Alain por su ayuda en todos los campos, pero sobretodo, en RMN y en el procesado “rápido” gracias a sus scripts maravillosos, y a Amaia la última en incorporarse, el aire fresco del grupo que enseguida se ha convertido en una buena amiga, tanto dentro como fuera del Biogune.

De aquí no me puedo ir sin mencionar a los Bioguners, tanto los que siguen aquí, como los que se han ido, y con los que he podido compartir mi vida dentro y fuera, con los que he pasado los mejores cafés de las 11, los mejores ratos, días, fines y vacaciones de mi vida. Gracias por ello también a los del principio Bego, Ander, Bea, Igor, Iker, Aberto, Mikel, June, Ben, Miguel, Albano y a los de después, Paula, Carlos, Elisa. Nunca habría podido ser mejor sin vosotros, y gracias por aguantarme y animarme en mis peores momentos.

Una pena perderme el gran futuro de la “Postgrad”, que ha conseguido hacer del bioGUNE una pequeña familia, tanto para los que ya estaban, como para los que han de venir. Gracias por llenar la semana de planes y, aunque es raro decirlo, por enseñarme cada rincón de Bilbao.

También quería dar las gracias a mis compis del 801, los Kypas y Vivancos, por esos cafés supermegarápidos, por ese apoyo cuando nada salía bien y por ayudarme hasta cuando no estaba en el labo ni en el país.

De nuestras colaboraciones quiero agradecer el apoyo y la ayuda tanto en el labo como fuera de él a mis compañeros, y más que compañeros, amigos del Biolab en el sincrotrone di Trieste, a Barbara, Adi, Marta, Giabba, Saša, Matteo, Alfredo y Silvia. Muchas gracias por enseñarme vuestro mundo y vuestra gastronomía.

Por último, quería agradecer a todas esas personas que han estado a mi lado fuera del laboratorio, mis grandes amigos los biolocos, mis amigas del txoko peter, a mi familia por ayudarme tanto y creer en mí y, por supuesto, Andrés, que con esa paciencia, hacía que todo se arreglase aunque pareciera imposible.

Eskerrik asko, muchas gracias, grazie mille.

RESUMEN

La familia ING (INhibitor of Growth) de supresores tumorales está formada por cinco proteínas homólogas que regulan el estado transcripcional de la cromatina por medio del reclutamiento de complejos remodeladores de cromatina a los lugares con la marca trimetilada en la histona H3 (H3K4me3). Esta modificación es reconocida por el dominio C-terminal conservado en las cinco proteínas ING, el Plant HomeoDomain (PHD). La unión a péptidos de histona trimetilados por ING5 ha sido caracterizada por RMN para la proteína completa mostrando una afinidad del mismo rango micromolar que la determinada para el dominio PHD aislado. El miembro de la familia ING4 dimeriza a través de su dominio N-terminal y se pliega de forma simétrica con estructura helicoidal de forma que es un lector bivalente de la marca H3K4me3. Los miembros ING4 e ING5 presentan una alta homología de secuencia en su dominio N-terminal y comparten regiones conservadas, motivos de secuencia y dominios estructurales. En esta tesis doctoral, mostramos que ING5 también es un dímero, tanto en solución como en células, y que, por lo tanto, también es un lector bivalente de la marca H3K4me3. Sin embargo, la estructura cristalográfica del dominio N-terminal de ING5 muestra una interfaz de dimerización distinta de la descrita para ING4, formando un dímero asimétrico. Resultados obtenidos mediante técnicas en solución, como RMN y SAXS, demuestran que ING5 forma dímeros simétricos como ING4, e indican que la estructura cristalográfica de ING5 es, probablemente, un artefacto de la cristalización. En esta tesis doctoral se ha caracterizado, por primera vez, la función de las regiones NLS de ING4 e ING5 como sitios de unión a ADN de doble hebra sin especificidad de secuencia de nucleótidos. Esta unión bivalente a ADN, junto a la unión bivalente a las colas trimetiladas de la histona H3, podría funcionar como un punto de anclaje reforzado para complejos remodeladores de la cromatina. Tres mutaciones puntuales en el dominio N-terminal de ING5 han sido descritas en cáncer de célula escamosa de boca: Q33R, I68V, and C75R. El análisis estructural de estos mutantes revela que son dímeros con estructura helicoidal al igual que ING5 nativa. Sin embargo, I68V y C75R están fuertemente desestabilizados, lo que sugiere un posible papel en el desarrollo de cáncer. El efecto funcional que tiene la expresión de estos mutantes en células NIH3T3 ha sido analizado mediante diversas técnicas de biología celular, mostrando variaciones en la sublocalización celular, en proliferación celular y en ciclo celular, siendo los mutantes I68V y C75R los de efectos más pronunciados.

SUMMARY

The INhibitor of Growth (ING) family of tumor suppressors consists of five homologous proteins that regulate the transcriptional state of chromatin by recruiting chromatin remodeling complexes to sites with histone H3 trimethylated at K4 (H3K4me3). This modification is recognized by the conserved Plant HomeoDomain (PHD) present at the C-terminus of the five ING proteins. ING5 binding to a trimethylated histone peptide has been characterized by NMR showing an affinity in the same micromolar range as its isolated PHD. ING4 dimerizes through its N-terminal domain, with a symmetric antiparallel coiled-coil structure, making it a bivalent reader of the H3K4me3 mark. ING4 and ING5 present a high sequence homology in their N-terminal domain, and they share conserved regions, sequence motifs and structural domains. Here, we show that ING5 is also a dimer, not only in solution but also in cells, and, therefore, it is also a bivalent reader of the H3K4me3 mark. However, the crystal structure of the N-terminal domain of ING5 shows a different dimerization interface from the previously described for ING4, forming an asymmetric dimer. Our solution NMR and SAXS data indicate that ING5 forms symmetric dimers as ING4, and suggest that the crystallographic structure of the N-terminal domain of ING5 is likely a crystallization artefact. In this thesis, for the first time, the specific role of the flexible NLS regions of ING4 and ING5 as DNA binding sites has been characterized, revealing that they bind dsDNA without nucleotide sequence specificity. The bivalent binding to DNA, together with the bivalent recognition of H3K4me3 marks, will result in a strong nucleosome binding and a stable scaffolding for recruited chromatin remodeling complexes. Three point mutations in the N-terminal domain of ING5 have been described in oral squamous cell carcinoma: Q33R, I68V, and C75R4. We have found that the N-terminal domains of the three mutants are dimeric coiled-coils. However, the I68V and C75R mutants are strongly destabilized, suggesting a possible role in cancer development. The functional effect of the expression of the mutants in NIH3T3 cells has been studied by different cell biology techniques, showing differences in the cell sublocalization, cell proliferation and in cell cycle progression, I68V and C75R being the most affected mutants.

INDEX

ABBREVIATIONS

INTRODUCTION

1. Chromatin Remodeling	3
1.1. Histone post-translational modifications: “ The Histone Code”	5
1.2. Chromatin Remodeling Complexes	7
1.3. Recognition of Histone modifications	8
2. Chromatin remodeling and Cancer	11
3. Inhibitor of Growth (ING) family of tumor suppressors	13
3.1. Sequence analysis and structural organization of ING proteins.....	14
3.2. INGs in H3K4me3 recognition by their Plant HomeoDomain	17
3.3. ING proteins are components of macromolecular acetylase/deacetylase complexes	18
3.4. ING4	20
3.5. ING5	22
4. Protein oligomerization in function regulation	24
OBJECTIVES.....	31

MATERIALS AND METHODS

1. Gene, plasmid, cloning and mutagenesis	
1.1. ING5, Nt, Nt mutants and PHD.....	35
1.2. ING4, Nt, PHD and ING4ΔNLS	36
1.3. JADE 1L IIb domain	36
2. Protein expression and purification	
2.1. Expression and purification of ING5, Nt, Nt mutants and PHD	38
2.2. Expression and purification of ING4, Nt, PHD and ING4ΔNLS.....	40
2.3. Expression and purification of JADE1L_IIb domain.....	42
3. Circular dichroism	42
4. Size Exclusion Chromatography-Multi angle static Light Scattering (SEC-MALS)	44
5. X-ray crystallography: Crystallization, Data Collection and Structure Determination. .	44
6. Small angle X-ray Scattering (SAXS)	48
7. Nuclear magnetic resonance (NMR) spectroscopy	50
7.1. Protein Backbone assignments	51
7.2. H3K4me3 peptide binding	52
7.3. Interactions with JADE-1L	53
7.4. DNA binding assays.....	54
7.5. Paramagnetic relaxation enhancement	54

8. Ion mobility coupled to electrospray ionization mass spectrometry (ESI-IM-MS) experiments under non denaturing conditions	55
9. Mammalian cell culture and transfection.	58
10. Cell proliferation assay	58
11. Cell Cycle Analysis	58
12. Immunoprecipitation and Western Blot analysis	59
13. Immunofluorescence	59
14. Fluorescence Electrophoretic Mobility Shift Assay (EMSA)	60

RESULTS

1. ING5 is a dimer and a bivalent reader of H3K4me3 mark	
1.1. Biophysical characterization of ING5	65
1.2. ING5 binds H3K4me3.....	67
2. The C-terminal PHD finger of ING5 recognises H3K4me3	68
3. DNA binding of ING4 and ING5 through the NLS disordered central region	72
4. The Nt domain of ING5 forms a symmetric coiled-coil dimer	78
4.1. Nt domain backbone NMR assignment.....	79
4.2. Paramagnetic relaxation effects.....	81
4.3. Structural analysis of ING5 Nt of three mutants designed to clarify the structural nature of the domain	83
4.4. ING5 Nt domain interaction with JADE1L	91
5. Oligomerization of ING5 inside living cells	
5.1. Homodimerization of ING5	93
5.2. Heterodimerization of ING5.....	93
6. Structural impact of ING5 N-terminal mutants detected in primary tumors	95
7. Functional impact of ING5 N-terminal mutants	
7.1. Cell morphology.....	98
7.2. ING5 N-terminal mutants sublocalization in cells	99
7.3. Cell proliferation effect of ING5 N-terminal mutants.....	99
7.4. Cell cycle profile of ING5 N-terminal mutants expressing cells	100
DISCUSSION.....	105
CONCLUSIONS	115
REFERENCES	119
APPENDIX	131

ABBREVIATIONS

ATP	Adenosine triphosphate
BRCT	Breast Cancer 1 C Terminus
BRPF	Bromodomain and PHD finger-containing proteins
BSA	Bovine Serum Albumin
CD	Circular Dichroism
CDK2	Cyclin dependent kinase-2
CHD1	Chromodomain Helicase DNA binding protein 1
CSI	Chemical shift index
CSP	Chemical shift perturbation
DAPI	4',6-diamino-2-phenylindole
D _{max}	Maximum particle diameter
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
ESI-IM-MS	Electrospray ionization Ion mobility mass spectrometry
FACS	Fluorescence-activated cell sorting
HAT	Histone Acetylase transferase
HDAC	Histone DeAcetylase Complex
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIF	Hypoxia Inducible Factor
HP1	Heterochromatin protein 1
HRP	Horseradish peroxidase
ING	Inhibitor of Growth
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KAT	Lysine Acetyl Transferases
KMT	Lysine methyltransferases
LB	Luria-Bertani broth
LID	Lamin interaction domain
LZL	Leucine Zipper Like
MALDI	Matrix-Assisted Laser Desorption/Ionization
MES	2-(N-morpholino)ethanesulfonic acid

MOPS	3-(N-morpholino)propanesulfonic acid
MALS	Multi Angle static Light Scattering
MCM	MiniChromosome Maintenance
MMP	Matrix metalloproteinase
MOZ	Monocytic leukemia zinc finger protein
MORF	MOZ related factor
MPD	2-Methyl-2,4-pentanediol
MTSL	S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonylthioate
NAD	Nicotinamide adenine dinucleotide
NLS	Nuclear Localization Signal
NMR	Nuclear magnetic resonance
NOE	Nuclear overhauser effect
NTS	Nucleolar translocation sequences
PAGE	Polyacrylamide gel electrophoresis
PBD	Partial BromoDomain
PDB	Protein data bank
PRE	Paramagnetic relaxation enhancement
PBR	PolyBasic Region
PBS	Phosphate-Buffered Saline
PCNA	Proliferating cell nuclear antigen
PCR	Potential chromatin regulatory
PEG	Polyethylene glycol
PHD	Plant HomeoDomain
PIP	PCNA-interacting protein
PRMT	Protein arginine methyltransferases
PTM	Post-translational covalent modification
RCI	Random coil index
R _g	Radius of gyration
RIPA	Radioimmunoprecipitation assay
SAXS	Small angle X-ray Scattering
SDS	Sodium dodecyl sulphate
TEV	Tobacco Etch Virus cysteine protease
TOF	Time-of-flight
TROSY	Transverse relaxation optimized spectroscopy
TSG	Tumor suppressor gene
WT	<i>Wild type</i>

INTRODUCTION



1. Chromatin Remodeling

The nuclear DNA of eukaryotic cells is packed with proteins into chromatin, a highly organized structure in which the nucleosome is the fundamental unit. In the nucleosome, two copies of each of the four core histones (H2A, H2B, H3 and H4) form an octamer in which two superhelical turns of 147 bp DNA are wound around for compaction [1] (**Figure 1**).

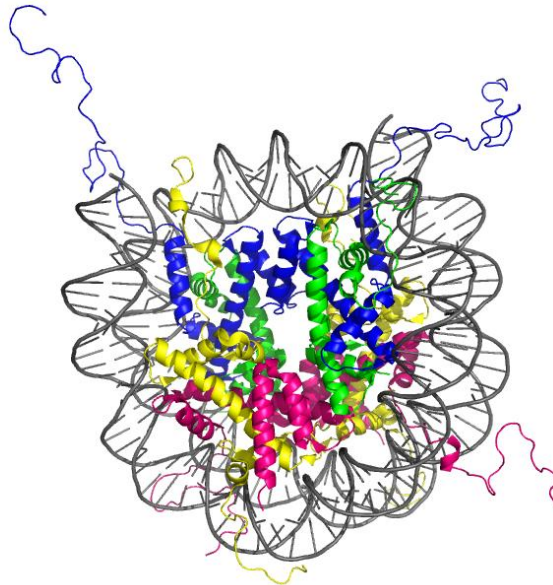


Figure 1. Model of the nucleosome core particle. Each nucleosome core particle consists of double-stranded DNA wrapped around two molecules of each of the four core histone proteins: H2A (yellow), H2B (magenta), H3 (blue) and H4 (green). The N-terminal tails of histone proteins protrude from the centre of the nucleosome core particle (Protein Data Bank entry 1kx5).

Histones are highly conserved proteins in eukaryotic organisms that have a C-terminal globular domain, necessary for oligomerization and nucleosome scaffold formation, and an N-terminal domain that is flexible and protrudes outward from the nucleosome core [2]. Those flexible N-terminal tails provide internucleosomal linkages and, together with other proteins, allow for higher levels of compaction of the chromatin [3] (**Figure 2**). These multiple protein-DNA interactions provide the nucleosome with the stability necessary for its packaging function.

The chromatin can adopt two different forms: heterochromatin, characterized by a high level of compaction and a low accessibility of genes, and euchromatin, which has a lower level of compaction and genes are more accessible. However, the nucleosome cannot be conceived as a simple static unit since it possesses dynamic properties regulated by various protein complexes related with DNA replication, repair and transcription [4, 5]. During these processes the nucleosome can act as a dynamic mechanical barrier in which the nucleosomal DNA spontaneously unwraps and rewraps

the histone core on a timescale of a few milliseconds, regulating the DNA accessibility for DNA-binding proteins and DNA processing enzymes [6]. This is achieved by the eukaryotic cell thanks to different and well-regulated mechanisms of chromatin remodeling.

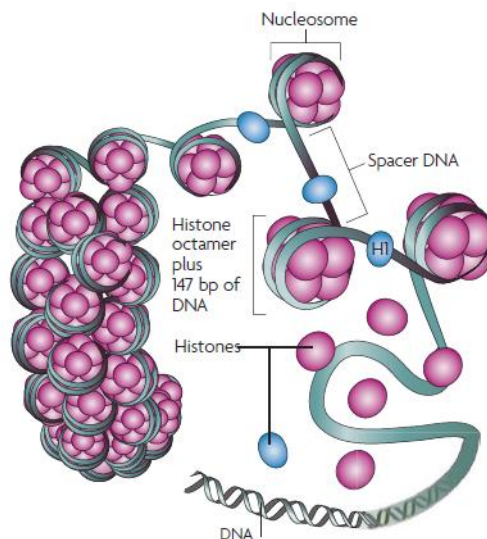


Figure 2. Structure of the chromatin. In the Eukaryotic cells the DNA is packed into chromatin, the nucleosome being its fundamental unit. Each nucleosome is composed of approximately 147 bp of DNA wrapped around an octamer of histones (two copies of H2A, H2B, H3 and H4, colored pink). Histone H1 (colored blue) binds to the DNA that links two nucleosomes allowing the higher levels of compaction of DNA into chromatin. (Figure obtained from Figueiredo *et al* 2009[7]).

The structural, mechanical and functional properties of nucleosomes have been related to the nucleosomal DNA sequence itself, which has an influence on the nucleosome positioning [8]. Differential methods have allowed analyzing relative free energies for nucleosome formation on one DNA sequence compared to another, showing that while nucleosomes can be formed with any DNA sequence, the relative affinities of certain DNA sequences are high compared to the affinities for arbitrary DNA sequences. The best known mechanism that also influences the chromatin structure and dynamics is the post-translational covalent modification (PTM) of histones within the nucleosome [9, 10]. Complementary to the chemical modification of histones that reorganize the nucleosome structures we find ATP-dependent chromatin remodelers that catalyze changes in nucleosome position and composition by inducing nucleosome sliding or partial disassembly/assembly of histones upon ATP hydrolysis [11]. Other factors may also influence the nucleosome structure and dynamics. For example, certain enzymes have to exert forces and torques in order to perform their tasks in processing the genome [12]. Besides, the existence of intermediate states in the nucleosome assembly process has been reported [13, 14] and related to changes in the structure and dynamics of chromatin

the genome [17]. For example, methylation at Lys4 of histone H3 is associated with regions active in transcription. Lysine methylation is one of the most stable histone marks, and the lysine can be monomethylated (me1), dimethylated (me2), and trimethylated (me3) at the amine group keeping its positive charge at physiological pH [20]. These modifications are carried out by lysine methyltransferases (KMTs) [21] and removed by histone demethylases [22]. Histone arginine residues can also be monomethylated (me1) and dimethylated (me2) in the guanidinium group (keeping its positive charge) by protein arginine methyltransferases (PRMTs) [23]. Dimethylation can occur in a symmetric (me2s on two separated nitrogen atoms) or asymmetric manner (me2a on the same nitrogen atom) [17]. Very little is known about arginine methylation relationship with chromatin dynamics. There are evidences reviewed by Tessarz and Koutzarides on 2014 [24], in which the asymmetric dimethylation at H3R42 positively correlates with transcription in vitro [25] by decreasing the nucleosome stability. The readout of methylation by reader modules regulates the chromatin dynamics and the transcriptional outcome [26]. Lysine methylation recruits “reader” and “eraser” modules, called respectively HATs (Histone Acetylase transferases) and HDACs (Histone DeAcetylase Complexes), to sites in the chromatin with this histone mark. These complexes introduce or remove acetylation marks from histone tails.

Lysine acetylation plays a vital role in regulating chromatin structure and transcriptional activity [17]. Histone acetylation neutralizes the positive charge of lysine residues, weakening the interactions between histones and nucleosomal DNA, linker DNA or adjacent histones, increasing accessibility of DNA for the transcriptional machinery [17]. Histone acetylation by HATs is generally related to transcriptional activation, whereas histone deacetylation by HDACs, promotes chromatin condensation and transcriptional repression [27]. On the contrary, histone deacetylation will repress transcription through an inverse mechanism [28].

Another important PTM of histones is the phosphorylation, which is introduced on serine, threonine and tyrosine residues by kinases; while this modification can be removed by phosphatases [17]. Phosphorylation has a similar role to acetylation in modulating nucleosome dynamics [29]. However, in mammalian cells, histone H3 phosphorylated at Ser10 have been related to two effects; is necessary for the high condensation of chromatin during mitosis, while it is also important in the activation of transcription [30]. Phosphorylation of histone H3 creates combinatorial patterns with lysine acetylation and methylation marks that are read by specific histone readers [17].

Ubiquitylation and sumoylation are large PTMs in which the C-terminus of ubiquitin or SUMO are covalently attached by a peptide bond to the ϵ -amino group of lysine residues.

The sequential and combinatorial nature of all these histone modifications, and their correlation with many chromatin-related processes led to the proposal of the existence of a “histone code” as the basis of a fundamental regulatory mechanism that extends the information potential of the genetic code [31]. Interestingly, more recent data has revived an old idea that histone modifications, and in particular histone “core” modifications, also have a direct effect on chromatin architecture by affecting histone-DNA and histone-histone interactions, as well as histone interactions with chaperones [24]. One of the best-studied histone core modifications with potential effect on histone-DNA and histone-histone interactions is the acetylation of Lys56 of H3. This histone modification has been shown to enhance the unwrapping of the DNA at the entry-exit site of the nucleosome [32]. H3K56 acetylation also influences chromatin structure at a higher order level by regulating tertiary contacts to keep nucleosome deprived chromatin regions accessible [33]. Acetylation can also directly affect the stability of the histone octamer as in the case of H4K91. This is the only modification described to date that regulates nucleosome stability by affecting histone-histone interactions. This residue is situated in the H3-H4 and H2A-H2B interaction surface, and acetylation decreases the association of H2A-H2B dimers with chromatin leading to nucleosome instability [34].

1.2. Chromatin Remodeling Complexes

There is at present much investigation about the detailed mechanism by which cells unravel the “histone code”. There currently are two models describing how this is carried out, the “direct” and the “effector-mediated” models. In the direct model post-translational modifications have a direct effect on chromatin compaction and this is the case of phosphorylation or acetylation. On the other hand, in the effector-mediated model, post-translational modifications are recognized by protein modules, facilitating a downstream signaling that recruits or stabilizes chromatin-related machinery. Thereby, histone tails are modified by specific enzymatic complexes of proteins, some of them with domains that recognize one or more of the possible modifications [35]. These epigenetic regulators can be divided into three different groups based on their functions (**Figure 4**). Epigenetic “writers” introduce modifications in histone tails, such as acetyl and methyltransferases.

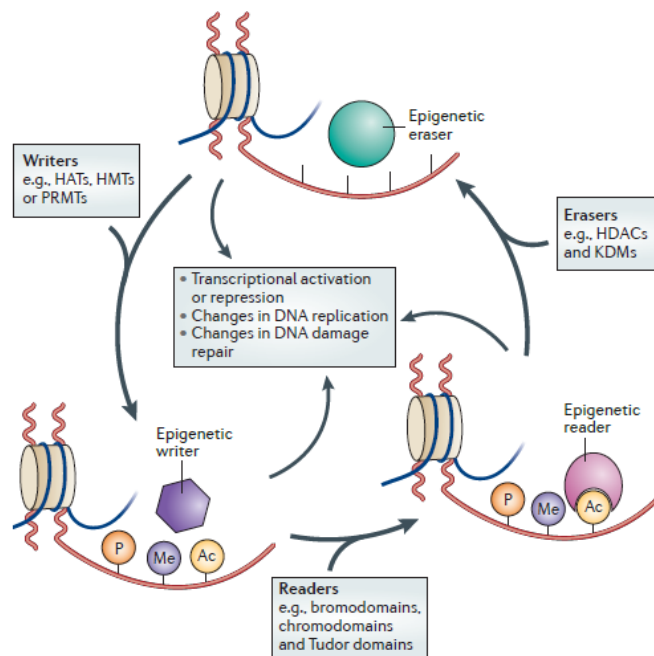


Figure 4. Epigenetic writers, readers and erasers. Epigenetic writers lay down epigenetic marks on amino acid residues on histone tails. Epigenetic readers with specialized domains bind to these epigenetic marks. Epigenetic erasers catalyse the removal of epigenetic marks. Addition and removal of these post-translational modifications of histone tails lead to the addition and/or removal of other marks in a highly complicated histone code. Together, histone modifications regulate various DNA-dependent processes, including transcription, DNA replication and DNA repair. Abbreviations: HAT, Histone AcetylTransferase; HMT, Histone MethylTransferase ; PRMT, protein Arginine MethylTransferase ; HDAC, Histone DeAcetylase; KDM, Lysine DeMethylase. Figure from Falkenberg *et al* 2014 [36].

These epigenetic regulators can be divided into three different groups based on their functions (**Figure 4**). Epigenetic “writers” introduce modifications in histone tails, such as acetyl and methyltransferases. These epigenetic marks can be removed by epigenetic “erasers”, such as deacetylases and demethylases [36, 37]. However, for these protein complexes to erase or write some modifications they all must have “reader” domains, which specifically recognize the histone modifications, recruiting writer and eraser protein complexes to particular chromatin positions, initiating signaling responses as transcription activation/inhibition. The combination of the different PTMs, together with the existence of a variety of “reader” domains in a protein or a protein complex, defines the histone code hypothesis.

1.3. Recognition of Histone modifications

The histone code “readers” are specific domains that recognize histone post-translational modifications. The recognition and deciphering of the histone code is done in a residue and a particular modification dependent manner. There are different groups

of “reader” domains including: bromodomains, Chromodomains, PHD (Plant HomeoDomain), BRCT (Breast Cancer 1 C Terminus) and Tudor domains (**Figure 5**).

- **Bromodomain**

Bromodomain is an acetyl-lysine binding module found in chromatin-associated complexes as nuclear Histone Acetyltransferases (HATs). These domain folds into a left-handed four helix bundle with a hydrophobic pocket that recognizes acetylated histones H2A, H2B, H3 and H4. Examples of HATS that contain this domain are Gcn5p that recognizes H4K16ac [38] (**Figure 5A**) and TAF-1 that contains two bromodomains that recognize H4K5acK12ac [39].

- **Chromodomain**

Chromodomain is a sequence motif implicated in methylated-lysine histone recognition. It folds in an antiparallel β -sheet crossed by one α -helix and the methylated tail of histone H3 binds as a β -strand in a conserved groove on the surface of the chromodomain, where the residues at the loops connecting the β -strands are the ones involved in methyl-lysine binding. Examples of proteins with Chromodomains are HP1 (heterochromatin protein 1) (**Figure 5B**), Polycomb proteins, and CHD1 (chromodomain Helicase DNA binding protein 1). While HP1 and Polycomb proteins use canonical chromodomains for the recognition of H3K9me_{2/3} and H3K27me₃, respectively [40, 41], CHD1 has two chromodomains that are both required and cooperate for the recognition of H3K4me₃ [42, 43]. In the case of CHD1 histone H3 binds to a groove at the inter-chromodomain junction [43].

- **Plant HomeoDomain**

The PHD is known to be involved in protein–protein interactions and is frequently present in proteins associated with chromatin remodeling functions [44], so it is commonly found in all eukaryotic genomes [45]. This domain, of about 60 amino acids, consists of a a two-stranded antiparallel β -sheet and a C-terminal α -helix stabilized by two Zn²⁺ cations with a Cys₄-His-Cys₃ motif [45] (**Figure 10B**). The PHD finger is the recognition module of histone H3 methylation state in Lys₄ (K4me_{0/1/2/3}) with different affinities [46] (**Figure 5C**), although there is published data showing that the PHD of SUMO ligase Siz1 in rice (OsSiz1–PHD) is able to recognize Arg₂ asymmetric

dimethylation in histone H3 tail (H3R2me2a) [47] and the PHD of DPF3b recognizes acetylation state of H3K14 (H3K14ac) [48].

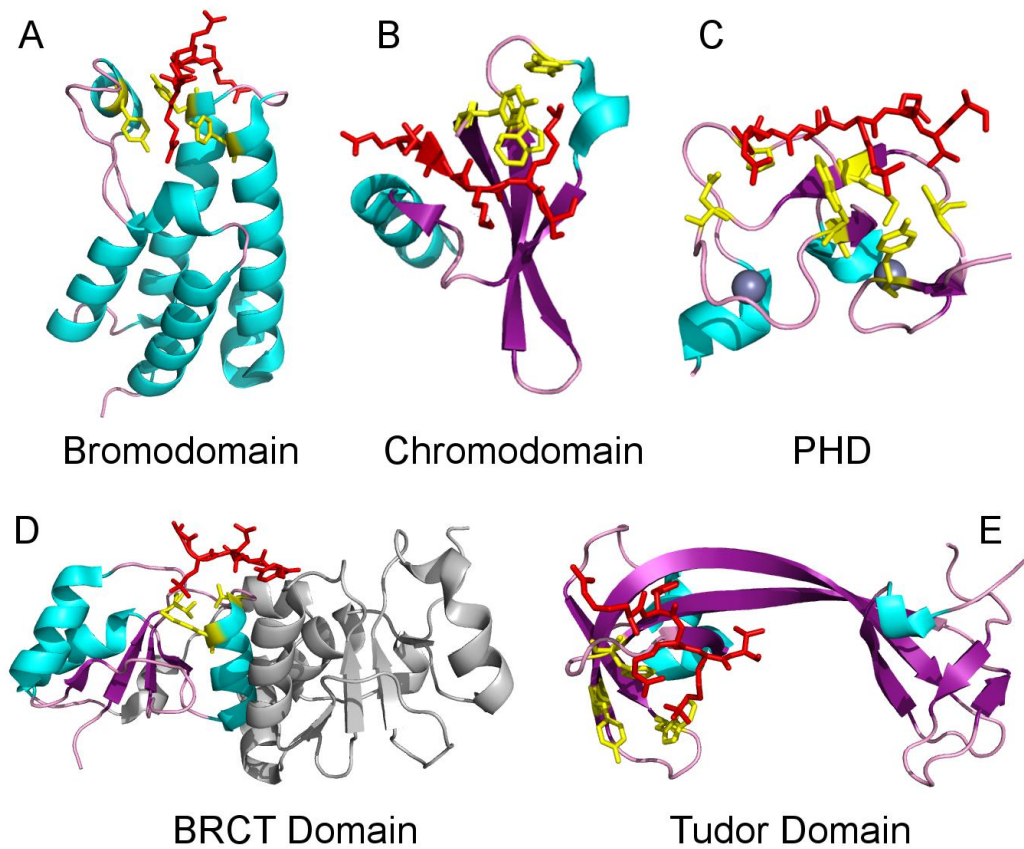


Figure 5. Structure models of the best characterized “reader” domains bound to modified histone tails. **A.** Bromodomain of Gcn5 bound to H4K16ac (PDB: 1E6I). **B.** Chromodomain of HP1 bound to H3K9me3 (PDB: 1KNA). **C.** PHD finger of ING5 bound to H3K4me3 (PDB: 3C6W). **D.** BRCT repeat of MDC1 bound to H2A.X S139ph (PDB: 2AZM). **E.** Tudor domain of JMJD2 bound to H4K20me3 (PDB: 2QQS). Side chains involved in the binding are displayed in yellow, while bound histone peptide are in red color.

- **BRCT (BRC1 C-Terminus) domain**

The BRCT domain functions as a phosphoserine/phosphothreonine-binding module. Its fold consists of a central four stranded β -sheet flanked by a single α -helix on one side and two α -helices on the opposite side [49] (**Fig. 5D**). The phosphopeptide binds in an extended conformation to a hydrophobic groove located in a highly conserved interface between the N- and C-terminal regions of the BRCT [50]. The BRCT domain may occur as isolated individual domain or as multiple tandem BRCT repeats, as it happens in several proteins involved in cell cycle regulation and DNA damage response [51]. One example of a protein with this domain is BRCA1 (Breast Cancer type 1 susceptibility protein), a tumor suppressor protein related to breast and ovarian cancer [51].

- **Tudor domain**

Tudor domain is a methyl-lysine and methyl-arginine reader. Four antiparallel β -strands form the core of the Tudor fold, which has a barrel-like structure with aromatic residues to accommodate methylated ligands [52] (**Figure 5E**). Tudor domain binds symmetric dimethyl-arginines [53]. With a lower affinity it has also been found to bind monomethyl-arginine and asymmetric dimethyl-arginines [52]. This domain was shown to have a unique property of binding trimethylated peptides from two different histone methylated sequences, H3K4me3 and H4K20me3, in the case of the lysine demethylase JMJD2A [54].

2. Chromatin remodeling and Cancer

Cancer is widely described as a complex genetic disease. In a multistep process a cell accumulates mutations leading to gene activity malfunction and to a transformation of a normal cell into a malignant phenotype that provides selective advantages [55, 56]. The mechanisms that lead to the transformation of a normal cell into a cancer cell are very different and it is necessary that a few of them converge for a cancer cell to escape from the control check points [55]. A cancer cell typically has defects in regulatory systems that govern the normal cell proliferation and homeostasis. It has been suggested that cancer genotypes are the result of ten essential alterations, “the ten hallmarks of cancer”, in cell physiology that collectively dictate malignant growth: sustaining proliferative signalling, evading growth suppressors, avoiding the immune destruction of malignant cells, enabling replicative immortality, tumor-promoting inflammation, activating invasion and metastasis, inducing angiogenesis, generation of genome instability and mutation, resisting cell death and deregulating cellular energetics [57] (**Figure 6**).

There are key genes that are usually altered in this process, the so called oncogenes and tumor suppressor genes. Oncogenes are the mutated version of a proto-oncogene, a group of genes that cause normal cells to become cancerous when they are mutated. Oncogenes induce hyperactive growth and division, protection against programmed cell death, and the ability to invade diverse tissue environments [58, 59]. Tumor suppressor genes (TSG) encode proteins that are involved in the regulation of many cellular processes as DNA repair, cell signaling, cell cycle arrest, senescence and apoptosis. They are classified into two different groups: Type I TSG, the “caretakers” of the genome, that usually are DNA repair genes which protect the genome from mutations; and Type II TSG, the “gatekeepers”, which prevent cancer through the control of cell growth [60]. In

the latter, we find proteins that are directly implicated in chromatin remodeling and gene expression regulation as proteins of the Inhibitor of Growth (ING) family [61].

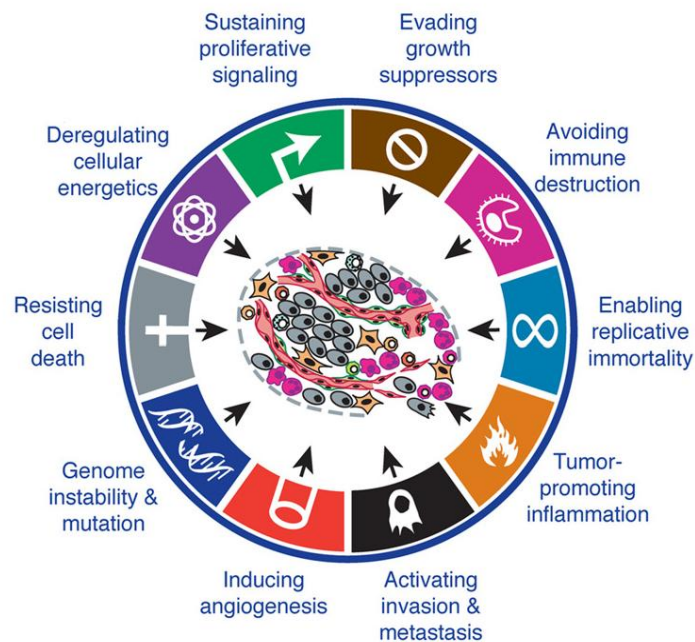


Figure 6. The ten Hallmarks of Cancer. Acquired capabilities of cancer that lead to essential alterations in cell physiology that collectively dictate malignant growth [55].

Alterations in the “histone code” and dysregulation of their “readers”, “writers”, and “erasers” are related to oncogenesis. Despite the difficulty of establishing a causal role of histone modifications in cancer, the global alterations of histone modification patterns appears to be linked to tumorigenesis [16]. For example, the global loss of trimethylation of histone H4 at K20 and a loss of acetylation at K16 have been observed to be associated with hypomethylation of DNA repetitive sequences, a known characteristic of cancer cells [62]. Moreover, histone serine phosphorylation is crucial for DNA damage repair, chromosome stability and apoptosis. Histone variant H2A.X phosphorylation is necessary for a normal radiation resistance and accumulation of DNA-damage response proteins for DNA repair [63], while histone H3S10 phosphorylation is related to chromosomal stability and aneuploidy during mitosis [64], and histone H2B is phosphorylated in S14 in response to apoptotic stimuli [65]. Although no other global histone epigenetic mark has been directly related to oncogenesis, impairments of chromatin remodeling machinery have also been reported [16]. Apart from that, disturbance of the balance of the acetylation status of chromatin through the disruption of HAT or HDAC activity can lead to tumorigenesis [66, 67]. Proteins of the Inhibitor of Growth (ING) family belong to this type of complexes.

3. Inhibitor of Growth (ING) family of tumor suppressors

The Inhibitor of Growth (ING) family of tumor suppressors consists of five homologous proteins (ING1-5) which play important roles in preventing tumor formation and progression [61, 68]. Members of the ING family are classified as type II tumor suppressors [60, 69] and are generally down regulated or even missing in different kinds of tumors, leading to a neoplastic growth. The founding member of the family, ING1, was discovered back in 1996 thanks to a new hybridization technique based on suppressive subtractive hybridization and selection of tumor suppressor elements when comparing normal cells and epithelial breast cancer cell lines from mice. The identified gene expressing ING1 was found to block proliferation in preneoplastic mammalian epithelial cells when it was overexpressed, being proposed as a tumor suppressor gene [70]. The rest of the members of the conserved ING family were identified progressively using sequence analysis showing a shared sequence homology between 32% and 76% [61, 71]. The sequence analysis and phylogenetic studies revealed that these genes are conserved from yeast to humans [72], that is, only present in eukaryotic organisms, highlighting the involvement of this family of proteins in nuclear processes. The localization of the five human genes of the ING family have been mapped in different chromosomes, and ING1, ING2, ING4 and ING5 are located in sub-telomeric regions while ING3 is located in the long arm of chromosome 7 [72]. Since ING1, ING2, ING4 and ING5 are located close to telomeric regions in the chromosome, the dysregulation of these proteins due to telomere aging could be one of the explanations for tumorigenesis [71]. There are different transcript variants for the ING genes produced through alternative mRNA splicing [72], but all ING proteins contain in their sequence a NLS (Nuclear Localization Signal) [68] and, in addition, several of ING members also contain a Nucleolar Translocation Sequence (NTS) within the NLS, which has been shown to translocate them to the nucleoli in response to DNA damage [73]. The mutation, alteration of the expression levels, and the subcellular localization of ING proteins have been found in several human cancer types. The characteristic pathways by which ING family proteins differentially affect the Hallmarks of cancer, and the various epigenetic mechanisms by which they regulate gene expression have been summarized in **Figure 7** [74]. In the figure, the different HAT and HDAC complexes of which the ING proteins form part (see below) are represented.

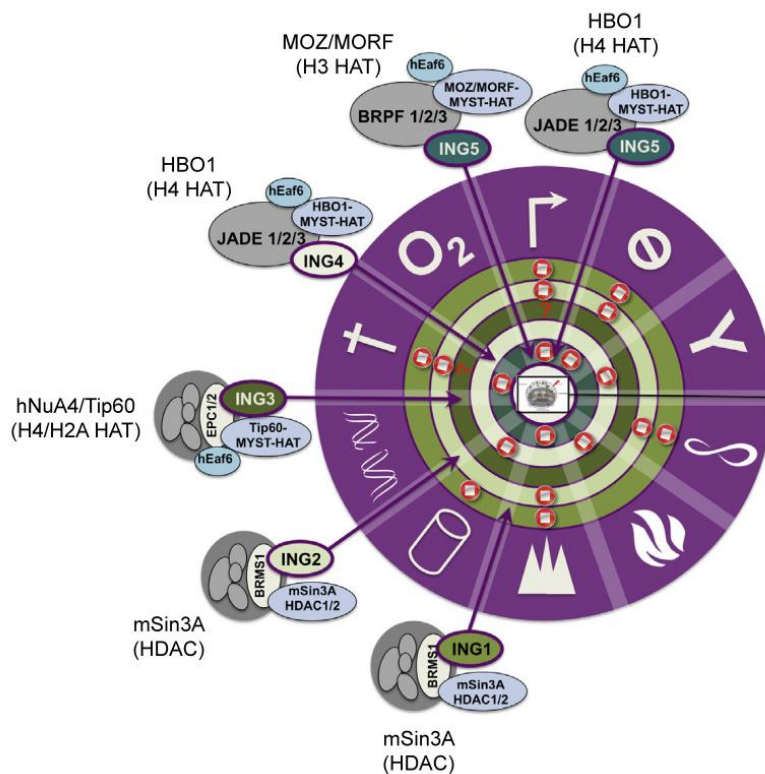


Figure 7. The hallmarks of cancer affected by ING HDAC and HAT complexes. White symbols in the purple ring are the ten hallmarks of cancer from Hanahan and Weinberg 2011 and summarized in figure 6. The red “stop”-signs mean that those hallmarks are inhibited by ING activity, and the question-mark indicates possibly inhibited by ING activity. For instance, ING5 has been found to inhibit aberrant proliferative signaling, dysregulated growth suppression, invasion & metastasis, resistance to cell death, Figure adapted from Tallen & Riabowol 2014 [74].

3.1. Sequence analysis and structural organization of ING proteins

The alignment of the amino acid sequences of the five members of the ING family shows three different regions: a conserved N-terminal domain, a non conserved central region containing the nuclear localization signal (NLS) and a highly conserved C-terminal region corresponding to the Plant HomeoDomain (PHD) (**Figure 8**).

The N-terminal region contains a Leucine Zipper Like (LZL) domain, rich in conserved leucines and isoleucines spaced seven amino acids apart, that can form a hydrophobic patch when folded into an α -helix [75]. This motif is present in all INGs except for ING1 and was proposed to be a homodimerization or heterodimerization region with INGs or other proteins. The PIP (PCNA-interacting protein) is a domain that is unique to ING1b and allows ING1 to bind to PCNA (proliferating cell nuclear antigen) in a DNA damage response [76]. Then, ING1 also presents a Partial BromoDomain (PBD), a motif which has been shown to interact with SAP30 of the Sin3-HDAC1 and HDAC2 complexes [74].

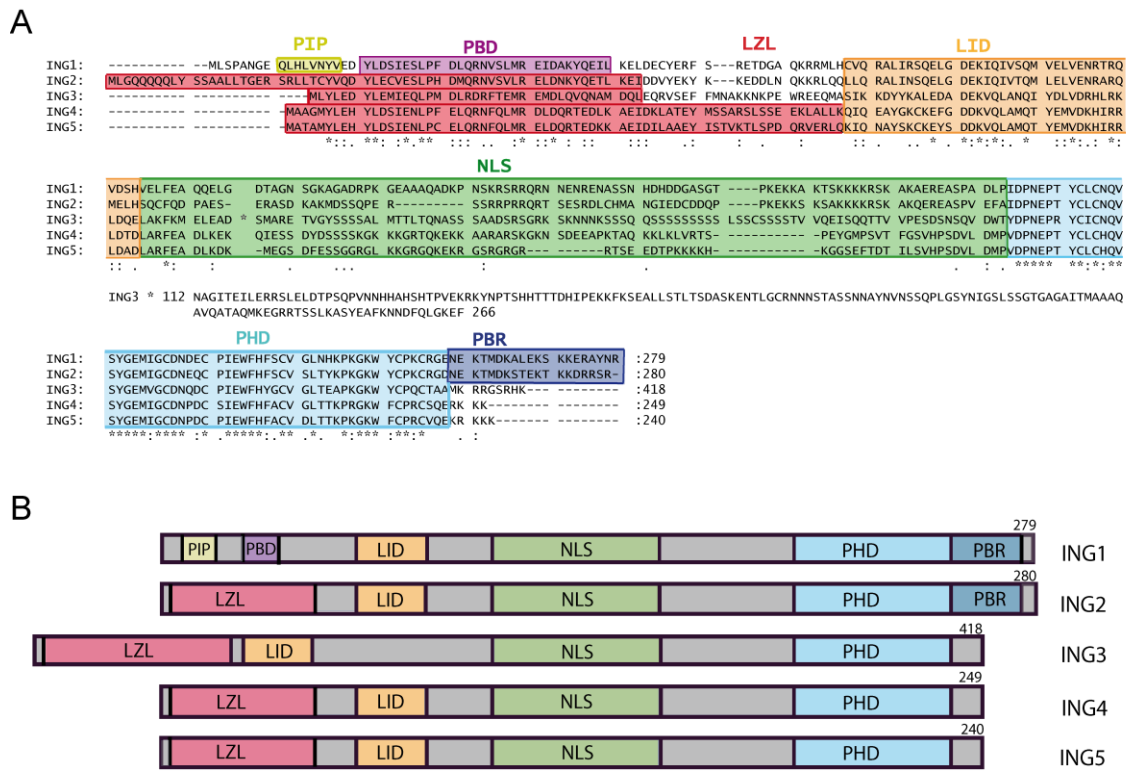


Figure 8. A. Sequence alignment of the human ING proteins. Sequences can be generally divided in three regions, which reflect structural motifs: a highly conserved N-terminal domain formed by LZL (red) and LID (orange) sequences (PIP in yellow and PBD in purple sequences are only present in ING1), a non conserved central region containing the NLS (Nuclear Localization Signal, green), and a highly conserved C-terminal region, containing a Plant HomeoDomain (PHD) (cyan). With a PolyBasic Region (PBR, blue) only in the case of ING1 and ING2. For ING1 the most frequently observed p33ING1b splicing variant was used. For the rest the full sequence is used. The alignment was performed using Clustal W. Asterisk at residue 112 of ING3 indicate non aligning sequence (residues 112-266). **B.** Scheme of the most representative sequence motifs of INGS.

All INGS have a lamin interaction domain (LID) in the N-terminal region [74]. This region was shown to bind lamin A and is believed to help tether ING1 in the nucleus, thereby localizing its functions [77]. This region was first referred to as potential chromatin regulatory (PCR) region by He *et al* 2005 and it was hypothesized to be involved in binding HAT/HDAC complexes during chromatin remodeling [72].

The central region of all INGS, rich in basic residues, contain a bipartite Nuclear Localization Signal: two clusters of basic amino acids, separated by a spacer of about 10 amino acids [68] (**Figure 9**). In addition, several of the ING proteins also contain short, highly basic nucleolar translocation sequences (NTS) within the NLS. However, these NTS have been only shown to be active for ING1, translocating it to the nucleoli in response to DNA damage for apoptosis activation [73]. These NTS are not well defined in ING3, ING4 and ING5.

```

ING1:  PNSKRS RRQR NNENRENASSNHDDDGASGTP KEKK AK-TS KKKK RSKAKA
ING2:  RSSRRP RRQR TSESRDLCHMANGIEDCDDQPP KEKK S-KSA KKKK RSKAKQ
ING3:  ADSRSG RKSK NNNKSSSQSSSSSSSSSS-LSS CSSS ST-VVQEISQQTTVVPE
ING4:  KKGRTQ KEKK AARARSKGKNSDEEAPKT-AQ- KKLK LVRTSPEYGMPSVTFGS
ING5:  LKKGRG QKEK RGSRGRGRRTSEED--TP---- KKKK HKGG----SEFTDTILS

```

Figure 9. Amino acid sequence alignment of the conserved bipartite NLS regions from each ING protein. Highlighted residues correspond to the related nucleolar translocation sequences.

At the C-terminal end all INGs have a PHD (Plant HomeoDomain) zinc finger, which is the most highly conserved feature of the ING family. The PHD is typically folded into a Cys4-His-Cys3 signature with characteristic cysteine spacing (**Figure 10**). The PHD finger of INGs directly bind histone N-terminal methylated tails with different affinities depending on the methylation state. The first time in which ING proteins were described to recognize histone post-translational modifications and, therefore, to be histone code “readers” was in 2006 by Peña *et al* and Shi *et al* [78, 79].

At the very end of the C-terminal region of ING1 and ING2, there is a PBR (PolyBasic Region) necessary for recognition of phosphoinositides in DNA damage signalling [80, 81]. The corresponding region in the rest of INGs is also rich in basic residues but notably shorter and, in the case of ING4 and ING5, they could work as additional nuclear localization signals [82].

The presence of different conserved domains and motifs in ING proteins makes them suitable to carry out many functions depending on their interaction with other partners. In this way, ING proteins are involved in a wide number of chromatin related events in the cell (**Figure. 11**).

A

```

ING1:  IDPNEPT Y C L C NQVSYGEMIG C DNDE C PIEWF H FS C VGLNHKPKGKWKY C PK C RGE
ING2:  IDPNEPT Y C L C NQVSYGEMIG C DNEQ C PIEWF H FS C VSLTYKPKGKWKY C PK C RGD
ING3:  YDPNEPR Y C I C NQVSYGEMVG C DNQD C PIEWF H YG C VGLTEAPKGKWKY C PQ C TAA
ING4:  VDPNEPT Y C L C HQVSYGEMIG C DNPD C SIEWF H FA C VGLTTKPRGKWF C PR C SQE
ING5:  VDPNEPT Y C L C HQVSYGEMIG C DNPD C PIEWF H FA C VDLTTKPKGKWF C PR C VQE

```

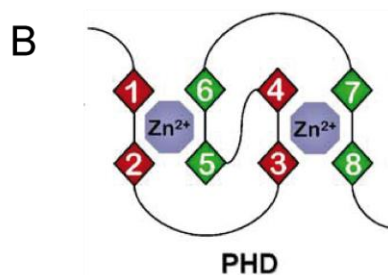


Figure 10. Amino acid sequence alignment of the INGs PHD motif. Highlighted residues in red correspond to the characteristic the zinc-binding cysteine and histidine residues that define the PHD motif. Modified from Russell 2006 [68].

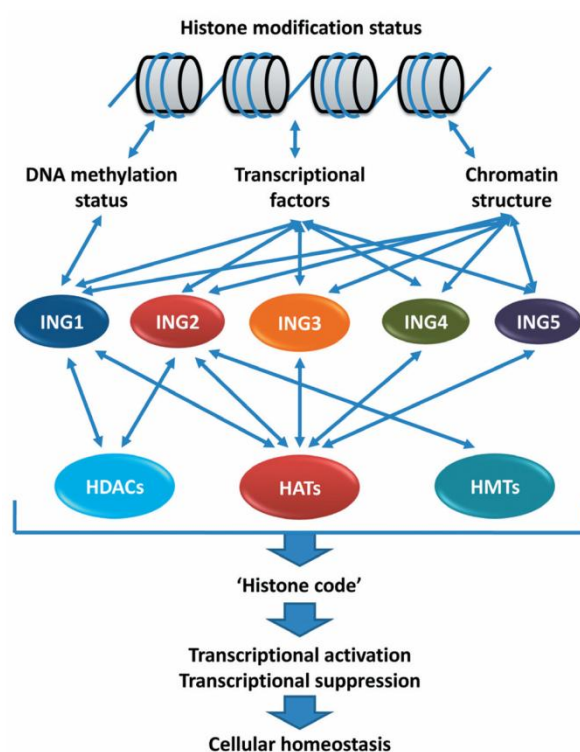


Figure 11. Network of interactions and functions of ING proteins [83].

3.2. INGs in H3K4me3 recognition by their Plant HomeoDomain

INGs regulate the transcriptional state of chromatin by recruiting remodeling complexes to sites with histone H3 trimethylated at K4 (H3K4me3). This mark is recognized by their C-terminal Plant HomeoDomain (PHD) [79]. Indeed, the structures of the PHD fingers of ING1 [84], ING2 [78], ING4 [85] and ING5 [86] bound to H3K4me3 peptides have been determined, displaying similar features with little differences in the conformation of the bound peptide [85]. The recognition of H3K4me3 by PHDs appears to be very specific since the interaction with unmodified H3K4 or other methylated lysine residues occurs with a much lower affinity [46]. The first member of ING family to be studied on their functionality as histone H3 methylation reader was ING2 which is related to transcription repression due to the fact that forms part of HDAC complex (mSin3/HDAC1/2) [79]. ING2 specifically recognizes H3K4me3 ($K_D = 1.5 \pm 1 \mu\text{M}$) with a higher affinity than H3K4me2 ($K_D = 15 \pm 4 \mu\text{M}$) and H3K4me1 ($K_D = 208 \pm 80 \mu\text{M}$); however, it does not recognize non methylated H3K4 ($K_D = 2240 \pm 350 \mu\text{M}$) [78]. In the complex formation, H3K4me3 peptide forms a third anti-parallel β -strand that pairs with the already existing double stranded β -sheet of the PHD finger. The binding region forms a pocket with two grooves connected by a narrow channel. The trimethylated Lys4 occupies the elongated groove formed by conserved hydrophobic and aromatic residues, whereas Arg2 is accommodated by the adjacent groove [46]. The H3K4me3 tail, which

involves the first six N-terminal residues, forms an extensive network of H-bonds and complementary surface interactions with the PHD that stabilize the complex [78]. The small size of the channel between grooves is probably involved in the specificity and unique recognition of the ARTK(me3)QT sequence in histone H3 tail by the PHD relative to other lysine methylation in the same or other histone tails. The presence of a bulky Arg or a Val in place of Thr3, as it happens for other methylated sequences, will disrupt the binding [46]. In the following years more information about the complex formation between the PHD of ING1, ING4, ING5 and H3K4me3 revealed that the peptide occupies analogous binding sites in all structures [84-86].

3.3. ING proteins are components of macromolecular acetylase/deacetylase complexes

Histone PTMs affect the compaction state of chromatin and, among them; acetylation has the highest potential to unfold chromatin, removing the basic charges of lysines. Histone tail acetylation is often associated with gene transcription activation and DNA repair and replication [37]. The acetylation state of chromatin is the result of the precisely regulated activity of protein complexes with acetylase or deacetylase activity, HATs (Histone acetyl transferases) or HDACs (Histone DeAcetylase Complexes) respectively.

The HDACs are a group of enzymes that catalyze the deacetylation of lysine residues in histones, allowing interaction between negatively charged DNA and positively charged histone proteins, which can result in heterochromatin formation and transcriptional silencing of genes [87]. In addition, HDACs have a number of non-histone proteins as substrates including proteins of transcription complexes that have a role in regulating gene expression, and proteins in pathways that regulate cell proliferation, cell migration, cell death and angiogenesis [88]. HDACs can be classified into eighteen groups, according to the functional and the phylogenetic, and they are subdivided into Zn²⁺-dependent (class I, II and IV), Zn²⁺-independent and NAD-dependent (class III) enzymes [89].

The human HATs can be classified as lysine (**K**) **A**cetyl**T**ransferases (KATs) and can be classified into cytoplasmic enzymes, which modify free histones in cytoplasm just after their synthesis [90], and nuclear enzymes, which are responsible for acetylation of histone and non-histone proteins in the nucleus [91]. Based on their sequence homology, nuclear HATs can be subdivided into five different families: GNAT (Gcn5 *N*-acetyltransferases), MYST, p300/CBP, transcription co-activators and steroid receptor co-

activators. The MYST (MOZ, YBF2/SAS3, SAS2, TIP60) family is the largest one and consists of TIP60 (KAT5), MOZ (KAT6A), MORF (KAT6B), HBO1 (KAT7) and MOF (KAT8) [91].

In the context of the chromatin, ING proteins recruit specific HAT and HDAC complexes to sites with the H3K4me3 mark. It is very interesting that INGs can be classified into the same three groups, (ING1/2, ING3, and ING4/5) not only according to the sequence homology and phylogenetics [72], but also according to the distinct types of association to HAT and HDAC complexes [92] (**Figure 7**). In the first group, ING1 and ING2 associate with mSin3A/HDAC1/2 histone deacetylase complexes [92]. While ING1 suppresses cell growth in a manner dependent on the mSin3A-HDAC interacting domain [93], ING2 directs HDAC1/2 complexes to actively transcribed genes as a rapid way of gene transcription suppression [79]. In the second group, ING3 forms part of hNuA4/Tip60 HAT complex, necessary for acetylation of histones H4 and H2A [92]. In the third group, ING4 and ING5 associated with HAT complexes containing MOZ/MORF (monocytic leukemia zinc finger protein/MOZ related factor) and HBO1 [46, 94]. HBO1-HAT is a complex composed of HBO1-JADE-ING4/5-hEAF6 subunits which regulate S phase progression and is likely responsible for the majority of histone H4 acetylation in higher eukaryotes [92, 95], although it also has H3 as a substrate to a lesser extent [92]. It was shown that the H3K4me3 recognition by the ING4-HBO1 complex drives acetylation on H3 at a set of genes in response to genotoxic stress [96]. In this context, the PHD finger of ING4 serves to sense the methylation state of H3K4 to facilitate HBO1-HAT activity on H3, thus, redirecting the HBO1 HAT complex activity from histone H4 to histone H3 [96]. Besides, the activity of HBO1 complex is related not only to transcription regulation, but also to apoptosis and proliferation regulation pathways due to its repression activity towards NF- κ B and p53 [97, 98]. Interestingly, ING5 is present in two independent HAT complexes with a different acetylation activity for histones. MOZ/MORF HAT acetylates histone H3 at Lys14, while HBO1 HAT acetylates histone H4 at Lys5, Lys8 and Lys12 [46]. ING5 will likely act as an adapter, recruiting these different complexes to sites in chromatin with H3K4me3 mark promoting distinct effects. Interestingly, ING5-HBO1 complex was co-purified together with MCM (minichromosome maintenance) complex, suggesting an important connection to DNA replication regulation [92].

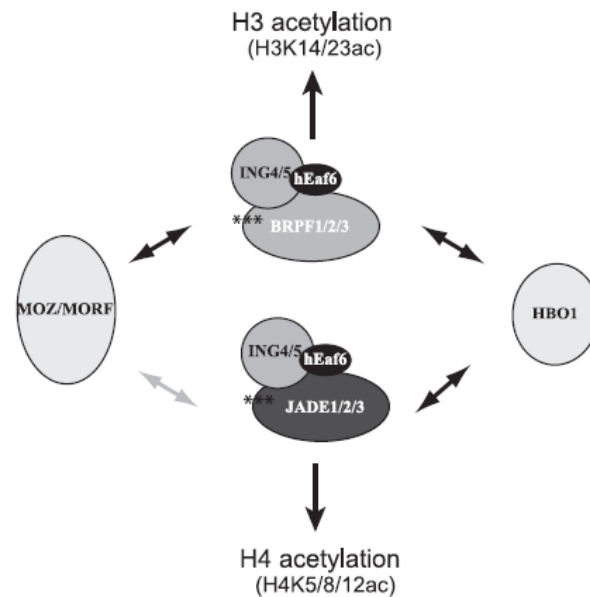


Figure 12. Model for MYST acetyltransferase assembly in alternate complexes, leading to different histone tail specificities. The HBO1 and MOZ/MORF catalytic subunits can be associated with different scaffold proteins, leading to a switch in histone tail specificity for acetylation of chromatin substrates. Thus, protein complexes associated with HAT proteins not only enable them to acetylate chromatin substrates, but also select which histone tail is targeted, specificity previously thought to reside in the acetyltransferase itself. The arrow between MOZ/ MORF and JADE1/2/3 is gray, since this interaction has only been reported in co-transfection experiments. Modified from Lalonde *et al* 2013 [94].

More recently, a new model has been proposed in which HBO1 acetyltransferase competes for BRPF or JADE scaffold subunits in cells. There is evidence in which HBO1–JADE and HBO1–BRPF HAT complexes coexist within HeLa cells [94]. HBO1–JADE complex targets mainly H4 residues, while HBO1–BRPF1 complex acetylates only H3 in the context of chromatin. These results highlight the new role of the associated scaffold subunits (JADE and BRPF) within MYST-ING acetyltransferases complexes in directing the acetylation of specific histone tails (**Figure 12**).

3.4. ING4

The ING4 gene is located at chromosome 12p 13.31 and consists of eight exons [99]. ING4 is a 29 kDa protein that was first identified by computational homology search [100].

The structure of the PHD of ING4 was determined in 2006 [101]. However, little was known about the full-length protein until it was purified and analyzed in 2010 [102]. The crystal structure of the N-terminal domain of ING4 was later reported in 2012 [103]. ING4 dimerizes in an antiparallel manner through its N-terminal domain with a coiled-coil structure and each protomer contains a helix-loop-helix structure [103]. In the full-

length version of ING4, the dimer results in a molecule arranged head-to-tail, with the two long and unfolded NLS regions pointing to opposite sides of the dimerization domain (**Fig. 13**), allowing the two PHD fingers to sample a great space around the protein and independently bind to two H3K4me3 histone marks. The binding of ING4 for the H3K4me3 peptide shows essentially the same binding site, an affinity of $K_D = 1.3 \pm 0.5 \mu\text{M}$, and discrimination between the different methylated forms of histone H3 at K4 as the isolated PHD finger [85, 102]. In the context of the chromatin, ING4 could bind simultaneously two histone tails on the same or different nucleosomes [103].

Dysregulation of ING4 or a decreased expression of ING4 is frequent in multiple cancers including head and neck squamous cell carcinomas, hepatocellular carcinoma, gastric adenocarcinoma, lung cancer, breast cancer, etc [104-107]. The exact mechanisms by which ING4 is downregulated have not been yet clearly understood. However, there are a few investigations that have revealed some of the implicated factors. The deletion of chromosome 12p 12-13 region is described for different cancers as head and neck squamous cell carcinomas and breast tumors [104]. Punctual mutations could also bring to ING4 function loss and tumorigenesis. ING4 N214D mutant attenuated inhibition of ING4-mediated cell proliferation, anchorage-independent growth and migration and induction of cell death [108]. Mislocalization or an altered ratio of nuclear and cytoplasmic ING4 expression can also lead to tumorigenesis in head and neck squamous cell carcinoma [109]. Post-transcriptional regulation may also be related with ING4 down-expression [110].

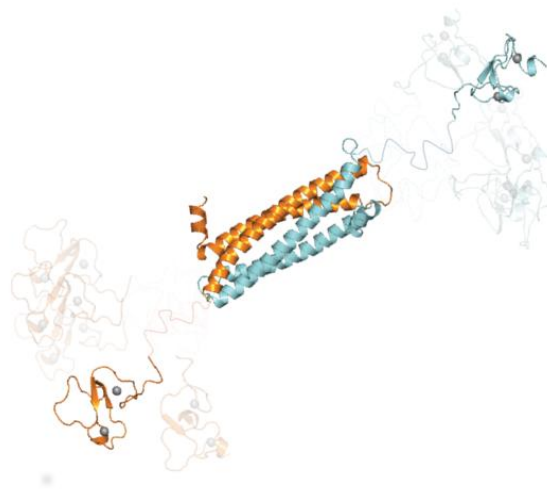


Figure 13. Model of ING4 dimer structure. The flexible nuclear localization sequence regions of ING4 were modeled as curvy lines. Figure modified from Culurgioni *et al* 2012 [103].

Although ING4 shares the capacity of chromatin modification and cell proliferation inhibition with other ING proteins, it is worth highlighting its unique role in the

regulation of angiogenesis and cell migration [99]. The correlation between ING4 and angiogenesis has been proved in different cancer types and depend on the interaction between ING4 and the nuclear factor NF- κ B. ING4 could decrease NF- κ B binding activity, transcriptionally repressing NF- κ B-responsive genes, including angiogenesis-related genes [111]. In addition to NF- κ B, the interaction of ING4 with Hypoxia Inducible Factor (HIF) has also been shown to play a key role in ING4-mediated suppression of myeloma-induced angiogenesis. The mechanism of this mode of action depends on the ability of ING4 to recruit chromatin-remodeling factors [112]. ING4 has also been related to the suppression of cell migration and contact inhibition through suppressing the activity of invasion-related genes such as MMPs (Matrix metalloproteinases), which are known to be regulated by NF- κ B [113].

3.5. ING5

ING5 was first identified by computational homology search and shares high homology sequence with ING4. The ING5 gene is located at chromosome 2p37.3 and consists of eight exons as ING4 does [114]. ING5, like other INGs, is a H3K4me3 sensor and it associated with HATs complexes. Depending on the biological context, ING5 may have suppressive or oncogenic abilities [74].

ING5 is involved in the p53-dependent stress signaling, partially by interacting with p53 and promoting its acetylation by p300 which is a member of a HAT complex, inducing activation of p53 downstream effectors and leading to chromatin remodeling [100]. ING5 also associates with minichromosome maintenance (MCM) proteins, with which it plays an essential role in DNA replication [92]. What is not yet clearly understood is the function of the recently discovered ING5 phosphorylation at threonine 152. This modification is carried out by CDK2 (cyclin dependent kinase-2) and it is known that, although the phosphorylation takes place in the NLS region of ING5 it does not affect its subcellular localization [115]. Very recently, ING5 has been related to MiRNAs, non coding RNAs being shown to be sometimes involved in proliferation and apoptosis. ING5 expression is negatively regulated by MiR-1307, promoting ovarian cancer tumorigenesis and chemoresistance [116].

ING5 deletion, mutation and/or down-regulation are frequent in different cancer types such as oral carcinogenesis or lung cancer [117, 118]. The reduction in nuclear ING5 expression and its cytoplasmic translocation has been observed in head and neck squamous carcinoma, and linked to tumorigenesis and aggressiveness in colorectal and gastric cancers [119, 120]. Three point mutations have been described in the N-terminal domain of ING5 in oral squamous carcinoma, C75R, I68V and Q33R [117]. Mutations in

this domain might influence the 3-dimensional structure of ING5 or may alter the interaction interface between ING5 and its partners on the formation of HAT complexes, thus, altering its role in downstream genes through chromatin remodeling.

The structure of PHD finger of ING5 bound to H3K4me3 peptide was determined [86], being similar to other PHDs of ING proteins [78, 84, 85, 96, 101]. In the crystal structure of the ING5 PHD in complex with H3K4me3 (**Fig. 14**), the overall architecture shows the characteristic features of a PHD finger. It contains two zinc-binding clusters, the first coordinated by one histidine and three cysteines, while the second is coordinated by four cysteines. In the complex, the H3K4me3 peptide binds to the PHD of ING5 via formation of a third antiparallel β -strand that pairs with the central β -sheet of the protein. The affinity of the binding is $2.4 \pm 1.0 \mu\text{M}$ as measured by fluorescence spectroscopy [86] and is favoured for trimethylated over di-, mono- and non-methylated histone peptides.

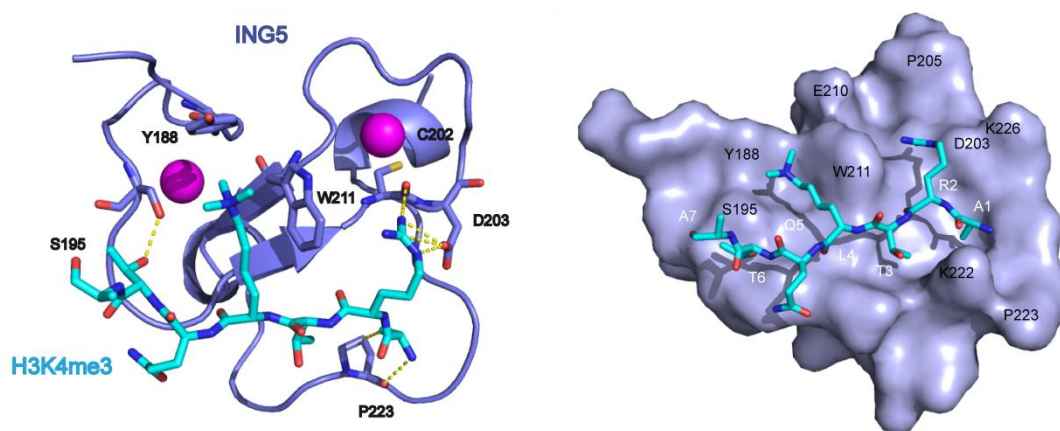


Figure 14. The ING5 PHD finger bound to H3K4me3 (PDB: 3C6W). Left: Recognition of the histone peptide by the ING5 PHD finger. Human ING5 PHD represented in *purple* bound to the H3K4me3 peptide in *cyan*, and the zinc ions are depicted in *magenta*. Right: The H3K4me3 binding groove. The histone peptide is depicted as a stick model.

There are not structural studies on the full-length version of ING5. However, the fact that ING5 shares a high grade of sequence homology with ING4 suggests that the structure is also conserved. Then, ING5 might be also folded into three domains: an N-terminal dimerization domain, a central and unstructured flexible NLS region, and a PHD finger in the C-terminal region.

Previous to this thesis work, the structure of the N-terminal domain of ING5 was determined in our group [121]. The overall crystallographic structure showed that the N-terminal domain of ING5 is a dimer in which each protomer is folded into 3 α -helices forming a coiled-coil structure similar to ING4. However, the crystal showed a completely different dimerization interface, being an asymmetric dimer while ING4 forms a symmetric dimer (**Figure 15B and C**). This result was unexpected if we take into

account previous results on ING4 N-terminal crystal structure and the high level of homology sequence between ING4 and ING5 N-terminal domains (75 % sequence identity; **Figure 15A**). In addition, usually homodimers are mainly formed in a symmetric manner [122]. On the other hand, some complexes have been reported to alternate between symmetric and asymmetric conformations to accomplish specialized functions [123], although the biologically relevant asymmetric complexes are only the 3% of the structure databases [124].

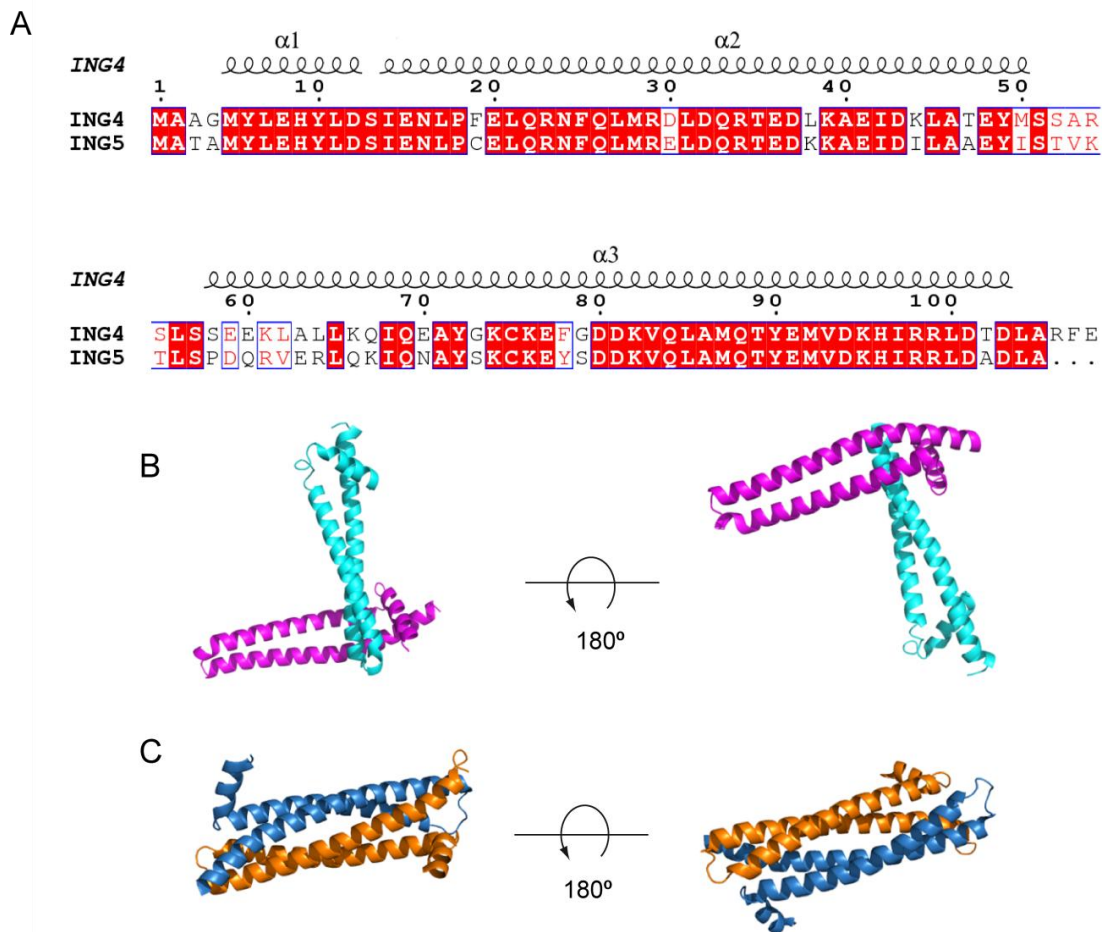


Figure 15. Coiled-coil assembly of ING4 and ING5 dimerization domain. **A.** ESPrict 3.0 [125] alignment of the ING N-terminal sequences of ING4 and ING5 with the experimentally determined α -helices in ING4 N-terminal crystal structure. **B.** Crystallographic structure of the N-terminal domain of ING5. Protomers arrangement forms an asymmetric dimer. Figure based on Rodriguez, J.A. 2014 [121]. **C.** Crystallographic structure of the N-terminal domain of ING4 (PDB: 4AFL) [103]. Protomers arrangement forms a symmetric dimer.

4. Protein oligomerization in function regulation

Oligomerization provides with several structural and functional advantages to proteins and leads to a better cellular economy; the advantages of modular complex formation are maintained without increasing genome size. Oligomerization not only contributes with higher stability, but also increases cooperativity and allows for multivalent binding to ligands.

For instance, the initial level of DNA compaction in eukaryotic cells occurs through its association with dimers of the four core histones, to form nucleosomes. The symmetry achieved by this dimerization pattern of the histone octamer generates repeated structural motifs on the surface that allows the DNA to bind and wrap around in a symmetrical manner. This organization might explain the preference of core histones for double stranded DNA [126, 127]. Many DNA binding proteins involved in DNA repair, DNA replication and gene expression form oligomers. An example is the case of type II restriction enzymes, which bind to palindromic DNA sequences. Oligomerization increases the DNA binding affinity not only through cooperativity, but also by doubling the length of the DNA site bound to the protein [128].

In other cases, self association of proteins related to DNA is accomplished by both structural and functional means. Protein oligomerization is particularly important for the assembly of protein complexes involved in gene expression as transcription factors. In eukaryotic cells many different proteins self-associate to form heteromeric transcriptional complexes. The function of these kind of complexes depends on its composition, and having multiple binding partners transforms the complex providing distinct properties to perform specific functions, thereby mediating different gene regulation [129]. Well known examples are: the Jun-Fos heterodimer, which increases its transcriptional activity upon binding to cofactor NFAT; and the Myc-Max and Mad-Max heterodimerization, which defines whether a large number of targeted genes will be expressed or silence. Members of these dimerizing families of proteins present conserved domains for dimerization, among them: the helix-loop-helix (HLH), a four helix bundle as a dimer, the leuzine-zipper-like (LZL), a coiled coil domain, and the ligand-binding domain (LBD) [130].

In the simplest form, oligomerization functions as a general mechanism for sensing the protein concentration. This may be a mechanism for enzyme activation, by altering the association/dissociation equilibrium. One of the best known example is caspase-9, which exists as an inactive monomer under physiological conditions but dimerizes during apoptosis [131]. On the contrary, protein dimerization can inhibit an active monomer. This is the case of the receptor-like protein tyrosine phosphatase- α , which exists as a downregulated homodimer in the cell surface, but it dissociates and activates upon ligand binding [132]. Oligomerization and activation of receptors in the cell surface in response to a ligand binding is a common feature in pathways of signal transduction across the cell membrane, as in the case of growth hormone, interferon, cytokine and tyrosine kinase receptor families [133].

The crystal structure of the ING4 N-terminal domain showed an antiparallel symmetric homodimer with each protomer folded into a helix-loop-helix structure [103, 134]. Dimerization of ING4 has a great impact on its tumour suppressor activity, monomeric mutants being unable to enhance apoptosis in response to DNA damage. This arrangement suggested that ING4 could bind simultaneously not only one, but two H3K4me3 on the same or different nucleosomes (**Figure 16**). In a bivalent model, it was hypothesized that ING4 might bind two different nucleosomes close in space, favouring at the same time the recognition of DNA sequences that are far away, activating distant enhancing elements.

The high sequence homology between ING4 and ING5 makes it possible to model the ING5 N-terminal structure based on ING4 N-terminal domain crystallographic model and perform analysis on the coiled coils with the help of helical-wheel models (**Figure 17**). Interestingly, the regions of highest homology keep the same or similar amino acid residue in positions *a*, *d* and *g*, essential for the formation of the heptad pattern typical of coiled-coil structures in leucine-zipper-like domain. Furthermore, the polar residues that stabilize the dimer formation are also highly conserved which suggests the possibility of ING5 to form symmetric homodimers in solution as ING4 does.

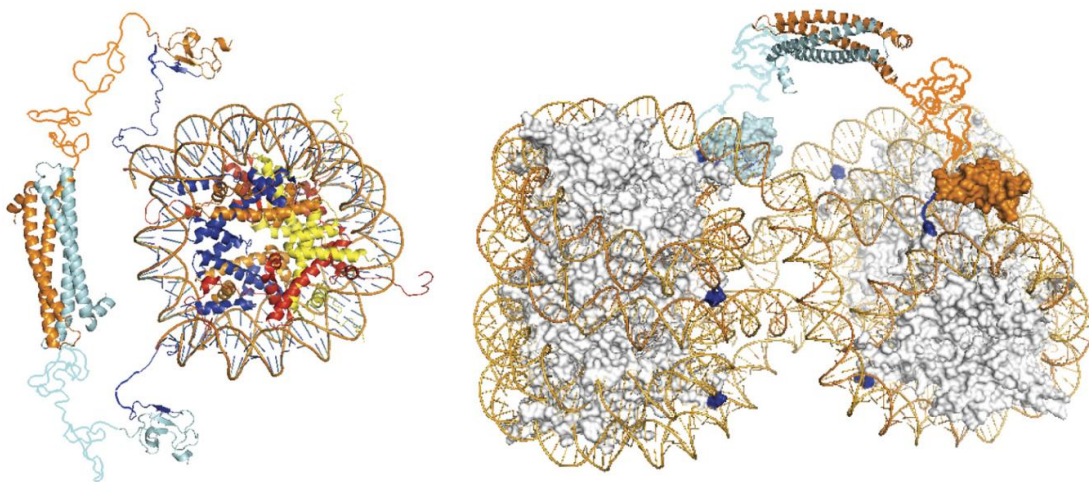


Figure 16. Model of ING4 dimer binding to chromatin. **A.** The ING4 dimer could bind the two H3 tails of the same nucleosome through its PHDs. **B.** ING4 could bind two nucleosomes due to its elongated structure. The depicted nucleosomes are positioned consecutively in the chromatin fiber, but they could also be non-consecutive if the distance is short enough. Figure from Culurgioni *et al* 2012 [103].

A quantitative estimation of the potential for dimerization was calculated by computation of the free energy of the interaction of the two protomers for ING5 model and the prediction was that ING5 may form a dimer as stable as ING4 (**Figure 17**) [103].

ING4 and ING5 have been shown to be part of the same chromatin remodeling complex, the HBO1 HAT [92]. Taking into account the high homology in sequence with ING4, we could hypothesize of a possible heterodimer formation between ING4 and

ING5. Molecular modelling suggests that ING4/ING5 heterodimers may be as stable as the corresponding homodimers (**Figure 17**) [103]. In this case, the heterodimerization could have several consequences and different regulatory effect, as it happens for other transcription factors as p63 and p73 (from the p53 family of tumor suppressors) that can form heterotetramers in a yet not fully understood alternative function [135]. Interestingly, ING4 and ING5 have been shown to be part of the same chromatin remodeling complex, the HBO1 HAT [92] (**Figure 7**).

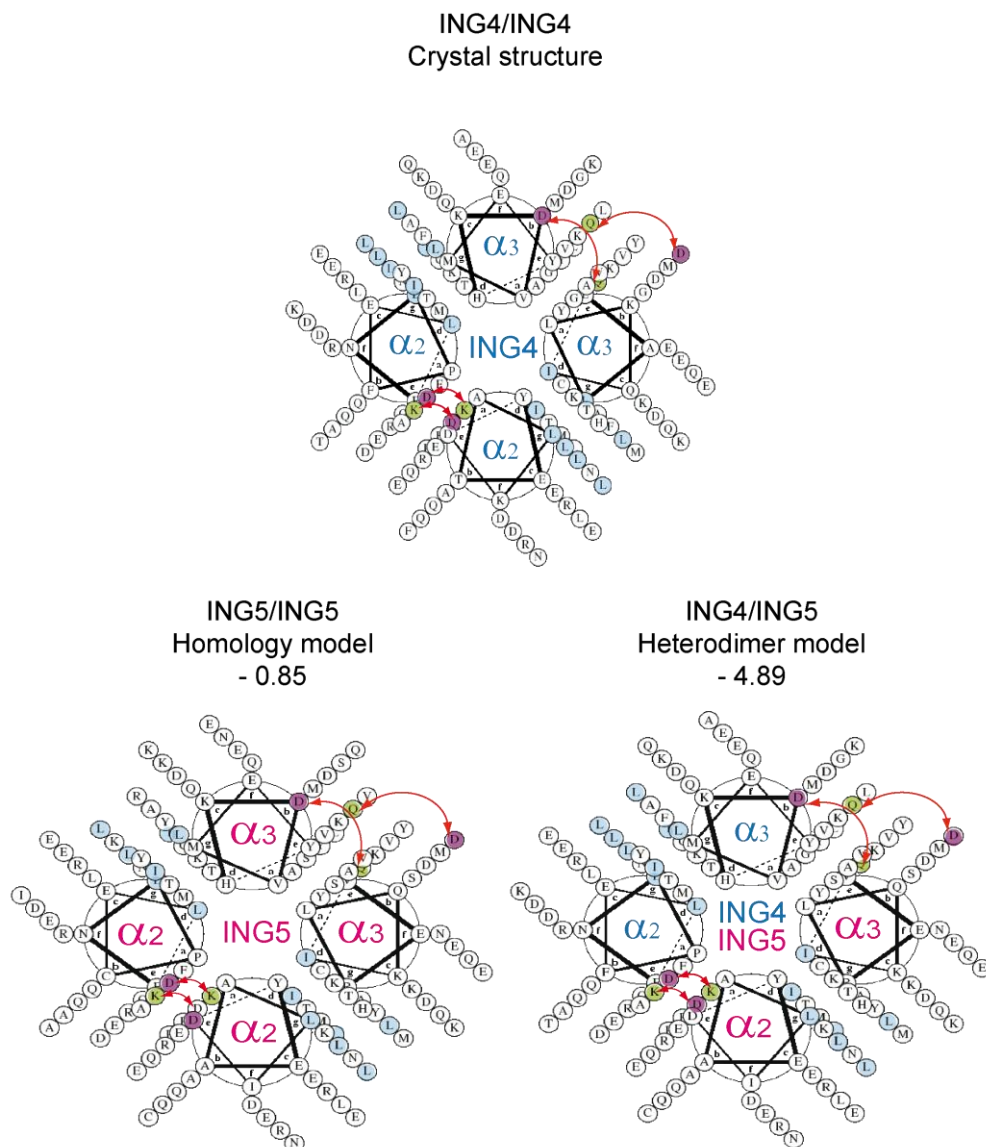
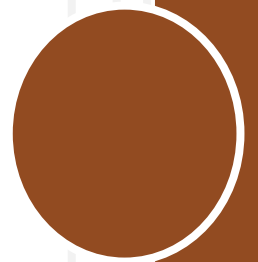



Figure 17. Coiled-coil assembly of ING4 dimerization domain and models of other putative ING dimers. Top: Helical wheel representation of the distribution of residues in the coiled coil and the heptad pattern $(abcdefg)_n$ of the N-terminal domain of ING4 crystal structure. *Light blue* indicate the leucine and isoleucine residues forming the zipper; *purple* and *green* circles indicate the negatively and positively charged (or hydrogen bond donor) residues, respectively, that interact, stabilizing the dimer. The salt bridge Asp-32/Lys-39 and the hydrogen-bonded pair Asp-94/Gln-69 are indicated by *red arrows*. **Bottom:** Models and relative predicted stability of ING5 dimer and ING5/ING4 heterodimer. Helical wheel representation of the distribution of residues in an idealized coiled coil homodimer structure of the N-terminal domain of ING5 (left) and the putative heterodimer model ING4/ING5 (right) were done based on the high homology with ING4. Coiled coil representation obtained from Drawcoil 1.0 program server (<http://www.grigoryanlab.org/drawcoil/>). The difference in free energy for the interaction between both protomers relative to ING4 dimer (in kcal/mol) was calculated with FoldX programme [136, 137]. According to this calculation an ING4/ING5 heterodimer would be even more stable than the two corresponding homodimers.

OBJECTIVES





The aim of this thesis work is to study the structural organization of the tumor suppressor ING5 and its function with the following particular objectives:

- 1- Study the structural organization of ING5.
- 2- Characterization of the full length ING5 binding to the H3K4me3 mark.
- 3- Study the solution structure of the N-terminal domain of ING5.
- 4- Characterize the DNA recognition by the NLS regions of ING4 and ING5
- 5- Evaluate the structural and functional impact of three point mutations at the N-terminal domain of ING5 detected in primary tumors.

MATERIALS AND METHODS



1. Genes, plasmids, cloning and mutagenesis

1.1. ING5, Nt, Nt mutants and PHD

A sequence analysis of the ING5 human DNA showed around one third of codons that correspond to rare ones in bacteria. Therefore, a synthetic gene of the full-length ING5 (Uniprot Q8WYH8, isoform 1) with codons optimized for expression in *E. coli* (Entelechon GmbH) was cloned into the expression vector pET11d (within NcoI and BamHI sites) and used to subclone different ING5 constructs. The full-length ING5 (residues 1-240) construct was designed with a His-tag at the N-terminus to facilitate protein purification by affinity chromatography followed by a specific sequence for TEV protease cleavage (ENLYFQG) for the His-tag removal after purification. Between the TEV site and the initial methionine of ING5 an alanine was inserted to prevent possible cleavage difficulties. The TEV protease cleavage is done between the glutamine and the glycine so the purified product after cleavage has two extra residues before the initial methionine of ING5.

The construct for the N-terminal domain of ING5 (residues 1-105, ING5 Nt₁₋₁₀₅) was obtained by inserting a stop codon in the amino acid sequence site corresponding to R106 in the ING5 cloned in pET11d. This clone was used as a template for the production of different ING5 Nt₁₋₁₀₅ mutants (Q33R, I68V, C75R, C19S, C75S, and $\Delta\alpha 1$) by using Quick Change Site-Directed mutagenesis kit (Agilent Technologies). Primers for mutagenesis (Thermo Scientific) were designed with Quick Change primer design (Agilent Technologies). We later designed the construct for the ING5 Nt₁₋₁₀₅ for expression without affinity His-tag. To obtain this construct, primers were designed to amplify and remove His-tag and TEV-site at the same time, using ING5 Nt₁₋₁₀₅ cloned in pET11d as template. We use ING5 Nt₂₋₁₀₅ to name this construct since it loses the first methionine when expressed in *E. coli*.

The C-terminal construct of ING5 (ING5 PHD, residues 178-240) was cloned into a modified pET29a expression vector. For that, the sequence between residues 178 and 240 of ING5 (of the full length ING5 construct on pET11d) was amplified with designed primers (Thermo Scientific) by PCR introducing two different sequences for restriction enzymes in each end of the gene (NcoI and BamHI sites). The amplified sequence corresponding to PHD was cloned into a plasmid based on pET39_Ub19 construct described in a recent publication [138] as universal tag for efficient protein production in bacteria. It consists of ubiquitin at the N-terminus with a His-tag sequence inserted in its first loop. The PHD sequence was preceded by a strep-tag sequence and TEV protease cleavage site was also included between the strep-tag sequence and the ING5 PHD

sequence. The purpose was to purify the protein by two sequential affinity chromatography steps.

For experiments on eukaryotic cells, the human ING5 clone was obtained from Addgene. The gene was amplified by PCR with oligonucleotides containing the N-terminal sequence for AU5 or HA tags (TDFYLK and YPYDVPDYA epitopes respectively) and cloned into the pLPC retroviral vector (using BamHI and EcoRI sites) which has puromycin resistance. The described cancer mutant constructs of HA-tagged ING5 (C75R, I68V, Q33R) were generated with the QuikChange site-directed mutagenesis kit (Agilent, United States).

1.2. ING4, Nt, PHD and ING4 Δ NLS

The synthetic gene of the full length untagged human ING4 (residues 1-249; Uniprot Q9UNL4, isoform 1) cloned into vector pET11d has been previously described [102]. The N-terminal construct (ING4 Nt) used in this study consists of ING4 residues 1-108 with a Strep-tag (WSHPQFE), inserted after the initial methionine by PCR. A clone of full length ING4 with the same Strep-tag inserted after the initial methionine was produced in the same way. The construct for the C-terminal PHD finger of ING4 (ING4 PHD, residues 188-249 with an extra methionine at the N-terminus), was previously described [101]. The synthetic gene of the ING4 central region deletion mutant (named ING4 Δ NLS and lacking residues 106-187) was purchased from Entelechon GmbH, and it was modified with the insertion of a His-tag and a TEV protease site at the N-terminus. All these ING4 constructs have been codon-optimized for expression in *E. coli*. The HA-ING4 and AU5-ING4 clones in pLPC vector have already been described [102].

1.3. JADE 1L Iib domain

For the binding analysis to ING5 Nt₁₋₁₀₅, the synthetic gene of JADE1L corresponding to the conserved Iib domain (Uniprot: Q6IE81, residues 489-534), was cloned as an N-terminal fusion with ubiquitin using vector pETM60_Ub3, a gift from Vladimir V. Rogov [138]. This JADE1L domain was reported to be the binding partner of ING4 and ING5 within the HBO1 HAT complex by Avvakumov *et al* [139]. Our JADE1L Iib construct (UBJADE) contains an ubiquitin as a tag for increasing solubility when expressed in *E. coli*, a His-tag and a TEV site for tag removal during purification (pETM60_Ub3_His_TEV_UBJADE1L_Iib).

The identity of all clones was confirmed by DNA sequencing and the amino acid sequences are described in **table 1**.

Table 1. List of protein sequences used in this thesis project. Sequences in bold correspond to affinity tags and TEV sites (in red).

Name	Sequence
ING5	MGSS HHHHHH SSG ENLYFQ GAMATAMYLEHYLDSIENLPCELQRNFQLMRELDQRTEDK KAEIDILAAEYISTVKTLSPDQRVERLQKIQNAYSCKKEYSDDKVQLAMQTYEMVDKHI RRLDADLARFEADLKDKMEGSDFESSGGRGLKKGRGQKEKRGSRGRGRRTSEEDTPKKK KHKGGSEFTDTILSVHPSDVLMPVDPNEPTYCLCHQVSYGEMIGCDNPDCPIEWFHFA CVDLTTKPKGKWFPCRCVQEKRRKK
ING5 Nt ₁₋₁₀₅	MGSS HHHHHH SSG ENLYFQ GAMATAMYLEHYLDSIENLPCELQRNFQLMRELDQRTEDK KAEIDILAAEYISTVKTLSPDQRVERLQKIQNAYSCKKEYSDDKVQLAMQTYEMVDKHI RRLDADL
ING5 Nt ₂₋₁₀₅	MATAMYLEHYLDSIENLPCELQRNFQLMRELDQRTEDK KAEIDILAAEYISTVKTLSPD QRVERLQKIQNAYSCKKEYSDDKVQLAMQTYEMVDKHIRRLDADL
ING5 PHD	MQIFVKTLTGKTTITLEVEPGSA HHHHHHHHHH AGSSDTIENVKAKIQDKEGIPPDQQR LIFAGKQLEDGRTLSDYNIQEKSTLELVLELQNESG WSHPQFEKS ENLYFQ GAMDMPVD PNEPTYCLCHQVSYGEMIGCDNPDCPIEWFHFACVDLTTKPKGKWFPCRCVQEKRRKK
ING4	MWSHPQFEK AAGMYLEHYLDSIENLPFELQRNFQLMRDLQRTEDLKAEIDKLATEYMS SARLSSEEKLALLKQIQEAYGKCKEFGDDKVQLAMQTYEMVDKHIRRLDLDLARFEAD LKEKQIESSDYDSSSSKGGKGRQKEKKAARARSKGKNSDEEAPKTAQKKLKVRTSP EYGMPSVTFGSVHPSDVLMPVDPNEPTYCLCH QVSYGEMIGCDNPDCSIEWFHFACVGLTTKPRGKWFPCRCVQERKKK
ING4 Nt	MWSHPQFEK AAGMYLEHYLDSIENLPFELQRNFQLMRDLQRTEDLKAEIDKLATEYMS SARLSSEEKLALLKQIQEAYGKCKEFGDDKVQLAMQTYEMVDKHIRRLDLDLARFE
ING4 PHD	MDMPVDPNEPTYCLCHQVSYGEMIGCDNPDCSIEWFHFACVGLTTKPRGKWFPCRCVQ ERKKK
ING4ΔNLS	MSYY HHHHHH DYDIPTT ENLYFQ GAMAAGMYLEHYLDSIENLPFELQRNFQLMRDLQ RTEDLKAEIDKLATEYMS SARLSSEEKLALLKQIQEAYGKCKEFGDDKVQLAMQTYEMV DKHIRRLDLDLADMPVDPNEPTYCLCHQVSYGEMIGCDNPDCSIEWFHFACVGLTTKPR GKWFPCRCVQERKKK
UBJADE	MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDY NIQKESTLHLVLQLESASGSG HHHHHH SAG ENLYFQ GAMGLQLFTRLQDLERVRNLT YMHVTRREKIKRSVCKVQEQIFNLYTKLLE

2. Protein expression and purification

All ING5 constructs were produced in *E. coli* BL21 (DE3) cells and purified by different chromatographic steps detailed below. All pure proteins were concentrated by ultrafiltration, flash-frozen in liquid N₂ and stored at -80 °C until used. The identity and labeling of each protein was confirmed by MALDI-TOF and the purity was checked by SDS-PAGE. Protein concentration was measured by UV absorbance using extinction coefficients calculated using the ExPASy ProtParam tool [140].

2.1. Expression and purification of ING5, Nt, Nt mutants and PHD

For ING5, Nt, Nt C75R, Nt I68V, Nt Q33R, Nt C19S, Nt₂₋₁₀₅, and Nt Δα1 expression, cells were grown in LB or in ZYP-5052 auto-induction rich medium [141]. Cells that were grown in LB at 37 °C were induced at O.D.₆₀₀ = 0.8 with 0.5 mM IPTG for 3 h at 37 °C, while cells that were grown in auto-induction ZYP-5052 medium for 2 h at 37 °C, were left for expression induction at 20 °C for 22-23 h. [U-¹⁵N]-labeled ING5, Nt (1-105), Nt (2-105), Nt C75S, Nt C19S and PHD proteins were produced for NMR experiments in a modified auto-induction media, P-5052 [142]. There were proteins as Nt, Nt C19S and PHD that were [U-¹⁵N]-labeled only or [U-¹³C, U-¹⁵N]-labeled by using Marley's method [143], in which cells were grown in LB until they reached O.D.₆₀₀ = 0.8, when they were pelleted and washed with a M9 salt solution and grown in ¹⁵N labeled minimal media for 30 min at 37 °C before inducing with 0.5mM IPTG for 3 hours.

All cultures were harvested by centrifugation and resuspended in lysis buffer (20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT (or 20 mM Tris pH 8.0, 1 mM DTT in the case of Nt₂₋₁₀₅), always in the presence of protease inhibitors (1 tablet *Complete EDTA-free* per 50 mL). After sonication and ultracentrifugation, proteins were found in the insoluble (ING5, Nt₁₋₁₀₅, Nt C75R, Nt Q33R, Nt I68V, Nt C19S, Nt C75S, Nt Δα1) or in the soluble fraction (Nt₂₋₁₀₅, PHD, Δα1, C75S). Insoluble proteins were recovered from the inclusion bodies by solubilization in 8 M urea containing lysis buffer and ultracentrifugation for 3 h. Supernatant was refolded by a 1:50-1:100 dilution into cold 20 mM, Tris pH 8.0, 300 mM NaCl, 1 mM DTT, and 50 μM ZnCl₂ (for the constructs including the PHD finger), containing protease inhibitors.

Refolded ING5 was purified by affinity chromatography with a HisTrap column (GE Healthcare) and eluted with linear imidazol gradient (0-500 mM) in 15CV. Selected fractions, without being concentrated, were loaded into a Superdex 200 26/60 column in different runs for polishing and a resulting yield of 0.7-1 mg/ L_{culture} (no TEV cleavage was done since this caused protein precipitation; **Figure 1**).

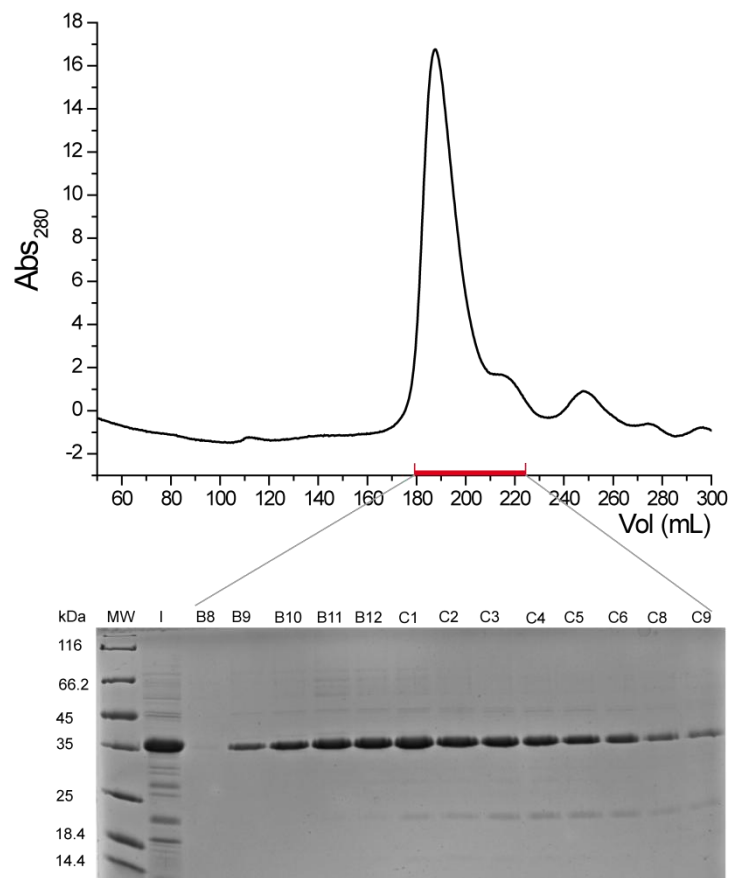


Figure 1. Last step of the purification of ING5. (Top panel) Selected fractions from the affinity chromatography step were pooled (lane I in the gel) and loaded into a gel filtration column (Superdex 200 26/60 column). (Bottom panel) A Tris-glycine 12 % SDS-PAGE was run with the indicated fractions from the chromatogram, at room temperature and 220 volts for 45 min. The fractions corresponding to the main chromatographic peak were pooled and concentrated for structural analysis or fast frozen and stored at -80°C .

Soluble or refolded Nt_{1-105} and all Nt_{1-105} mutants (C19S, C75S, $\Delta\alpha 1, \text{C75R}$, I68V, Q33R) containing a His-tag were purified by affinity chromatography with a HisTrap 5 mL column (GE Healthcare) and eluted with an imidazol linear gradient (0 - 0.5 M) in 5 CV. Selected fractions were incubated with TEV protease in a 1:30 ratio (1 mg TEV: 30 mg protein) and dialyzed against 20 mM, Tris pH 8.0, 40 mM NaCl, 1 mM DTT at 4°C for 16 h. Dialyzed samples were loaded again on a HisTrap column (GE Healthcare) in order to remove the cleaved His-tag. The flow through was loaded on a Hitrap Q HP 5 mL column (GE Healthcare) and the elution was done with NaCl linear gradient (0.04 - 1 M) in 5 CV. Selected fractions were concentrated and separated by gel filtration (Superdex 75 16/60, GE Healthcare) in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT with a yield of 6-18 mg/L_{culture}. The degree of purity of Nt_{1-105} sample is shown in **Figure 2**. ING5 Nt_{2-105} was purified from the soluble fraction with a first step on a Q-Sepharose 26/10 column equilibrated in 20 mM Tris pH 8.0, 1 mM DTT and a gradient elution from 0 M to 1 M NaCl in 9.5 CV. Selected fractions were pooled and dialyzed o/n at 4°C against 20 mM Tris pH 8.0, 40 mM NaCl, 1 mM DTT. Dialyzed sample was loaded onto

a Mono Q 5/50 HR equilibrated in 20 mM Tris pH 8.0, 40 mM NaCl, 1 mM DTT. Elution was done with a linear gradient to 1 M NaCl in 25 CV. Selected fractions were polished by successive gel filtration chromatography on Superdex 75 26/60 and Superdex 75 16/60 columns equilibrated in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT. The final yield was 21 mg/L_{culture}. The degree of purity of Nt₂₋₁₀₅ sample is shown in **Figure 2**. Soluble PHD was purified first by affinity chromatography with a HisTrap column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT and eluted with one step of imidazol of 500 mM. Selected fractions were diluted 2 times with 20 mM Tris pH 8.0 and loaded into a StrepTactin column equilibrated in 100 mM Tris pH 8.0, 150 mM NaCl. Elution was done with one step of 2.5 mM Desthiobiotin. Eluted fractions were selected and digested with TEV in a 1:30 ratio (at the same time as a dialysis was done o/n at 4 °C to remove Desthiobiotin. After injection in the StrepTactin column the flow through was concentrated for gel filtration separation on a Superdex 75 16/60 column equilibrated in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT. The yield of pure proteins was 2 mg/L_{culture}. The degree of purity of this sample is shown in **Figure 2**.

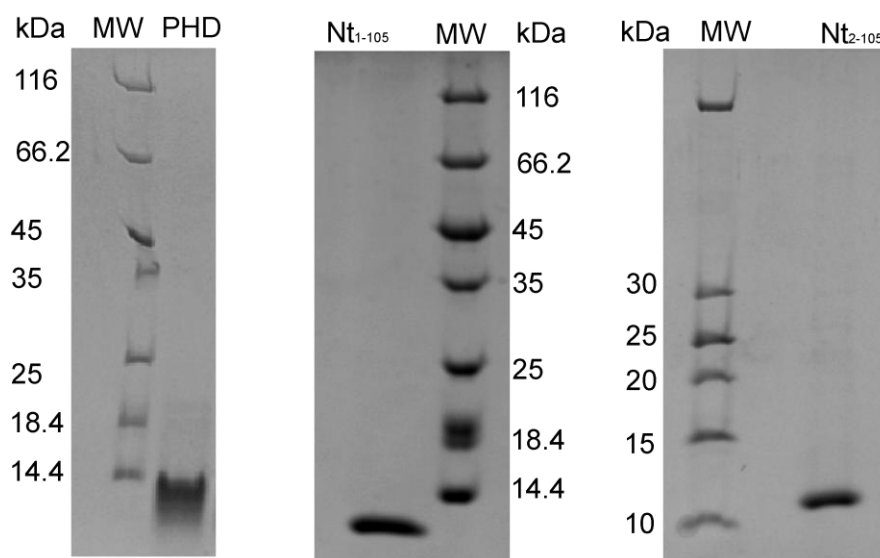


Figure 2. Representative purity grade of protein samples. SDS-PAGE on a Tris-tricine 10 % gel was loaded with PHD (2 µg) protein and run at room temperature at 120 volts for 80min. Tris-glycine 12 % gels were used in the case of Nt₁₋₁₀₅ (3 µg) and Nt₂₋₁₀₅ (2µg) and run at room temperature at 220 volts for 45 min . All gels were stained with coomassie brilliant blue and destained with 30% acetic acid in ethanol.

2.2. Expression and purification of ING4, Nt, PHD and ING4ΔNLS

For Nt, PHD and ING4ΔNLS expression, cells were grown in auto-induction medium [141]. Uniformly ¹⁵N enriched ING4 was expressed in a modified auto-induction medium [142] and in minimal media as described [143]. Cultures were harvested by centrifugation and resuspended in lysis buffer (20 mM Tris pH 8.0, 1 mM DTT), with the addition of 150 and 300 mM NaCl in the case of N-t and ING4ΔNLS respectively,

always in the presence of protease inhibitors (1 tablet Complete EDTA-free per 50 mL). After sonication and ultracentrifugation, proteins were predominantly found in the insoluble fraction (ING4, Nt, PHD, and ING4 Δ NLS) or in both the insoluble and soluble fractions (Strep-ING4, Nt). Insoluble proteins were solubilized in lysis buffer with 8 M urea and separated by ultracentrifugation at 35000 rpm for 3h. Supernatant was refolded by a 1:10-1:100 dilution into cold 20 mM, Tris pH 8.0, 1 mM DTT, and 50 μ M ZnCl₂ (needed for PHD finger folding).

Refolded ING4 was loaded onto a HiLoad 26/10 Q Sepharose anion exchange column equilibrated in 20 mM Tris pH 8.0, 1 mM DTT and elution was carried out with a 0-0.5 M NaCl gradient in 4.7 column volumes. Eluted fractions were diluted 3 times in 20 mM Tris pH 8.0, 1 mM DTT and loaded onto a Hi-Trap SP FF column equilibrated in 20 mM Tris pH 8.0, 50 mM NaCl and 1 mM DTT for cation exchange chromatography. Elution was done with a 0.05-1 M NaCl gradient in 20 CV. Selected fractions were concentrated and separated by gel filtration in a Superdex 75 26/60 column equilibrated in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT. The degree of purity of this sample is higher than 90 % (**Figure 3**).

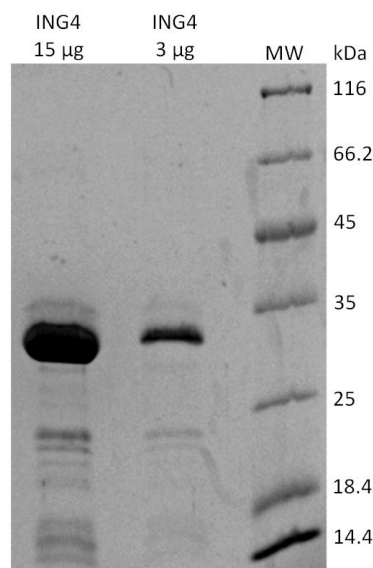


Figure 3. SDS-PAGE analysis of the purity of ING4 protein. A 12 % gel was loaded with 15 (left lane) or 3 (right lane) μ g of ING4 protein and run at room temperature and 220 volts for 45 min. The gel was stained with coomassie brilliant blue and destained with 30% acetic acid in ethanol. Densitometry analysis of the gel indicates that the major, corresponding to ING4, represents between 90 (left lane) and 94 % (middle lane) of the total protein.

Soluble Strep-ING4 was purified on a Strep-tactin 5 mL column (IBA-lifesciences) equilibrated in 100 mM Tris-HCl, pH 8.0, 150 or 300 mM NaCl, and eluted with 2.5 mM desthiobiotin. Selected fractions were concentrated and loaded onto Superdex 75 16/60 equilibrated with 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1mM DTT. Soluble Strep-

ING4 Nt was purified on a Strep-tactin 5 mL column in the same way as the full length protein. Refolded PHD was purified as previously described [101]. Refolded His-ING4 Δ NLS was loaded into a His-Trap FF crude 5 mL column equilibrated in 20 mM Tris, pH 8, 300 mM NaCl, 1 mM DTT and washed with 50 mM imidazol. Elution was done with a 50-300 mM gradient in 20 CV. Selected fractions were diluted 1:2 in 20 mM Tris pH 8.0, 1mM DTT and loaded onto a HiTrap Q HP 5 mL column equilibrated with 20 mM Tris, pH 8, 30 mM NaCl, 1 mM DTT. Protein was eluted with a 0.05-1 M NaCl gradient in 50 CV. Selected fractions were concentrated and loaded on a Superdex 200 26/60 column equilibrated in 20 mM Tris, pH 8, 300 mM NaCl, 1 mM DTT.

2.3. Expression and purification of JADE1L_Iib domain

The expression of UBJADE construct, cloned in into pETM60_Ub3, was done at low temperatures using *E. coli* Artic (DE3) RIL grown in LB medium. With this approach we tried to increase the expression yield of soluble recombinant UBJADE since this *E. coli* strain expresses cold-adapted chaperonins which confer an improved protein processing at low temperatures. However, after the induction of expression with 0.5 mM IPTG, the majority of the protein was insoluble, and the little amount of soluble expressed UBJADE resulted in very low purification yields due to its low solubility in the purification conditions as well as protein degradation. Therefore, this strategy for obtaining JADE1L_Iib was abandoned.

3. Circular dichroism (CD)

Circular dichroism is defined as the difference in absorption of left-handed and right-handed circularly polarized light. When optically active molecules such as proteins interact with the light, they absorb right and left hand circularly polarized light differently. CD spectra are reported in units of mean residue molar ellipticity ($[\theta]$, deg.cm².dmol⁻¹) [144]. This technique is an excellent method for evaluating secondary structure in solution because when the chromophores of the polypeptide backbone are aligned forming regular secondary structure motifs, they have characteristic far UV (178-250nm) CD spectra. For example, α -helical proteins have negative bands at 222 nm and 208 nm and a positive band at 193 nm. Proteins with antiparallel β -sheets have negative bands at 218 nm and positive bands at 195 nm, and disordered proteins have very low ellipticity above 210 nm and negative bands near 195 nm (**Figure 4**) [144].

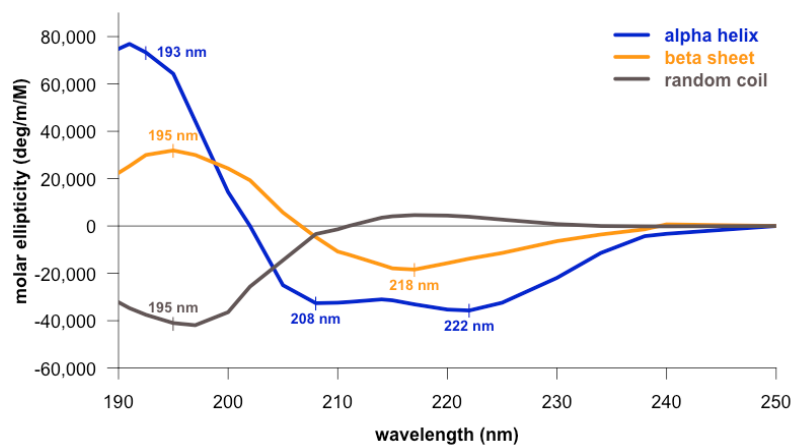


Figure 4. Circular dichroism reference spectra for each type of secondary structure motif. Figure based on old experiments from Greenfield & Fasman, 1969[145].

CD measurements were performed in a Jasco J-815 spectropolarimeter (Jasco International, Japan) equipped with a Peltier temperature controller. The CD spectra were recorded on protein samples at 11-40 μM in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT using a 0.1 cm path length quartz cuvette at 25°C. A reference spectrum with buffer alone was subtracted before data were converted to mean residue molar ellipticity ($[\theta]_{\text{MRE}}$). Thermal denaturation from 5 to 95 °C was recorded on protein samples in the range of 1-40 μM using a stoppered 2 mm path length cuvette by increasing temperature at a rate of 1 °C/min and measuring the change in ellipticity at 222 nm. The midpoints of the thermal-denaturation curves (T_m) were determined from the first derivative, the second derivative, or by fitting to a sigmoidal transition curve.

Dimer dissociation K_D values at 25 °C were derived from thermal denaturation analysis followed by CD curves were following a method described by Marky and Breslauer[146]. This method assumes a two-state transition (i.e. the denaturation showed by the loss of CD signal at 222 nm is concomitant with dimer dissociation) and that the change in heat capacity during melting is negligible [147]. For this purpose, a series of CD thermal denaturation curves across a range of peptide concentrations (40, 20, 10, and 2 μM) were recorded. The midpoints of the thermal-denaturation curves (T_m) were determined for each curve and plotted as their reciprocal versus the natural logarithm of the protomer concentration. The K_D values at the temperature of interest can be obtained by extrapolation from linear regression of the obtained plot.

4. Size Exclusion Chromatography-Multi angle static Light Scattering (SEC-MALS)

Light scattering is a powerful technique for determining the molar mass of proteins in solution and studying the formation of oligomers. In a typical MALS experiment the sample is irradiated by a laser light at different angles and the intensity of the scattered light by the sample is measured at several angles give direct information about the molar mass of the macromolecule independently of its shape [148]. SEC may also be used to determine the size of a protein after calibration of the column, however the results are only valid when the protein is globular and approaches the shapes of a sphere. SEC coupled with “on-line” MALS detector serves as a further step of fractionation to avoid ambiguity that might result from the light scattered by a mixture of species, and the molecular mass determination by MALS will be independent of the SEC elution position [149].

These experiments were performed at 25 °C using a Superdex 200 Increase 10/300 GL column (GE HealthCare) attached in-line to a DAWN-HELEOS light scattering detector and an Optilab rEX differential refractive index detector (Wyatt Technology, California, USA). The column was equilibrated either with 20 mM Tris pH 8.0, 300 mM NaCl, 0.5 mM TCEP, 0.03% NaN₃, (0.1 μm filtered) in the case of ING5 constructs or PBS pH 7.4 (0.1 μm filtered) in the case of ING4ΔNLS. The SEC-MALLS system was calibrated with a sample of Bovine Serum Albumin (BSA) at 1 g/L in the same buffer. A sample of 100 μL of protein at 1-3 g/L was injected at 0.5 mL/min. Data acquisition and analysis employed ASTRA software (Wyatt Technology). Based on numerous measurements on BSA samples at 1 g/L under the same or similar conditions we estimate that the experimental error in molar mass is around 5%.

5. X-ray crystallography: Crystallization, Data Collection and Structure Determination.

X-ray crystallography is a technique that requires de generation of protein crystals under precipitant conditions [150]. Crystals are ordered arrays of atoms related by translations in one dimension (fibers), two dimensions (sheets) and three dimensions (lattice) [151]. The smallest repeating unit of the crystal is the unit cell and its shape is defined by the length of three axes (a, b, c) and three angles (α , β , γ). The unit cell has also internal symmetry and the smallest portion of structural information required to reconstruct the lattice through crystallographic symmetries and translations is the asymmetric unit. When X-rays are applied to a crystal they are diffracted by the atoms

from multiple parallel planes simultaneously and can be constructive (in phase) or deconstructive (out of phase). The planes that scatter X-rays are determined by the wavelength of the incident X-rays, the unit cell parameters, and the orientation of the crystal. In the unit cell, the size, shape and orientation of the molecule will determine the reflections that will occur on the detector. The arrangement of the atoms in the unit cell will determine the intensities of the diffracted X-rays. Those positions control the amplitude and phase. However, data collection only measures the intensities while the relative phase information necessary for the calculation of the electronic distribution in the unit cell is missing. This is the so called “phase problem” that has to be solved in order to determine the structure by using one of three different methods: experimental methods, direct methods and molecular replacement [151].

Crystallization screenings for ING5 Nt₁₋₁₀₅ mutants C19S and $\Delta\alpha 1$, and Nt₂₋₁₀₅ were performed at 21 °C with a MOSQUITO robot (TTP Labtech) using the vapor diffusion of sitting drop technique. The initial screens tested for Nt C19S were JCSGII, JCSGIII, JCSGIV, PACT suite, AmSO₄ suite, PEGs suite (Qiagen), PEG/ ION, PEGR_x, SALTR_x, Index, additive screen (Hampton Research) and Morpheus, JCSG plus (Molecular dimensions) in 96-well MRC plates (Molecular Dimensions). Drops consisted of 0.1 -0.3 μ L protein solution plus 0.1- 0.3 μ L reservoir solution and a reservoir volume of 60 μ L. The already known good conditions for crystallization of ING5 Nt were tested for screening on ING5 NtC19S (0.1 M MES pH 6.5, 10-30% MPD, 0-6% PEG3350) in 48-well MRC plates (Molecular dimensions). Drops consisted of 1-2 μ L protein solution plus 1-2 μ L reservoir solution and a reservoir volume of 200 μ L. In all plates protein concentration was 10 mg/mL. Crystals were only obtained in very few conditions in nanodrops (**Figure 5, A-B**). Condition JCSGII n.50 was selected as the best one for ING5 Nt C19S based on the similitude with the already tested and successful condition for ING5 Nt crystallization, and it was the starting point for optimization. The crystals were difficult to reproduce in bigger sitting drops but they were finally obtained in 0.1 M MES pH 6.5, 20-30% MPD and were cryoprotected with the reservoir condition plus 20% glycerol before collecting data at ID29 beamline at ESRF. These crystals were very delicate and depicted no diffraction (**Figure 5, C-D**).

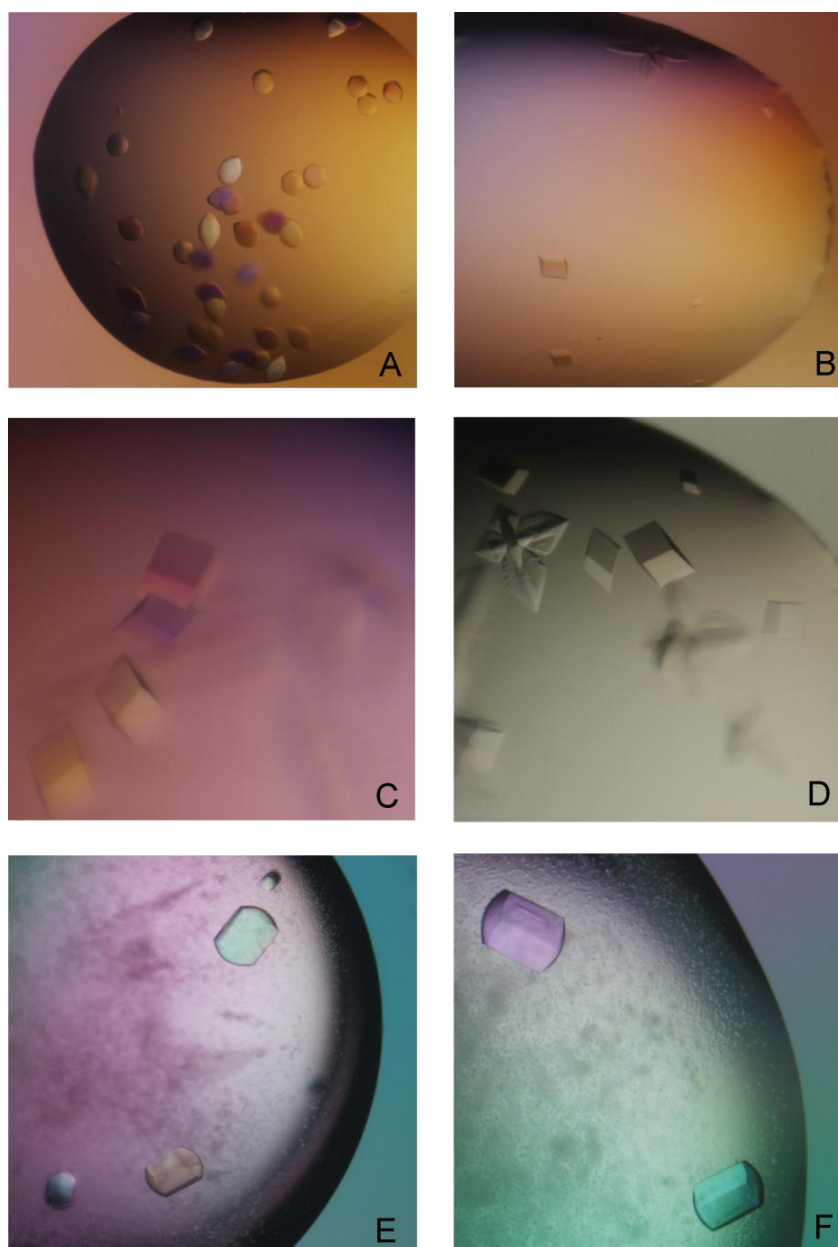


Figure 5. ING5 Nt₁₋₁₀₅ C19S crystallization. **A, B.** Crystals obtained in JCSGII commercial screening conditions no. 37 and no. 50, respectively. Condition JCSGII n.50 (0.1M MES pH6.0, 10% MPD) was selected as the best one based on the successful condition for ING5 Nt crystallization. This condition was further optimized: **C, D.** Crystals obtained in the optimized condition of 0.1 M MES pH 6.5, 20-30% MPD. These crystals were taken to ID29 beamline at ESRF synchrotron but did not diffract. A new set of crystals were obtained in Morpheus commercial screening condition no. C8 and this crystallization condition was optimized: **E, F.** Crystals obtained in the optimized condition of 0.1 M sodium MOPS/HEPES pH 7.5, 37.5% MPD_PEG1K_PEG3350, 0.03 M NaNO₃, 0.03 M Na₂HPO₄, 0.03 M (NH₄)₂SO₄. These crystals were taken to PROXIMA1 beamline at SOLEIL synchrotron depicting 3.1 Å diffraction.

A new set of crystals was obtained in Morpheus crystal screening plates. Conditions n.C4 and n.C8 (surprisingly both containing MPD) were selected as the best crystallization conditions and these crystals were reproduced in bigger sitting drops on 12-well plates (**Figure 5, E-F**). Best crystals grew on Morpheus n.C8 condition (0.1 M sodium MOPS/HEPES pH 7.5, 37.5% MPD_PEG1K_PEG3350, 0.03 M NaNO₃, 0.03 M Na₂HPO₄, 0.03 M (NH₄)₂SO₄). These were big crystals with 77% solvent which were cryoprotected directly with the reservoir condition and data was collected at beamline PROXIMA1 in SOLEIL synchrotron with a DECTRIS PILATUS3 6M detector, depicting 3.1 Å diffraction (**Figure 6A**).

The initial screenings tested for ING5 Nt Δα1 were JCSGI, JCSGII, JCSGIV, PACT Suite, AmSO4 Suite (Qiagen), SALTRx, PEG/ ION, Index (Hampton Research), and JCSG plus, Morpheus, MIDAS, Structure Screen, Proplex, NR-LBD (Molecular Dimensions) in 96-well MRC plates (Molecular Dimensions). Crystals were obtained in several conditions in nanodrops but the best crystallization conditions (JSCGIV n.65, JCSG Plus n.F1, INDEX n.D2, MIDAS n.2-5) were selected based on the quality of the crystal (best ordered, biggest ones, and most reproducible). Surprisingly all of these conditions contained jeffamine M-600 (a 600 molecular weight polypropylene glycol monoamine) and were reproduced in bigger sitting drops in 48-well plates. Best crystals of ING5 Nt Δα1 appeared in both, 0.1 M Tris or HEPES pH 7.0, and 35% jeffamine with Ionic Liquid Screen #20 from Hampton Research (1-Butyl-3-methylimidazolium trifluoroacetate) (**Figure 7**). These crystals were cryoprotected with 20% ethylenglycol and data was collected at XALOC beamline in ALBA synchrotron with very low resolution (5.09 Å or lower) (**Figure 6B**).

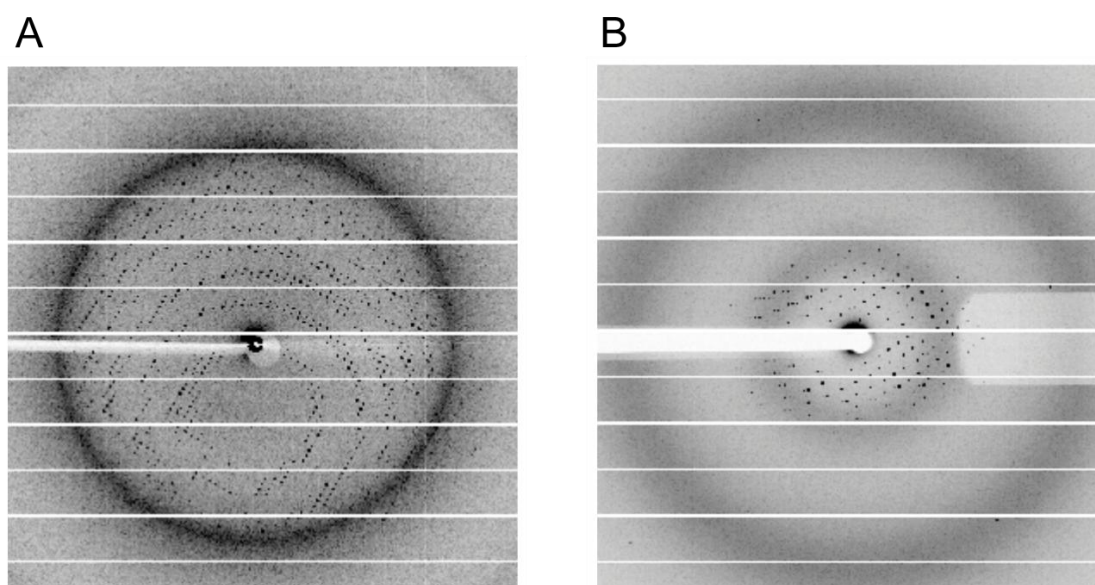


Figure 6. **A.** Diffraction pattern (3.1 Å) of the obtained crystals for C19S mutant using synchrotron radiation at the PROXIMA1 beamline (SOLEIL, Paris). **B.** Diffraction pattern (5.09 Å) of the obtained crystals for the Δα1 mutant using synchrotron radiation at the XALOC beamline (ALBA, Barcelona).

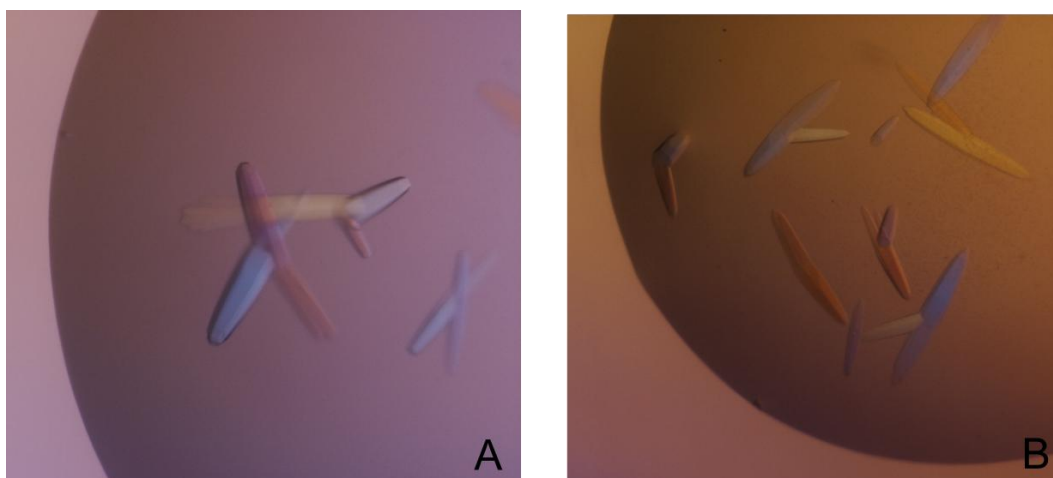


Figure 7. ING5 Nt₁₋₁₀₅ Δα1 crystallization. A-B. Best crystals obtained in the optimized condition of 0.1 M Tris pH 7.0, and 35% jeffamine with Ionic Liquid Screen #20 from Hampton Research (1-Butyl-3-methylimidazolium trifluoroacetate). These crystals were taken to XALOC beamline at ALBA synchrotron depicting low resolution diffraction.

ING5 Nt₂₋₁₀₅ initial screens were Morpheus, JCSG plus, Structure screen 1 and 2 (Molecular dimensions) PEG ION 1 and 2, PEG Rx1 and Rx2, Index (Hampton Research) and Pact (Qiagen). Crystals were obtained in several conditions in nanodrops but the best crystallization conditions were found in JCSG plus condition F4 (2.16) and Structure Screen conditions F1 and F12. All of these conditions contained jeffamine M-600 and were reproduced in bigger sitting drops in 48-well plates. These crystals were cryoprotected with 20% glycerol and brought to XALOC beamline in ALBA synchrotron where they did not diffract.

Diffraction data of the ING5 Nt C19S mutant obtained in SOLEIL was reduced using the MOSFLM [152] and scaled with SCALA [153]. The crystal structure of the wild type (PDB entry 5ME8) was used for molecular replacement with MOLREP [154]. The final model was reconstructed with Coot [155] and refined with REFMAC [156]. The final structural models were validated for geometry and stereochemistry restrictions [157] with MOLPROBITY [158], and were analyzed with PISA [159], CoilCheck [160], and Socket [161]. Figures were produced with PYMOL [162]. The coordinates and structure factors have been deposited at the Protein Data Bank (PDB) with the entry code 5MTO.

6. Small angle X-ray Scattering (SAXS)

SAXS is a method for low resolution structural characterization of proteins and oligomers in solution [163]. It is a complementary technique to other high resolution structural techniques such as X-ray crystallography or NMR; however, the sample does not need any intermediate step of preparation like crystallization or isotopic labeling, so

in a synchrotron scattering data can be collected in seconds, allowing for a rapid characterization of the sample. In practice, the sample is exposed to highly collimated X-rays and the scattered radiation is registered by a detector at small angles. The X-ray scattering curve (intensities versus scattering angle) is used to create a low resolution model (“SAXS envelope”) of the macromolecule. Besides, several parameters can be extracted directly from the scattering curve such as: molar mass (MM), radius of gyration (Rg), and maximum particle diameter (Dmax).

SAXS data was collected at 20 °C on BM29 beamline at The European Synchrotron (ESRF), in Grenoble, France with a 2D detector (Pilatus 1M) over an angular range $q_{min} = 0.01 \text{ \AA}^{-1}$ to $q_{max} = 0.5 \text{ \AA}^{-1}$ [164]]. Prior to measurements the samples were loaded onto a PD10 column equilibrated in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP and concentrated by ultrafiltration. Aliquots of the samples were prepared by dilution at different concentrations and the buffer was used as a blank for the SAXS experiments. The protein concentrations used were 4 and 7 g/L for ING5 Nt and 5 and 12 g/L for the ING5 Nt C19S mutant. The analysis of the data sets at different concentrations gave similar results. Data collection, processing and initial analysis were performed using the beamline software BsxCUBE. Further analysis was performed with ATSAS [165]. The Guinier approximation was used to calculate the radius of gyration using the PRIMUS software [166]. The distance distribution function, $p(r)$, was computed with the program GNOM [167] by optimizing the maximum particle dimension, D_{max} , to 89 Å. Low resolution *ab initio* reconstructions of the Nt domain and C19S mutant were built from the scattering data using GASBOR [168]. Ten independent GASBOR reconstructions were performed, and the most probable model was filtered with DAMSEL [169]. The low resolution shape of the ING5 Nt and C19S mutant are represented as an assembly of 230 dummy residues inside a search volume with a maximum diameter of 89 Å. Ten independent GASBOR reconstructions were performed and the most probable envelop was determined with DAMAVER [169]. The crystal structure of ING5 Nt C19S (PDB: 5MTO), or the homology model we have modeled based on the crystal structure of ING4 Nt (PDB: 4AFL), were fitted to the SAXS data with CRY SOL [170]. The goodness of the fitting is characterized by the discrepancy value χ^2 . Superposition of the bead model on the structures was carried out with SUBCOMB [171] and Chimera [172]. The model of ING5 Nt based on the crystal structure of ING4 Nt was built with FOLDX [136] and the 3 or 4 missing residues at the N-termini of the two protomers (absent in the crystal structure of ING4) were modeled with Coot in an extended conformation.

7. Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy is based on the fact that certain atomic nuclei have a quantum mechanical magnetic property known as spin [173]. The most common nuclei analyzed in proteins by NMR are ^1H , ^{13}C and ^{15}N because of their favorable magnetic properties (spin quantum number $I=1/2$) and their ubiquity in the polypeptide chain. However, while ^1H is naturally abundant in 99%, the natural abundance of ^{15}N and ^{13}C are less than 0.4% and 1%, respectively. For this reason, proteins studied by NMR are usually produced isotopically labeled in ^{15}N and ^{13}C . The nuclei with spin $I=1/2$ have two possible which in the presence of a magnetic field will display an energy difference between them proportional to the intensity of the magnetic field and dependent on the properties of the nuclei.

In a simple NMR experiment, the sample is irradiated with very short (μs) radiofrequency pulses (polychromatic radiation) at a determined power necessary to excite all the nuclei of a given kind under a magnetic field. In this way a net magnetization of the sample nuclei is created and the relaxation back to equilibrium is registered as an oscillating current in the coil of the spectrometer, this is the NMR signal. Fourier transformation of the time-domain signal yields the frequency-domain NMR spectrum with the signals of the different nuclei appearing at slightly different frequencies due to their different chemical environment (the chemical shift) [174] In protein NMR, the frequency of the nuclei will depend on the type of the amino acid and also on its particular chemical environment, which is modulated by the secondary, tertiary, and quaternary structure of the protein, as well as the binding of ligands and, to a lesser extent, by the solvent conditions (pH, salt and temperature).

In the case of proteins and other large molecules one dimensional experiments are not sufficient for analysis since there is a strong signal overlap. It is, then, necessary to use multidimensional NMR facilitating the analysis of protein spectra when two or more different nuclei are combined in the study. In triple resonance experiments, ^1H , ^{13}C and ^{15}N nuclei are studied generally for proteins. These experiments can provide information about how the atoms are chemically linked, how close they are in space and how rapidly they move with respect to each other.

For this thesis work, all NMR experiments were recorded at 25 °C on a Bruker Avance III spectrometer operating at 18.8 T (800 MHz of ^1H Larmor frequency) equipped with a triple resonance cryo-probe with z gradients. Chemical shifts were measured relative to internal 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS) for ^1H and calculated for ^{15}N and ^{13}C using IUPAC references [175].

7.1. Protein Backbone assignments

One of the most useful NMR experiments in protein NMR is the Heteronuclear Single Quantum Coherence spectrum [176], in which the signals correspond to the correlation of two different nuclei that are directly bonded. The most frequently used is the ^1H - ^{15}N -HSQC, which correlates NH groups, that in the case of proteins, will give information about peptide bonds of amino acids and, in addition, about the side chains of arginines, tryptophans, glutamines, asparagines and lysines. The previous assignment of the signals obtained in a ^1H - ^{15}N -HSQC is necessary to identify the amino acid residue to which that signal belongs; and this is the basis for protein-ligand interaction studies, particularly to map the region of interaction and measuring the affinity. Heteronuclear ^1H - ^{15}N correlations can be measured in the Transverse Relaxation Optimized Spectroscopy mode, which provides increased resolution at the expense of some sensitivity loss. TROSY is especially suitable for slowly tumbling deuterated proteins, but also it may also be helpful in non-deuterated proteins as well. With this strategy the average life of the signals that relax very fast is increased to facilitate their observation.

For the backbone amide assignment of ING5 Nt₁₋₁₀₅ and PHD finger of ING5, triple resonance experiments were acquired. These experiments allow for the frequency separation of overlapping signals in the 2D experiments, and provide useful information to identify the secondary structure of the protein from the chemical shifts using available methods such as TALOS [177] RCI (Random coil index) [178] and CSI [179]. The backbone assignments of ^{13}C , ^{15}N -ING5 Nt₁₋₁₀₅ and ^{13}C , ^{15}N -PHD finger of ING5 were obtained from 2D ^1H - ^{15}N -TROSY, ^1H - ^{13}C -HSQC spectra and from TROSY-based 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HN(CA)HA, and HN(COCA)HA experiments, acquired with non-uniform sampling (NUS) [180] in the case of ING5 PHD. Backbone ^1H - ^{15}N heteronuclear NOE measurements [181] were performed in the case of ING5 Nt₁₋₁₀₅ and each ^1H - ^{15}N NOE spectrum (both saturated and nonsaturated one) was acquired with 352 scans and 64 t_1 points in the indirect dimension with an overall recycling delay of 5 s to ensure the maximal development of NOEs before acquisition and to allow solvent relaxation, thus minimizing transfer of saturation to the most exposed amide protons of the protein from scan to scan [182]. The heteronuclear NOEs were calculated from the ratio of cross-peak intensities in spectra collected with and without amide proton saturation during the recycle delay.

The ING5 Nt₁₋₁₀₅ NMR sample was 250 μM protein in 50 mM MOPS pH 7.3, 1 mM DTT, 5% $^2\text{H}_2\text{O}$, 3 μM DSS (2,2-dimethyl-2-silapentane-5-sulfonate sodium salt) and protease inhibitors. ING5 PHD NMR sample was prepared at 220 μM in 20 mM Tris pH

8.0, 300 mM NaCl, 1 mM DTT, 5% $^2\text{H}_2\text{O}$, 44 μM DSS (2,2-dimethyl-2-silapentane-5-sulfonate sodium salt) and protease inhibitors.

NMR signal assignment was initially done automatically using MARS [183] and completed manually. Uniformly sampled data (the standard mode) spectra were processed with TOPSpin (Bruker) and NUS spectra with MddNMR [184]. The processed spectra were analyzed using Sparky [185].

To study the possibility of a different oligomerization mode of ING5 at acidic pH, titrations of pH were performed by stepwise addition of 0.4 – 4 μL of a HCl solution (37%) into a 400 μL sample of ^{15}N ING5 Nt at 180 μM in 50 mM MOPS pH 8.0, 1 mM DTT, 5% D_2O , 3 μM DSS and 0.5 mg of protease inhibitors. ^1H - ^{15}N - TROSY spectra were recorded for each step for comparisons.

7.2. H3K4me3 peptide binding

The synthetic peptide H3K4me3 corresponds to residues 1-15 of histone H3 plus an extra tyrosine residue at the C-terminus (ARTKQTARKSTGGKAY) to measure peptide concentration by ultra violet (UV) absorbance. This peptide has free N- and C-termini and was purchased from NeoMPS (Strasbourg). For H3K4me3 peptide preparation 3.2 mg of lyophilized powder were dissolved into 300 μL of 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT adjusted to 8.0 with concentrated NaOH and HCl. Concentration was measured by UV absorbance using extinction coefficients calculated with ExPASy ProtParam tool [140].

For the analysis of the interaction of ING5 with H3K4me3 we used ^1H - ^{15}N -HSQC spectra, which provide information of the chemical shift perturbations (CSP) of the NH groups caused by the presence of the peptide. For this experiment, titrations with histone synthetic peptide H3K4me3 were performed by stepwise addition of the concentrated peptide stock solution (3.8 mM) into 450-500 μL samples of 11 μM ^{15}N ING5 or 10 μM ^{15}N ING5 PHD in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, 5% (v/v) $^2\text{H}_2\text{O}$, 0.01% NaN_3 , 4 μM DSS. After each addition step a ^1H - ^{15}N -HSQC spectrum was recorded.

The measurement of the chemical shift perturbation (CSP) in a ligand concentration dependent manner by using ^1H - ^{15}N -HSQC allowed us to fit the data points to the equation that describes the binding equilibrium to a single set of binding sites. The error in CSP, estimated from the spectral resolutions in the processed data, is 0.006 ppm. Dissociation constants (K_D) were determined by data fitting (Origin, Microcal) to the

equation: $CSP = (K_D + [P] + [L] - \sqrt{(K_D + [P] + [L])^2 - 4 \cdot [P] \cdot [L]}) / (2 \cdot [P]) \cdot CSP_{max}$, where [L] is the concentration of the peptide, [P] is the concentration of PHD, CSP is the measured chemical shift perturbation, and CSP_{max} is the maximum CSP at saturation [101]. The adjustable parameters are the maximum CSP (CSP_{max}) and the dissociation constant (K_D). These values were determined by fitting of the data to the equation using Prism (GraphPad software). CSP was calculated from the equation: $CSP = \sqrt{((\Delta\delta_H)^2 + (\Delta\delta_N/5)^2) \cdot 0.5}$, where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift changes in the 1H and ^{15}N resonances, respectively, upon peptide addition. In this thesis K_D values were simultaneously measured for several residues with large CSP values and are reported as the average plus the standard deviation of those measurements.

7.3. Interactions with JADE-1L

Since the purification of soluble JADE 1L Iib domain was unsuccessful, we designed three peptides with overlapping sequences (**Table 2**) corresponding the region of the conserved Iib domain of JADE 1L that interacts with ING4. This sequence was RRLQLFTHLRQDLERVRNLTYMVTRR [139]. This strategy pretends to obtain peptides that are more soluble than the Iib domain and, perhaps, at least one of them containing all the necessary residues for binding. The peptide design was based on distributing the sequence stretches that are more hydrophobic (underlined in the sequence above) in three overlapping peptides with similar length and with a tyrosine residue to measure concentration by UV absorbance.

Table 2. Sequences of the overlapping peptides spanning the Iib conserved domain of JADE1L protein. Residues underlined correspond to hydrophobic clusters of residues.

name	sequence
A	RRLQLFTHLRQDLERVRNLT <u>Y</u> <u>M</u> <u>V</u> <u>T</u> <u>R</u> <u>R</u>
B	RQDLERVRNLT <u>Y</u> <u>M</u> <u>V</u> <u>T</u> <u>R</u> <u>R</u> EKIKRS
C	RREKIKRSVCKVQE <u>Q</u> <u>I</u> <u>F</u> <u>N</u> <u>L</u> <u>Y</u> <u>T</u> <u>K</u> <u>L</u> <u>L</u> <u>E</u>

The peptides were purchased from Apeptide and were found to be soluble in aqueous buffer. The lyophilized peptides were dissolved in water and their concentration was measured by absorbance at 280 nm ($\epsilon_{280} = 1490 \text{ M}^{-1} \text{ cm}^{-1}$ for all three peptides). From each peptide stock solution, a sample of 100 μL at 1.5 mM was prepared and their pH was raised to 7.0-8.0 with NaOH before being added to the ING5 Nt sample. ^{15}N labeled ING5 Nt sample was prepared at 150 μM in 800 μL of 20 mM Tris pH 8.0, 300 mM

NaCl, 1 mM DTT. From that stock, 4 different samples of 200 μL of ^{15}N ING5 Nt at 150 μM were obtained and, to three of them, 100 μL of a different peptide solution were added in order to have three different ING5 Nt / peptide (100 μM : 500 μM) samples of 300 μL and a reference sample of ING5 Nt at 100 μM in 300 μL without peptide. To diminish the possible mismatching of buffers, the four samples were dialyzed separately against 50 mM MOPS pH 7.0, 1 mM DTT o/n @ 4°C in Float-A-Lysers of 100-500 Da cut-off membrane (SpectrumLabs). Prior to NMR measurements, 5% D_2O and 4 μM DSS were added to all the samples and were introduced in 5 mm NMR tubes. ^1H - ^{15}N - TROSY spectra were recorded for each sample.

7.4. DNA binding assays

For the studies on ING4 binding to dsDNA, a titration of increasing amounts of a concentrated stock of the DNA (642 μM in 20 mM MES pH 6.5, 100 mM NaCl, 5 mM MgCl_2 and 1mM DTT) was done into a 25 μM sample of U- ^{15}N - labeled ING4 in 400 μL of the same MES buffer with 5 % $^2\text{H}_2\text{O}$ in a 5 mm Shigemi NMR tube. ^1H - ^{15}N -HSQC spectra were recorded (128 indirect points, 2 h total acquisition time) after each addition. To avoid differences in the chemical shifts due to changes in the ionic strength or pH in the DNA and protein samples, the oligonucleotides were previously desalted with a PD10 column (GE Healthcare) into 20 mM MES pH 6.5, 100 mM NaCl, 5 mM MgCl_2 and 1mM DTT, annealed, and concentrated by ultrafiltration (3 kDa cut off), and the protein was dialyzed in the same batch of buffer used to prepare the DNA. The oligonucleotides used for the NMR experiment did not have any 6-FAM fluorescent probe. TopSpin (Bruker) and Sparky [185] were used for NMR data processing and spectral analysis, respectively. Dissociation constants were determined by the combined ^1H and ^{15}N CSP fitting as described above using Prism (GraphPad software). The ^1H - ^{15}N -HSQC spectrum for the characterization of ^{15}N ING4 ΔNLS mutant was measured on a 27 μM sample in PBS pH 7.4.

7.5. Paramagnetic relaxation enhancement

Paramagnetic relaxation enhancement (PRE) can be employed to probe the structure and dynamics of proteins. One of the most used spin label reagent is the MTSL (*S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl) methyl methanesulfonylthioate). This compound is also known as MTSSL, methanethiosulfonate spin label), and serves to attach a spin label selectively to cysteine residues via a disulphide bond. The reaction will introduce a group at a solvent exposed region of the protein that won't significantly alter the protein structure [186]. MTSL possesses an unpaired electron that reduces the intensity of NMR signals of atoms in a distance dependent manner, with an approximate distance

range of around 20 Å [187]. In this thesis PRE analysis has been used to differentiate between two models of the ING5 Nt dimer in solution.

For these measurements, the solvent exposed cysteine 19 was labeled with the thiol reactive MTSL using a modification of a previously described protocol [188]. For that purpose the C75S mutant was created to eliminate this other reactive thiol. The Nt₁₋₁₀₅ C75S protein sample was buffer exchanged to 50 mM MOPS pH 7.0 with a desalting PD10 column (GE Healthcare) to remove DTT that was present in the purification buffer. Sample was then concentrated with an amicon filtering unit of 3K cut off (Millipore) until 800 µL at 200µM and divided into two different samples. 1 mM DTT was added to one of the samples that we used as “reference” and the other sample was taken for labeling reaction. Site-directed spin labeling was achieved by incubating the second sample containing ¹⁵N ING5 Nt₁₋₁₀₅ C75S with a 5-fold molar excess of the spin label reagent MTSL at 25 °C for 2 h in the absence of light. Prior to the NMR experiment the homogeneity of the sample and the presence of MTSL molecule covalently bound to ¹⁵N ING5 Nt₁₋₁₀₅ C75S was confirmed by MALDI-TOF (the MTSL molecule increments the molecular weight of the polypeptide chain in 186.3 Da). 5% ²H₂O and 4 µM DSS were added to the samples prior to NMR experiments and ¹H-¹⁵N-TROSY spectra were recorded at 800 MHz and 25 °C. Spectra were processed with TopSpin (Bruker) and the intensity analysis was done using Sparky [185].

8. Ion mobility coupled to electrospray ionization mass spectrometry (ESI-IM-MS) experiments under non denaturing conditions

With the aim of detecting the possible heterodimerization of ING4/ING5, ESI-IM-MS experiments under non denaturing conditions were performed. The use of ion mobility (IM) experiments coupled to mass spectrometry (MS) allows to differentiate molecules according to their charge state (CS), collision cross-section (CCS) and shape by measuring mass spectra and drift time distribution of m/z ions continuously and simultaneously and thus, reflecting on the gas phase conformation of the proteins. This joint technique is particularly useful to get insights on the oligomerization state of proteins as it has the capability to separate in the gas phase ions of the same m/z, such as monomers, dimers, trimers or higher order aggregates. The technique can be performed in experimental conditions which preserve the non covalent protein-protein complexes and their conformational ensemble in the gas phase and is therefore a powerful tool for the study of macromolecular structures and protein-protein interactions [189].

ESI-IM-MS experiments were conducted on a Synapt HDMS (Waters) quadrupole-traveling wave IMS-oaTOF mass spectrometer equipped with an Advion TriVersa NanoMate (Advion Biosciences). Prior to IM-MS experiments, the samples were buffer exchanged to 100 mM Ammonium Acetate (NH₄OAc), pH 6.92 and both proteins were then incubated o/n @ 37°C at equal molar concentration (85 μM). Before sample injection, the samples were further desalted using a Micro Bio-Spin 6 column (Biorad) using NH₄OAc 100 mM as the exchanging buffer. ESI-IM-MS spectra of ING5 Nt₁₋₁₀₅, ING4 Nt and the mixture of both proteins were recorded under non denaturing instrumental conditions in the positive ion mode. The Advion source was working in direct infusion mode allowing automated chip-based nanoelectrospray. Spray voltage was 1.75 kV and delivery pressure 0.5 psi. Backing pressure, sampling cone voltage, trap collision energy, transfer collision energy, and source temperature were set to 5.83 mbar, 40 V, 6V, 4 V and 20 °C, respectively. The IM gas flow was kept at 23 mL/min, gas pressure in ion mobility cell at 4.49·10⁻¹ mbar and gas flow rate in the trap cell at 5 mL/min. Wave height and wave velocity in the ion mobility cell were set to 9V and 350 m·sec⁻¹, respectively. External mass calibration was achieved using a cesium iodide solution in m/z range of 500 to 8000. Data was acquired and processed with MassLynx software v 4.1 (SCN 704). Mass spectra were deconvoluted to the average masses with integrated algorithms in Masslynx and IM-MS data was processed with Driftscope software vs. 2.4. Individual protein samples gave experimental molecular weights of 12541.74 ± 4.71 Da for ING5 Nt¹⁻¹⁰⁵ and 13755.41 ± 2.86 Da for ING4 Nt, in agreement with the theoretical values (12545.170 Da and 13754.48 Da respectively, for the monomeric proteins). The ING5 Nt₁₋₁₀₅:ING4 Nt protein mixture at molar ratio of 1:1 was used for the analysis of the heterodimer formation. The assignment by mass to the heterodimer in the mixture was done based on the observed masses in the independent homodimer mixtures (**Table 3**).

Table 3. Observed masses in the spectra.

Deconvoluted average masses detected	Assignment
ING4 Nt sample	
13755.41 ± 2.86	monomer ING4
27508.17 ± 2.66	homodimer ING4
41279.66 ± 1.63	homotrimer ING4
ING5 Nt sample	
12541.74 ± 4.71	monomer ING5
25087.37 ± 0.91	homodimer ING5
37632.09 ± 2.70	homotrimer ING5
50248.53 ± 7.34	homotetramer ING5
ING4/ING5 sample	
12543.59 ± 2.94	Monomer ING5
25087.37 ± 5.48	Homodimer ING5
13755.34 ± 2.38	Monomer ING4
27508.70 ± 2.23	Homodimer ING4
26303.02 ± 7.55	Dimer (1:ING4-1:ING5)
40054.93 ± 5.30	Trimer (2:ING4-1:ING5)
38848.09 ± 2.19	Trimer (1:ING4-2:ING5)

9. Mammalian cell culture and transfection.

Human cell line HEK-293T (embryonic kidney epithelium) and NIH3T3 cells (immortalized mouse fibroblasts) were cultured in DMEM (Gibco) all supplemented with 10% FCS and 1% penicillin/ streptomycin (10.000 U/mL, Gibco) at 37 °C and 5% CO₂. Cells were passaged when they reached 70-80% confluence at 1:8 dilution with 0.05% trypsin. Transfection of 293T and NIH3T3 cells was performed using a standard calcium phosphate method after seeding 3×10^6 cells on 10 cm plates. Selection of stable transfected NIH3T3 cells was performed at 2 µg/mL puromycin during 2-3 weeks after transfection and the antibiotic was maintained on the media at a low dose (0.5 µg/mL).

10. Cell proliferation assay

To evaluate and compare the effect on cell proliferation stably NIH3T3 cells expressing ING5 or ING5 mutants were seeded at a density of 900 cells per well in 96-well plates and incubated overnight.. Cells were rinsed with phosphate buffered saline and stained with crystal violet for counting after 0, 24, and 72 hours of proliferation. The stain was solubilized in 10% acetic acid after air drying and diluted in water for absorbance measurement at 595 nm.

11. Cell Cycle Analysis

To analyze the effect of the overexpression of ING5 and its mutants on cell cycle profiles, NIH3T3 cells with stable expression of the corresponding genes were harvested prior to confluence from a 6-well plate, fixed with 70% (vol/vol) ethanol, and stored at 20°C for >12h. After ethanol was removed by centrifugation, pellets were resuspended in PBS buffer containing 40 µg/mL propidium iodide and 100 µg/mL RNase A and incubated for 30 min at room temperature. The DNA contents were measured as the propidium iodide signal to determine cell cycle profiles using FACSCanto (BD Biosciences) and then analyzed by FACS Diva software. At least 20000 cells were analyzed from each sample.

12. Immunoprecipitation and Western Blot analysis

Transiently transfected HEK-293T cells on a 10 cm plate were washed twice with PBS, and total cell extracts were prepared by lysis in 1 mL of cold radioimmunoprecipitation assay (RIPA) lysis buffer (EMD Millipore) with complete mini EDTA-free protease and phosphatase inhibitor cocktail tablets (Roche), PhosphoStop (Roche), and 0.1% SDS (Life Technologies), as previously described [190]. The lysates were incubated for 10 min on ice and then centrifuged for 12 min at 15,000 g. The supernatants were divided into fresh tubes on ice. For immunoprecipitation, mouse monoclonal anti-HA (12CA5 clone, Roche) and mouse monoclonal anti-AU5 (Abcam) primary antibodies were used. One μg of the antibody was incubated with 1 mg of the lysate for 16 h at 4 °C. Then, 20 μL of BSA blocked agarose-conjugated protein A/G beads (Santa cruz) were incubated for 1 h at 4 °C, and after washing 3 times with RIPA buffer, samples were probed by Western Blotting. Samples were resolved by 12% SDS-polyacrylamide gel electrophoresis (PAGE) and transferred onto Protran (Whatman) nitrocellulose membranes of 0.45 μm . Nonspecific binding was blocked by 1 h incubation with blocking buffer before membranes were probed overnight at 4 °C with primary antibodies diluted in blocking buffer (5% milk in Tris-buffered saline without tween20). After extensive washing with TBS-T, specific bands were detected on Hyperfilm (GE Healthcare) using horseradish peroxidase (HRP)-conjugated goat anti mouse Fc γ fragment specific secondary antibody (1:10000; Jackson ImmunoResearch) and the ECL detection system (Biorad).

13. Immunofluorescence

NIH3T3 cells stably expressing ING5 and its mutants were plated in coverslips in 24-well plates at a density of 40000 cell per well. Cells were fixed with 4% paraformaldehyde in PBS for 20 min at RT, then treated with 0.1% Triton X-100 in PBS for 10 min at RT, for membrane permeabilization and incubated for 1 h in blocking solution (2% BSA in PBS, 0.01% NaN_3 , glycine 50 mM). ING5 and mutants were detected with anti-HA (12CA5 clone, Roche). Primary antibody was diluted 1:200 in blocking solution. The secondary antibody used was Alexa Fluor 594 antimouse (Invitrogen A11005, dilution 1:5000). For the nuclear shape studies of NIH3T3 cells stable expressing ING5 and ING5 mutants, Hoechst 33342 (reference H3570, Invitrogen) for the nuclei and Phalloidin-Tetramethylrhodamine B isothiocyanate (reference P-1951, Sigma) for the cytoplasm were used. Washing between the different steps were carried out with PBS. Cells were finally mounted on to microscope slides using Vectashield

containing DAPI (Vector) in the cases where Hoechst 33342 was not used. Slides were examined using a Zeiss Axiomager D1 fluorescence microscope.

14. Fluorescence Electrophoretic Mobility Shift Assay (EMSA)

Oligonucleotides used for the EMSA assays (**Table 4**) were designed to generate single strand (ss), double strand (ds), or primed (p) DNA ligands labeled with (6-Carboxyfluorescein (6-FAM) at the 5' end. To avoid any quenching effect the dsDNA was designed to have two extra bases at the 5' end. Oligonucleotides were chemically synthesized and HPLC-purified by Thermo Scientific or Sigma-Aldrich, and were solubilized in Tris-EDTA buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, pH 8.0) to a concentration of 100 μ M. For the preparation of the different dsDNA ligands 1:1.2 ratio of labeled: unlabeled oligonucleotides were mixed and diluted in annealing buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) to the desired final concentration and annealed by incubation in boiling water for 5 min and slow cooling down to room temperature. EMSA experiments were performed by incubating increasing concentrations of ING4, ING5 or their different domains with fluorescent DNA ligands at a final concentration of 0.1 μ M in a 15 μ l reaction mixture containing 20 mM Tris-HCl pH 8.0, 5 mM MgCl₂, 50 mM NaCl, 2 mM DTT and 0.1 g/L BSA. After incubation for 10 min at room temperature, 5% glycerol was added and the reaction products were separated on a 6% native polyacrylamide gel run at room temperature in cold 0.5 X Tris-Borate-EDTA buffer for 50 min at 80 V. Labeled nucleic acid fragments were detected by fluorescence imaging (ImageQuant LAS4000, GE Healthcare) and quantification of protein-nucleic acid complexes was performed with ImageQuant TL image analysis software (GE Healthcare). The apparent equilibrium dissociation constants were determined using a Hill equation [191] with a single-site binding model using Prism (GraphPad software) from the mean of three independent experiments.

Table 4. Oligonucleotides (5' to 3' sequences) used for DNA binding studies. The dsDNA substrates with 18 bp, 32, 22, and 10 bp were obtained by annealing oligonucleotides A1 with B1, A3 with B2, C1 with C2, and D1 with D2, respectively. The pDNA substrate contained an 18 bp long dsDNA region and a 12 base long ssDNA region, and was obtained by annealing oligonucleotides A2 and B1. The oligonucleotides used for the NMR experiment (dsDNA substrates of 32bp and 10bp) did not have any 6-FAM fluorescent probe.

Name	Sequence
A1	TGGCCTGCAGGCATGCAA
A2	GTCCGTACGTTCTGGCCTGCAGGCATGCAA
A3	GATGAGATTGAGGCTGGCTGGCCTGCAGGCATGCAA
B1	6-FAM-GCTTGCATGCCTGCAGGCCA
B2	6-FAM-GCTTGCATGCCTGCAGGCCAGCCTCAATCTCATC
C1	GAGTGTGGTGTACATGCACTAC
C2	6-FAM-GTAGTGCATGTACACCACACTC
D1	ATACGATGGG
D2	CCCATCGTAT

RESULTS



1. ING5 is a dimer and a bivalent reader of H3K4me3 mark

1.1. Biophysical characterization of ING5

To study the structural organization of ING5 the domain boundaries were defined based on those previously identified for ING4 [102]. ING5 and ING4 share high sequence homology and secondary structure predictions give very similar results: a conserved N-terminal domain with two long α -helices forming a coiled-coil, a long and disordered central region rich in positively charged amino acids that contains the Nuclear Localization Signal and is, therefore, named NLS, and a highly conserved C-terminal domain folded into a PHD finger [72]. The high sequence identity at the N-terminal domain suggests that ING5 will also form dimers as ING4 does.

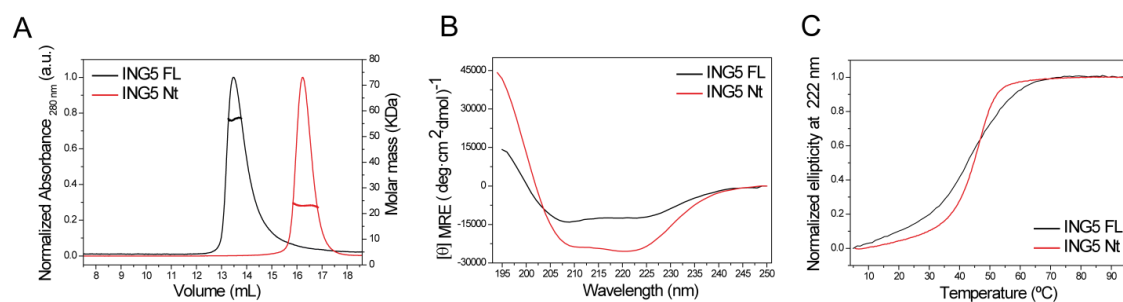


Figure 1. ING5 dimerizes in solution through its N-terminal helical domain. (A) SEC-MALS analysis of ING5 (black) and Nt (red). The thin line corresponds to the axis on the left and the thick line to axis on the right. All data were obtained in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, at 25 °C. The exclusion volume of the column is 8.7 mL. (B) CD spectra of ING5 (black) and Nt (red) in the same conditions as in A. (C) Thermal denaturation of ING5 (black) and Nt (red) in the same buffer as in A.

ING5 has a molar mass of 57 kDa as measured by SEC-MALS (**Figure 1A**), and this is consistent with the theoretical calculated mass of a dimer (60 kDa). However, the elution volume of ING5 is smaller than the expected one for a dimer, indicating that ING5 has an elongated shape and/or large regions that are flexible and disordered [192]. In both situations, a protein with these characteristics will elute at a smaller volume than a globular protein of the same mass.

The circular dichroism spectrum of ING5 shows two minima at 209 and 223 nm and a positive ellipticity below 200 nm, which is indicative of a high content of helical structure (**Figure 1B**). The low absolute ellipticity indicates, however, that a large part of the chain is not helical. The N-terminal domain of ING5 (ING5₁₋₁₀₅, hereafter named Nt) is also a dimer (**Figure 1A**) with a measured molar mass of 23 kDa (the theoretical calculated molar mass of the ING5 Nt dimer is 25 kDa). Both ING5 and the Nt domain show cooperative thermal denaturations, with similar midpoint melting temperatures (43 and 45 °C, respectively), indicating the presence of a defined tertiary fold for both proteins (**Figure 1C**). The denaturation is, however, more cooperative in the case of the

isolated Nt domain. The presence of different structural regions in ING5 is probably the reason for the apparent less cooperative thermal denaturation. The thermal denaturation of ING5 is irreversible, with all the protein precipitated in the cuvette, while the denaturation of the Nt domain is mostly reversible. The shape of the Nt spectrum at 25 °C is recovered after renaturation by cooling down, with a small loss of signal intensity due to partial protein precipitation (about 15 %).

The ^1H - ^{15}N HSQC spectrum of the 60 kDa ING5 dimer is similar to that of ING4 (**Figure 2**) [102]. It shows a similar pattern of a set of sharp and dispersed backbone amide resonances corresponding to the PHD finger (**Figure 4B**) and a few non-dispersed resonances in the ^1H dimension (**Figure 2**). The small number of non-dispersed signals in the spectrum of ING5 is likely due to the higher pH (necessary to have it in soluble form), which increases the rate of exchange with the solvent of the disordered regions. The observation of a single set of resonances for the two PHD fingers of the dimer indicates that they are independent and chemically equivalent. Altogether these results suggest that the structural organization of ING5 is very similar to that of ING4: an N-terminal dimeric coiled coil domain and a C-terminal PHD connected by a flexible NLS segment, the three regions being structurally independent with no interactions among them.

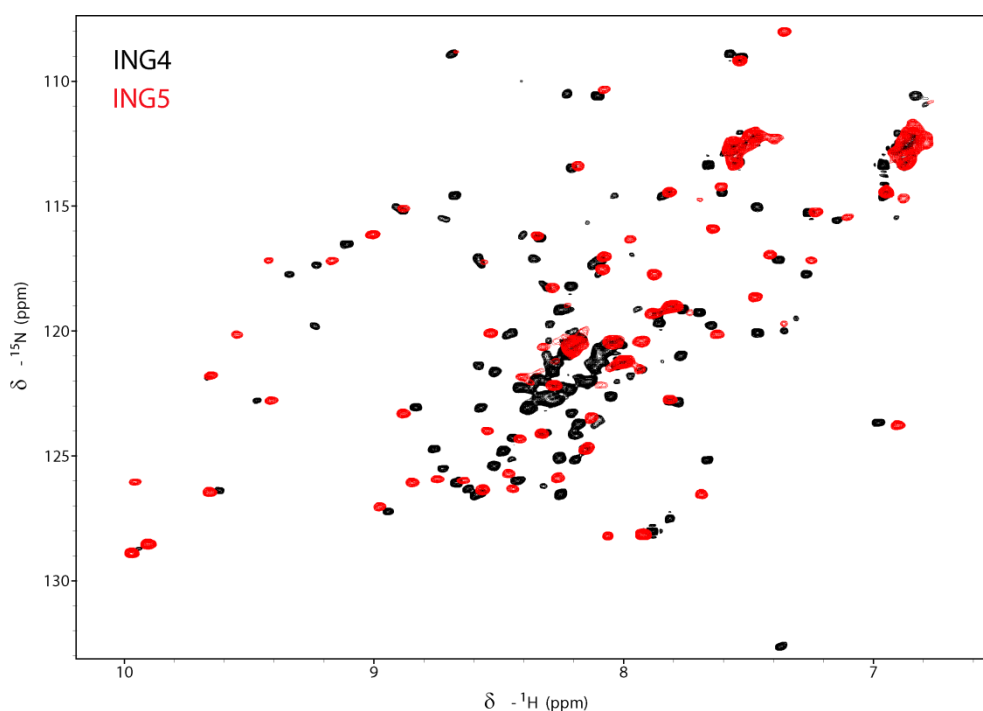


Figure 2. Structure of ING5 in solution. Superposition of ^1H - ^{15}N HSQC spectra of ING5 (red) and ING4 (black) at 25 °C and 800 MHz. The dispersed signals belong to the PHD and there is only one set of PHD signals. Therefore ING5 structure is similar to ING4, a dimer with a disordered NLS domain and two PHD fingers chemically equivalent. The spectrum of ING5 was obtained on an 11 μM sample in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT. The spectrum of ING4 was measured on a 30 μM sample in 20 mM sodium phosphate pH 6.5, 50 mM NaCl.

1.2. ING5 binds H3K4me3

The binding of ING5 to H3K4me3 peptide was examined in solution by NMR as previously done for ING4 [102]. Due to the low solubility of ING5, we had to use different buffer conditions than those used for ING4 (pH 8.0 and 300 mM NaCl, while it was pH 6.5 and 200 mM NaCl in the ING4 sample). We measured the chemical shift perturbations (CSP) caused on ING5 signals upon binding to increasing concentrations of the peptide (**Figure 3A**). During the titration, many dispersed signals, corresponding to those of the PHD domain, suffered a CSP until the binding reaction reached the saturation point. Although the majority of the signals can be observed in all titration points, there are a few whose intensities are strongly reduced at intermediate titration points (**Figure 3B**).

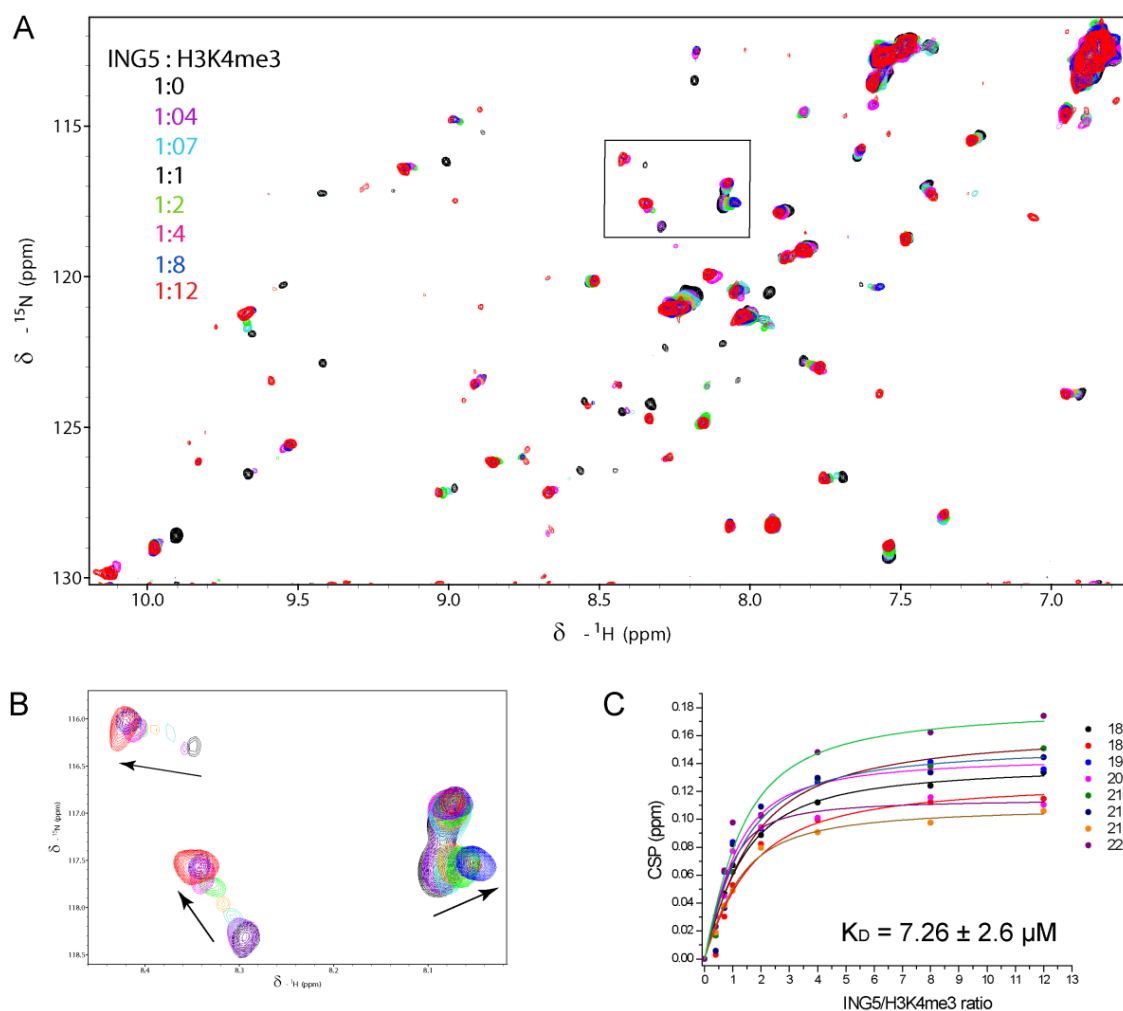


Figure 3. A. ING5 binding to H3K4me3. (A) Superposition of eight ^1H - ^{15}N -HSQC spectra of ^{15}N ING5 at 11 μM in 450 μL of 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, 3% D $_2\text{O}$, 0.01% NaN_3 and 4 μM DSS in the presence of increasing amounts of histone H3K4me3 peptide at 25 $^\circ\text{C}$. Each point of the titration is shown with a different color, from 1:0 (in black) to 1:12 (in red). (B) Zoom of the framed region in A with some of the signals suffering chemical shift perturbations upon the addition of increasing amounts of histone H3K4me3 peptide. (C) Chemical shift perturbation of the eight signals that shift the most (CSP higher than the average plus one standard deviation). Each dot represents a point along the titration and the line the fitting to a model of one set of binding sites. The numbers on the right correspond to the individual residues, and the dissociation constant (K_D) is the average of the values calculated for them with the standard deviation. The error in the CSP measurement is estimated to be ± 0.006 ppm.

For the analysis of the data, the CSP were plotted versus ING5:H3K4me3 ratios (**Figure 3C**) and the curves were fitted to a model of one set of binding sites, as described in the methods section. The dissociation constant was calculated as an average of the eight values obtained from the fitting of eight different signals whose CSP were higher than the average plus the standard deviation (**Figure 3C**). ING5 binds H3K4me3 with a $K_D = 7.3 \pm 2.6 \mu\text{M}$, a value that is around 3 times higher than the K_D measured previously [86] for the isolated ING5 PHD finger ($K_D = 2.4 \pm 1.0 \mu\text{M}$ by fluorescence). We have to take into account the differences in the experimental conditions and the used techniques. In addition, the peptide used in the NMR titration is a bit longer (residue 1-15 instead of 1-12) and contains a tyrosine residue at the C-terminus to measure the concentration with high precision (the peptide used by Champagne et al. 2008 [86] does not contain any tyrosine or tryptophan and no information on how the concentration was measured is available). The affinity of ING5 for H3K4me3 is on the same order that measured for the isolated PHD and full length ING4 by NMR ($3.9 \pm 1.0 \mu\text{M}$ and $1.3 \pm 1.0 \mu\text{M}$, respectively), considering the errors and the different buffers used [85] [102].

2. The C-terminal PHD finger of ING5 recognises H3K4me3

The crystal structure of the isolated ING5 PHD finger has been previously reported bound to an H3K4me3 peptide [86], and consists of the canonical PHD fold, with its characteristic C_4HC_3 zinc finger motif. We have analyzed the structure of the PHD in solution by MALS and NMR and found that it is a monomeric protein (**Figure 4A**) with a ^1H - ^{15}N NMR spectrum (**Figure 4B**) very similar to the set of dispersed signals in the spectrum of ING5 (**Figure 2 and 3A**). After assignment of the backbone resonances of the isolated PHD it was found that for 83% of the backbone amide signals of the PHD, the ^1H and ^{15}N resonance frequencies closely match with those of a corresponding signal in the spectrum of full-length ING5 (chemical shift deviations < 0.065 ppm). This result indicates that the Nt and the PHD are independent and non-interacting domains linked by a flexible NLS segment.

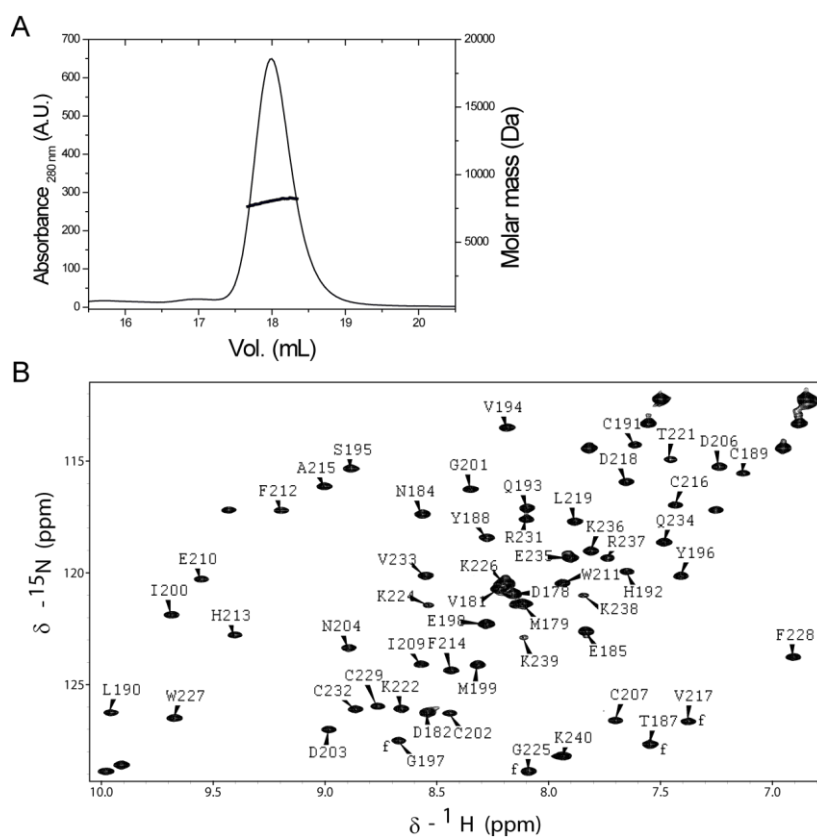


Figure 4. (A) SEC-MALS of the PHD finger of ING5 at 25 °C in 20 mM Tris pH 8.0, 300 mM NaCl, 0.5 mM TCEP. The thin line corresponds to the axis on the left and the thick line to the axis on the right. (B) ^1H - ^{15}N -HSQC spectrum of the ING5 PHD finger with backbone amide signal assignment. The spectrum was recorded at 25 °C and the ING5 PHD sample was prepared at 220 μM in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, 5% $^2\text{H}_2\text{O}$, 44 μM DSS (2,2-dimethyl-2-silapentane-5-sulfonate sodium salt) and protease inhibitors. The signals labeled with “f” are folded in the ^{15}N dimension and their real chemical shift is 18.5 ppm smaller. Non assigned peaks correspond to side chain NH resonances.

We have analyzed by NMR, as it was done for full length ING5, the binding of the isolated ING5 PHD finger to an H3K4me3 peptide (**Figure 5**), and mapped the binding site (**Figure 6**). Many of the assigned dispersed signals of ING5 PHD exhibit chemical shift perturbations upon addition of H3K4me3 peptide which mimic those observed with the full length protein (**Figure 5A, B**). The titration curves obtained for H3K4me3 peptide are shown in **Figure 5C**, together with the calculated dissociation constant (K_D). The K_D was calculated as an average of the ten values from the fitting of those different signals whose CSP were higher than the average plus one standard deviation (**Figure 5C**). This calculation yields a dissociation constant of $K_D = 17.9 \pm 2.5 \mu\text{M}$ at 25 °C (average and standard deviation; **Figure 5C**). This value is about two-fold larger than the measured for full length ING5 ($K_D = 7.3 \pm 2.6 \mu\text{M}$), perhaps due to the loss of avidity in the isolated PHD as compared with the ING5 dimer.

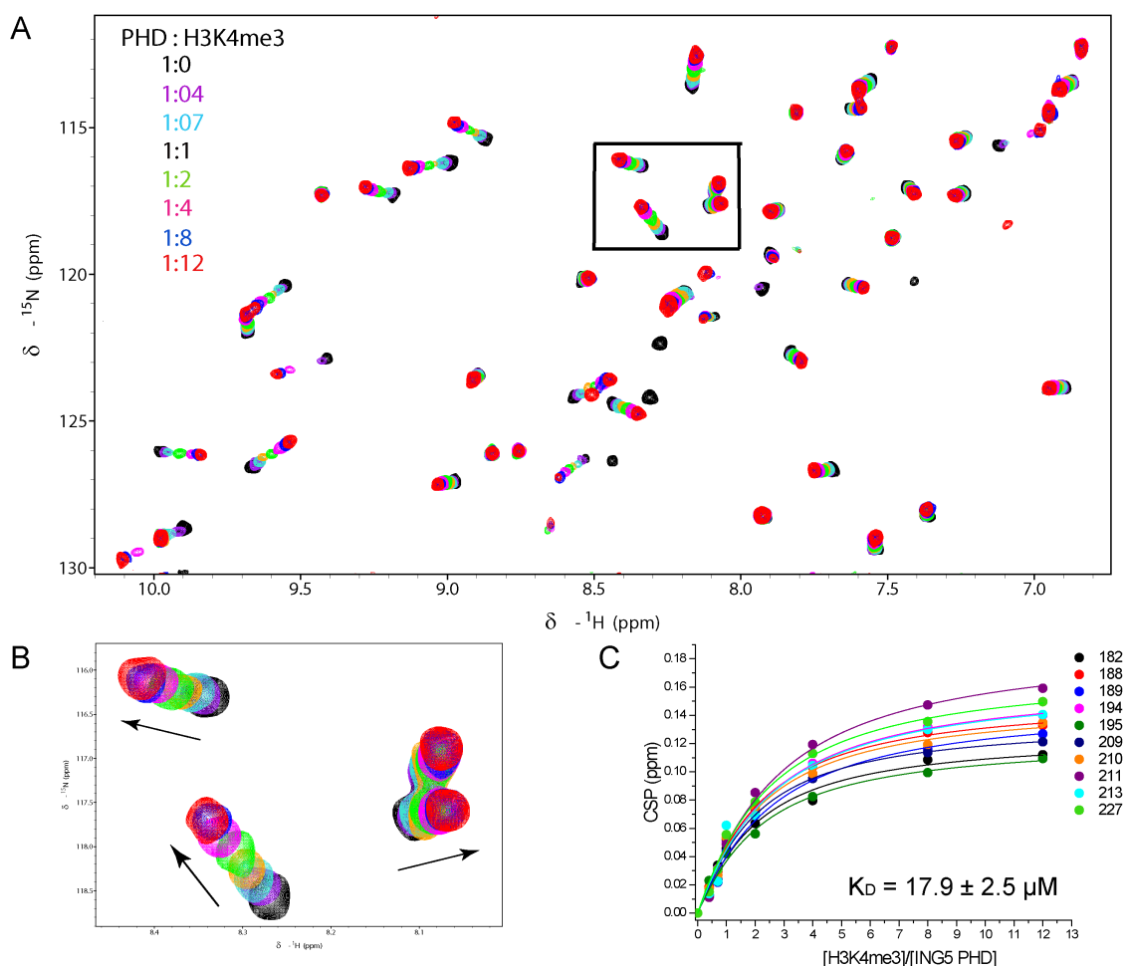


Figure 5. A. ING5 PHD binding to H3K4me3. (A) Superposition of eight ^1H - ^{15}N -HSQC spectra of ^{15}N PHD of ING5 at $10\ \mu\text{M}$ in $500\ \mu\text{L}$ of $20\ \text{mM}$ Tris pH 8.0, $300\ \text{mM}$ NaCl, $1\ \text{mM}$ DTT, 3% D $_2\text{O}$, 0.01% NaN_3 and $4\ \mu\text{M}$ DSS in the presence of increasing amounts of histone H3K4me3 peptide at $25\ ^\circ\text{C}$. Each point of the titration is shown with a different color, from 1:0 (in black) to 1:12 (in red). (B) Zoom of the framed region in A with some of the signals suffering chemical shift perturbations upon the addition of increasing amounts of histone H3K4me3 peptide. (C) Chemical shift perturbation of the signals that shift the most. Each dot represents a point along the titration and the line is the fitting to a model of one set of binding sites. The dissociation constant (K_D) is the average of the calculated values plus one standard deviation.

The CSP measured in the presence of a 12 molar excess of peptide is represented for each PHD residue in **Figure 6A** and on the surface of the crystal structure of the PHD in **Figure 6B**. There are many residues that experience large perturbations in their chemical shifts, indicating a large interaction surface. There is a strong similarity with the pattern of changes experienced by ING4 PHD [101], indicating that the binding mode in solution is similar in both proteins (as suggested by the crystal structures), and that the differences in the calculated K_D , are mostly due to the different experimental conditions and methods used.

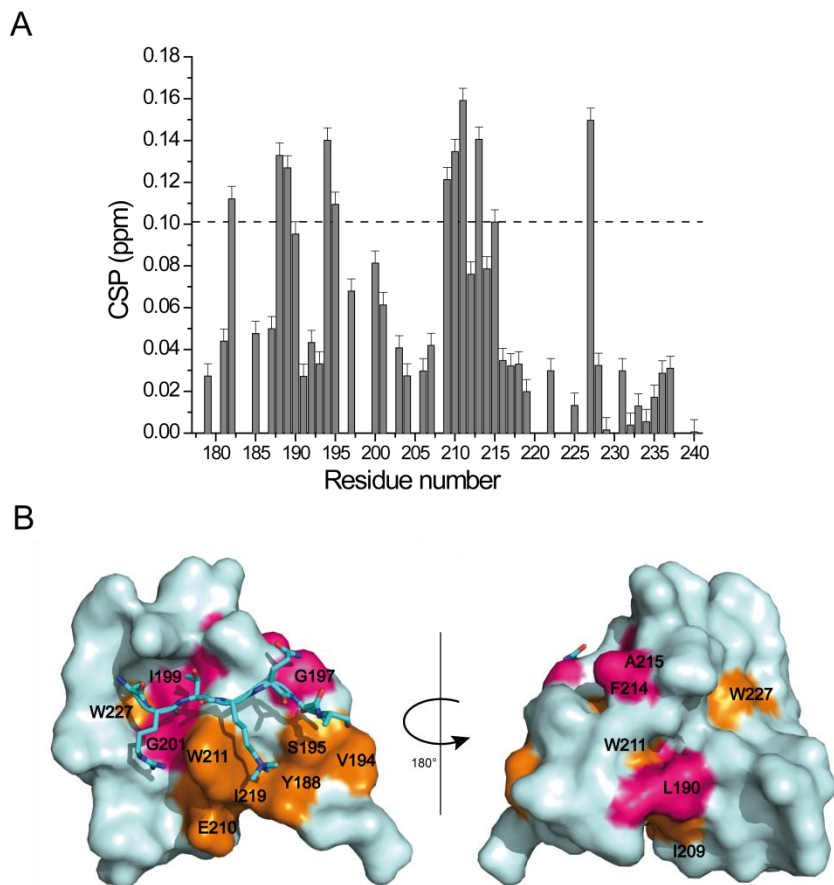


Figure 6. ING5 PHD CSP mapping and comparison with the crystal structure of the complex with a histone H3K4me3 peptide. (A) Binding histograms showing the CSP observed for each residue in the ^1H - ^{15}N HSQC spectrum of ING5 PHD in the presence of 1:12 excess H3K4me3 peptide. The dashed line represents the average plus one standard deviation. The experimental error in the CSP is estimated to be 0.006 ppm. (B) Surface representation of the ING5 PHD finger (in light cyan). Those residues with a CSP upon binding to peptide H3K4me3 larger than the average plus one standard deviation (182, 188, 189, 194, 195, 209, 210, 211, 213, 227) are highlighted in orange, and those with a CSP larger than the average (190, 197, 200, 201, 212, 214, 215) in magenta.

3. DNA binding of ING4 and ING5 through the NLS disordered central region.

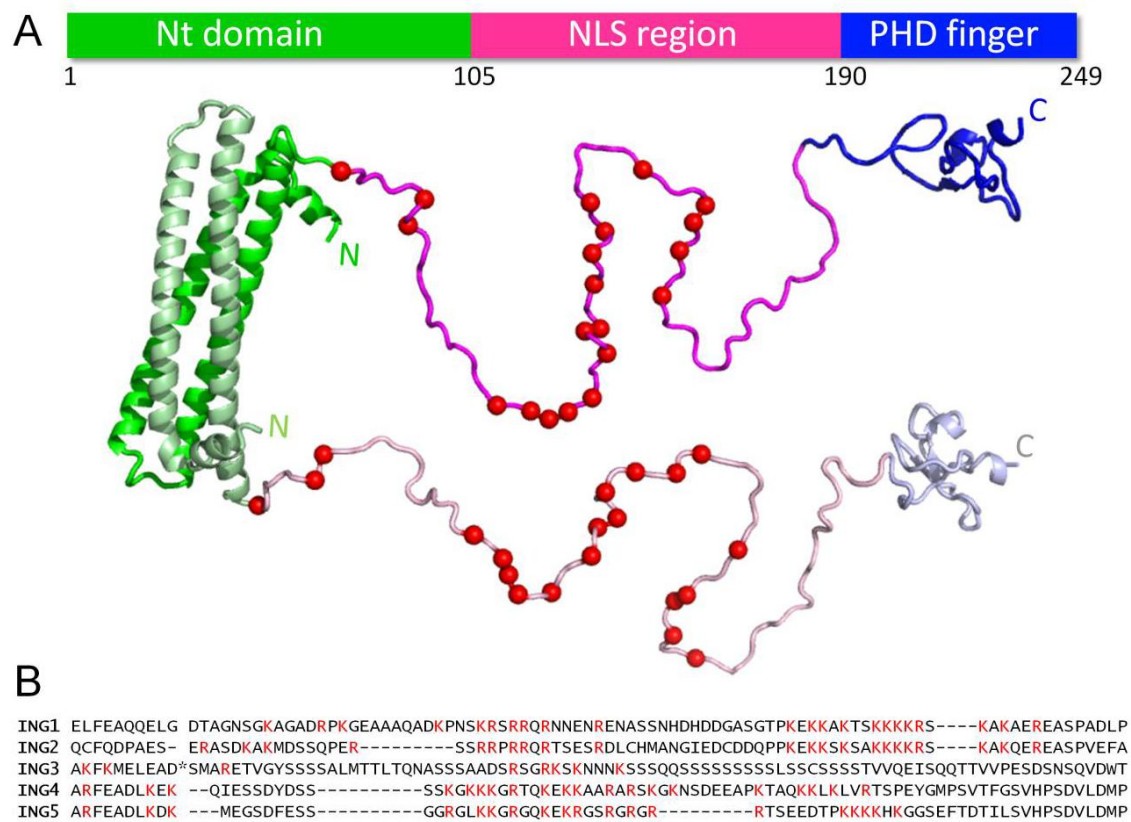


Figure 7. The NLS region of ING4 and other ING proteins. (A) Schematic representation of the domain structure of human ING4 dimer: in green the N-terminal coiled-coil domain responsible for the antiparallel dimerization, in blue the PHD domain responsible for binding to the histone H3 tail, and in magenta the connecting region containing the NLS. Below the scheme are the crystal structures of a dimer of the N-terminal domain [103] (PDB ID: 4AFL) and the PHD finger [85] (PDB ID: 2VNF). The NLS region was randomly built to connect the two domains, and the positions of lysine and arginine residues are indicated by red spheres. One of the two ING4 protomers is represented with paler colors than the other one. (B) Sequence alignment of the NLS regions of the ING proteins. The alignment was done with CLUSTAL W using the longest isoforms of the proteins (except ING1, for which p33ING1b isoform was used). The sequences shown correspond to the limits of ING4 NLS region as defined in A (starting at residue 106). The asterisk in the sequence of ING3 indicates that a long insertion (residues 112-266) is omitted for the sake of clarity.

ING4 can be described as an elongated dimer with two PHD fingers pointing to opposite directions and tethered by a disordered central region approximately 85 residues long (Figure 7A) [102, 103]. This central region is rich in basic amino acids (Figure 7B) and contains the nuclear localization signal, and is commonly named the NLS region. Previously in the lab, it was found that human ING4 produced in bacterial cells copurifies with DNA (as seen by ultraviolet absorbance and native PAGE stained for protein or DNA), pointing to a stable interaction with DNA. This observation prompted us to identify the ING4 domains involved in the interaction as well as the binding determinants regarding structure and length of the DNA molecule [193] (Appendix IV).

ING4 was shown to bind weakly to an 18 bp DNA duplex (dsDNA18) and the binding was enhanced when the longer dsDNA32 substrate was used (Figure 8). We could not detect any binding for the isolated N-terminal or PHD domains so these results point to

the central region as the binding site for the DNA. This was confirmed by the deletion mutant ING4 Δ NLS not binding to any of the two dsDNAs (**Figure 8**). This mutant lacks residues 106-187, which encompass most of the central NLS region connecting the two folded domains of ING4. Biophysical characterization of the mutant shows a symmetric dimer with predominantly helical coiled-coil structure and C-terminal PHD fingers, as seen by SEC-MALS and CD (**Figure 9**).

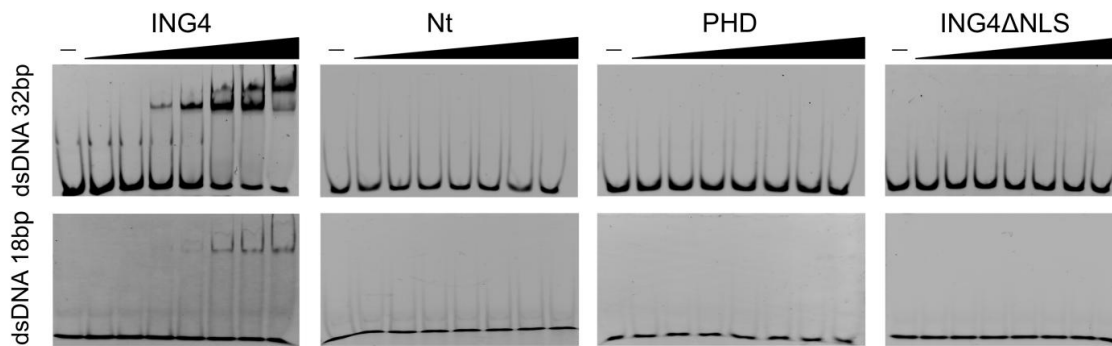


Figure 8. Titration of dsDNA32 (top) and dsDNA18 (bottom) with increasing amounts of ING4 proteins and monitoring complex formation by EMSA. The concentration of DNA was 0.1 μ M and the protein concentration was (lanes from left to right) 0, 0.05, 0.1, 0.25, 0.5, 0.75, 1 and 2 μ M.

An avidity effect was expected since ING4 is a dimer with two DNA binding sites and the long dsDNA32 can reach both sites independently in a way that binding to the first site will increase the likelihood for the second interaction to occur. The dimeric nature of ING4 makes it also possible that the long dsDNA32 binds more than one ING4 dimer, giving rise to multiple modalities of binding seen at high protein concentrations in **Figure 8**.

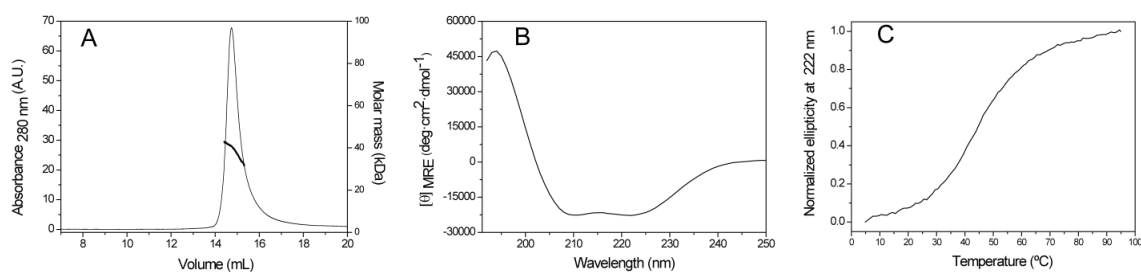


Figure 9. Biophysical characterization of mutant ING4 Δ NLS. (A) SEC-MALS analysis of the molecule. The molar mass at the centre of the chromatography peak is 40.7 kDa. The value calculated from the amino acid composition of the monomer is 19.5 kDa, indicating that the protein forms a dimer. (B) Secondary structure by far-UV circular dichroism. The ratio of the ellipticity at 222 and 208 nm is larger than 1, suggesting that it contains a large proportion of coiled coil structure. (C) Tertiary structure analysis by thermal denaturation followed by the change in the CD signal at 222 nm, showing a midpoint melting temperature of 45 $^{\circ}$ C.

As an approach for quantifying and interpreting the affinity of the interaction we fitted the data to the Hill equation [191] and calculated an apparent dissociation constant of $K_D = 0.6 \pm 0.03 \mu\text{M}$ (**Figure 10**). ING4 also binds to single stranded DNA molecules (**Figure 11**), but more weakly than it binds the corresponding duplexes (**Figure 8**). ING4 also binds to a primed DNA (**Figure 11**), but not as strongly as to dsDNA32. All these results indicate that ING4 has a preference for double strand DNA binding.

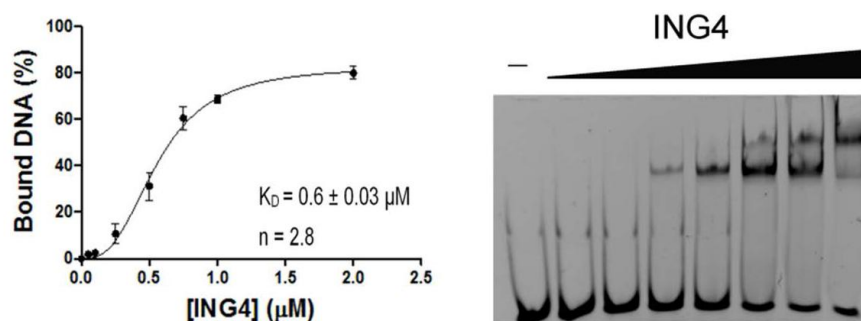


Figure 10. Quantitative analysis by gel densitometry of the binding of ING4 to fluorescent dsDNA32 measured by EMSA. The symbols and error bars are the mean and standard deviations of three independent experiments. The gel shown on the right corresponds to one of these three experiments. The curves are the fitting to a Hill equation corresponding to a single-site binding model as implemented in Prism (GraphPad software). The apparent equilibrium dissociation constant (K_D) and cooperativity (n) are indicated.

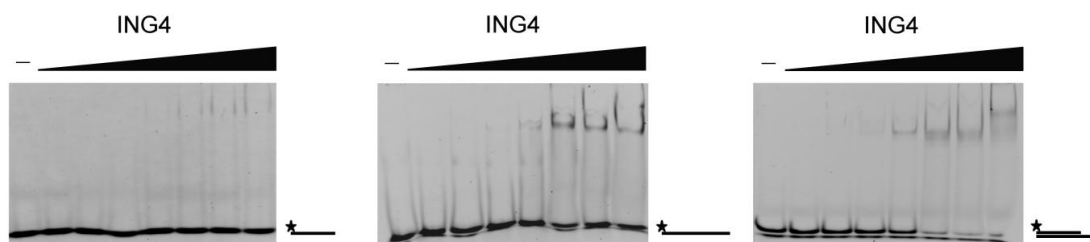


Figure 11. Titration of fluorescent DNA molecules with increasing amounts of ING4 and monitoring complex formation by EMSA. From left to right: 20 base-long ssDNA, 34 base-long ssDNA, and the corresponding primed DNA. The concentration of DNA was $0.1 \mu\text{M}$ in the three experiments, and the protein concentration was (lanes from left to right) 0, 0.05, 0.1, 0.25, 0.5, 0.75, 1 and $2 \mu\text{M}$. A scheme of the structure of the DNA probe used is drawn at the right hand side of the gels, with the asterisk indicating the position of the fluorescent tag.

We used NMR to observe the interaction of ING4 with the dsDNA32 directly in solution (**Figure 12**) as a complementary experiment to EMSA. We can use this technique to do so because, although ING4 has a large size (it is a 55 kDa dimer), most of the signals observed at $25 \text{ }^\circ\text{C}$ belong either to the small PHD finger, which has been specifically assigned [101], or to the central disordered region. The size and shape of the dimeric N-terminal domain make its NMR signals mostly unobservable unless the concentration and the temperature are increased [102], so this will help in identifying the signals to residues of the NLS region. In the experiment we measured the chemical shift perturbations (CSP) caused on ING4 signals upon binding to increasing concentrations of

the dsDNA32 (**Figure 12A**). Along the titration, the backbone amide signals that experience CSPs are in the central region of the proton frequency dimension (**Figure 12A**).

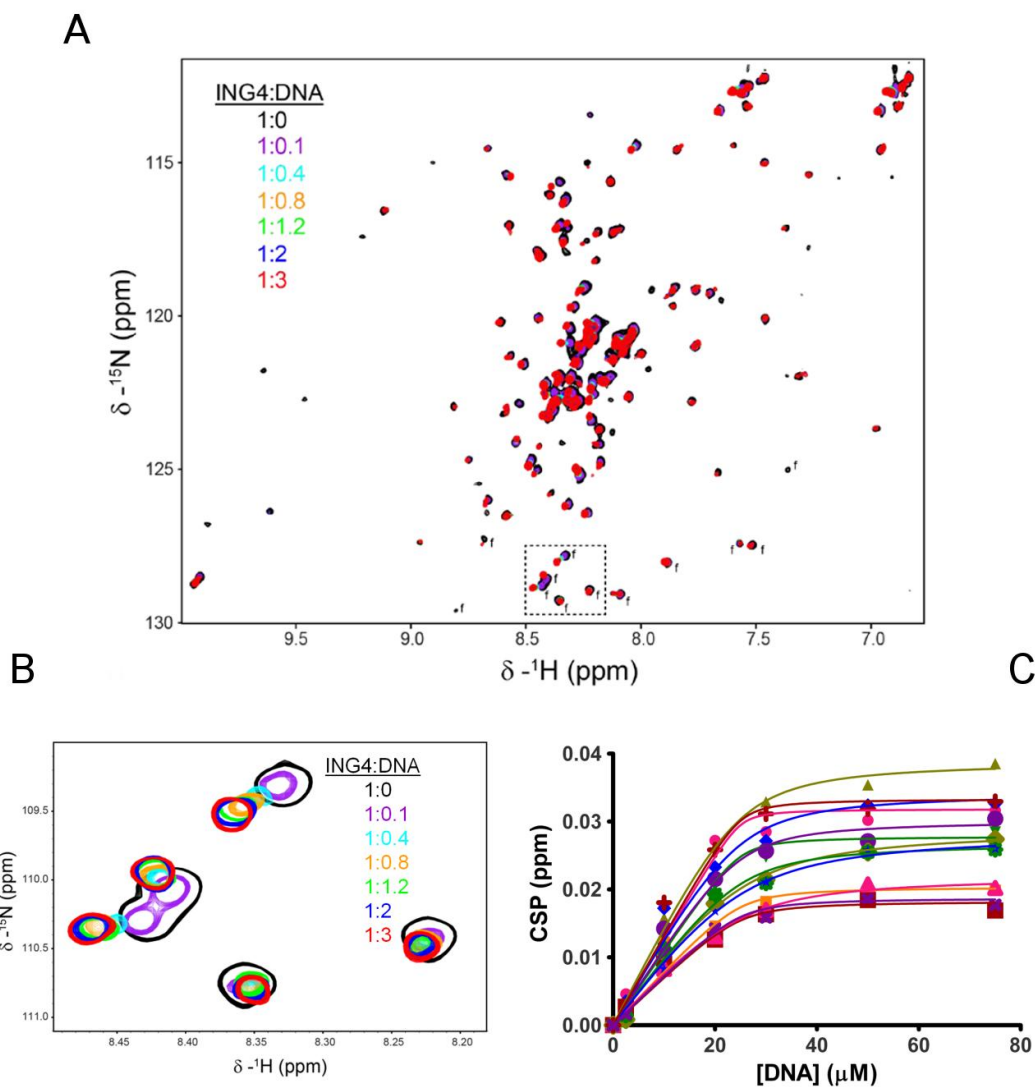


Figure 12. NMR analysis of the ING4-dsDNA32 interaction. (A) Overlay of NMR spectra of ING4 in the presence of increasing amounts of dsDNA32 as indicated by the color code (molar ratio ING4 protomer:dsDNA32). The concentration of the sample was 25 μM in 20 mM MES pH 6.5, 100 mM NaCl, 5 mM MgCl_2 , 1 mM DTT. Spectra were recorded at 25 $^\circ\text{C}$. The signals labeled with “f” are folded in the ^{15}N dimension and their real chemical shift is 18.5 ppm smaller. The dashed rectangle indicates the region expanded in panel B. (B) Zoom of the overlaid NMR spectra of ING4 shown in panel A, which contains the backbone amide signals of 5 residues tentatively assigned to the 5 glycine residues of the NLS region, three of them experiencing large CSP in the presence of saturating amounts of dsDNA32. (C) Plot of the CSP of the thirteen ING4 backbone amide signals with values larger than 0.015 ppm. The lines correspond to the fittings to a one site binding model, and yield dissociation constants in the 0.6-3.8 μM range.

It is important to mention that in this region of the NMR spectrum of ING4 there are 5 signals that are absent in the spectrum of ING4 Δ NLS and that have ^{15}N chemical shifts typical of glycine residues (**Figure 13**). We tentatively assigned these signals to 5 glycine residues that are present in the NLS region of ING4 (**Figure 7B**). Three of them experience relatively large shifts and the other two shift very little (**Figure 12B**). This is consistent with three of the NLS glycines being located close to lysine and arginine residue clusters in the central region of the NLS and two of the glycines being in the C-terminal end of the NLS, with no positively charged residues in their vicinity (**Figure 7B**).

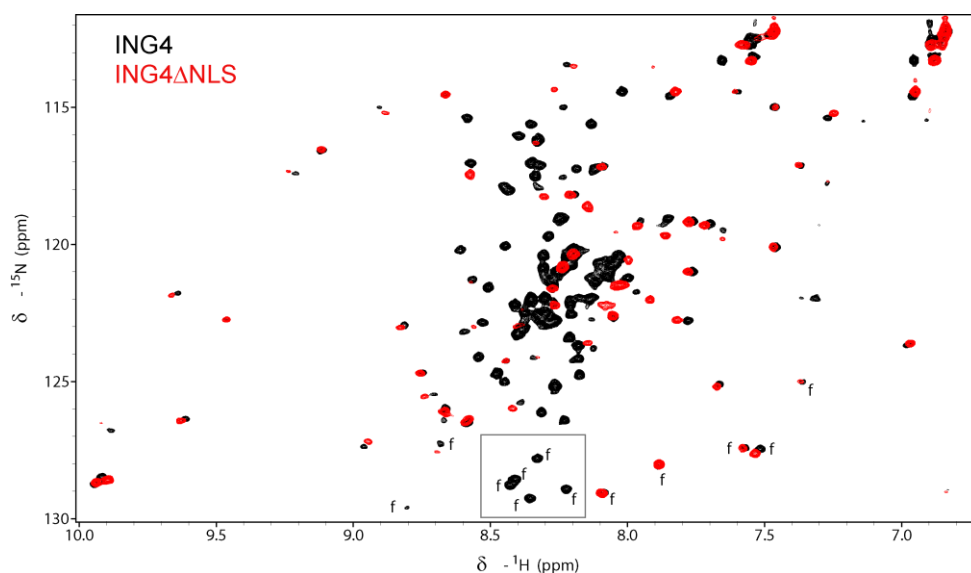


Figure 13. NMR spectrum of ING4 Δ NLS. Superposition of ^1H - ^{15}N HSQC spectra of ING4 (in MES pH 6.5) and ING4 Δ NLS (in PBS pH 7.0) at 25 °C showing dispersed (PHD finger) and non-dispersed (other parts of the protein) signals in the proton dimension. Framed signals correspond to the 5 glycine residues present in the NLS region of ING4 that are absent in ING4 Δ NLS. The signals labelled with “f” are folded in the ^{15}N dimension and their real chemical shift is 18.5 ppm smaller.

For ratios above 1:2 (ING4 protomer:dsDNA32) the calculated CSP reached a plateau (**Figure 12C**). Since the stoichiometry of the complex is unknown, data from 13 signals experiencing the largest changes upon de addition of dsDNA32 was fitted to a simple one site binding model (one DNA duplex bound to one ING4 protomer) and apparent dissociation constants were calculated to be in the 0.6 – 3.8 μM range. These numbers are only indicative of a low micromolar apparent affinity, similar to the value derived from the EMSA for the same DNA molecule. Even at high ratios the CSP are small indicating that in the DNA-bound form the backbone of the NLS region remains, at least partially, disordered. This observation is consistent with a weak binding based on electrostatic interactions, mediated by the positively charged side chains of the NLS and the phosphate backbone of the DNA, as has been reported for other disordered proteins rich in positively charged residues [194].

Our results indicate that the interaction is electrostatic, favoring long duplexes and not particular nucleotide sequences. Experimental data with dsDNA substrates with different sequence and lengths is in agreement with this conclusion. A 22 bp duplex with an unrelated sequence binds ING4 but more weakly than dsDNA32 (**Figure 14**), and a short 10 bp duplex still binds to ING4 but much more weakly (**Figure 15**). Therefore, the major determinant of DNA binding is not the nucleotide sequence but the length of the duplex; although we cannot exclude that there may be certain sequence preferences.

Our findings on ING4 [193] can be probably extrapolated to ING5 and other ING proteins, because an alignment of their NLS regions shows several clusters of positively charged residues (**Figure 7B**), perhaps with the exception of ING3 whose central region has much less positively charged residues than the others [68, 72].

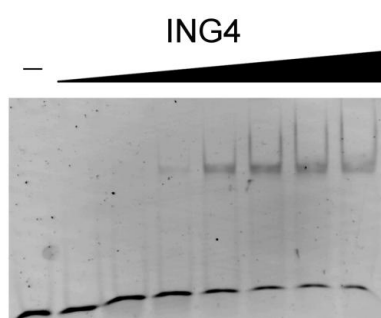


Figure 14. Titration of dsDNA22 with increasing amounts of ING4 protein and monitoring complex formation by EMSA. The concentration of DNA was 0.1 μM and the protein concentration was (lanes from left to right) 0, 0.05, 0.1, 0.25, 0.5, 0.75, 1 and 2 μM .

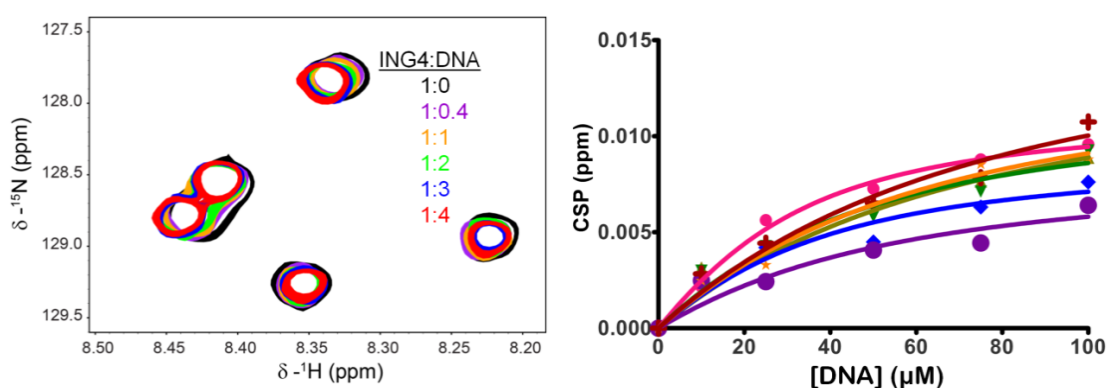


Figure 15. NMR analysis of the ING4-dsDNA10 interaction. (Left) Zoom of the overlaid NMR spectra of ING4 in the presence of the indicated molar ratios of dsDNA10, with the same backbone amide signals shown in figure 12B. (Right) Plot of the CSP of the six ING4 backbone amide signals with values larger than 0.005 ppm. The lines correspond to the fittings to a one site binding model, and yield dissociation constants in the 17-58 μM range.

ING5 binds dsDNA32 substrate as seen by EMSA (**Figure 16**); however, the binding is weaker than in the case of ING4 (**Figure 8**). The weak binding makes it difficult to quantify and interpret the affinity of the interaction from EMSA experiments. An attempt to do so using a Hill equation [191] yields an apparent dissociation constant of $1.18 \pm 0.37 \mu\text{M}$ (**Figure 16**) that is in the same micromolar range as the calculated for ING4 with the same substrate. We could not detect any binding for the isolated N-terminal or PHD domains so these results point to the central region of ING5 as the binding site for the DNA as it happens for ING4 [193]. ING5 also binds to a single stranded DNA molecule 32 nucleotides long (**Figure 16**, Bottom panel), but even more weakly than it binds the corresponding duplexes.

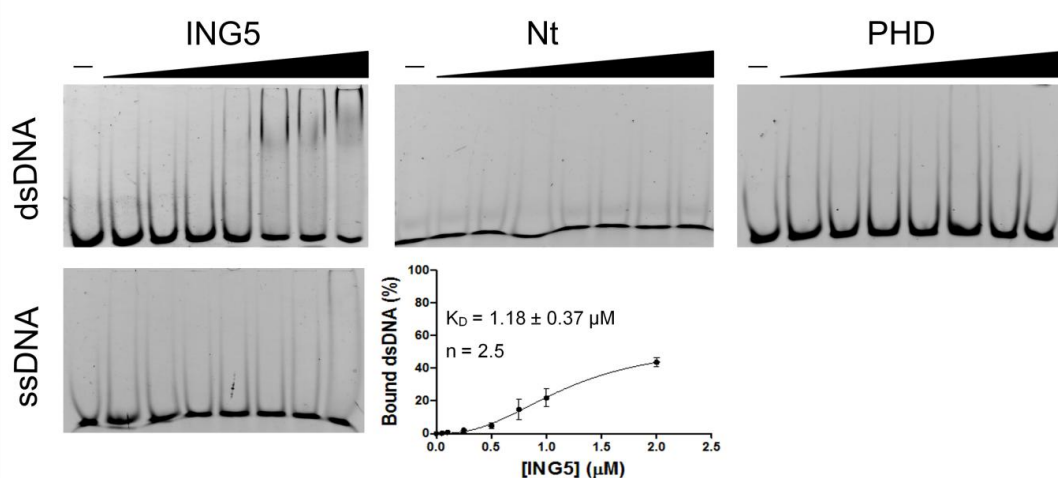


Figure 16. EMSA analysis of the ING5-DNA interaction. Top panels: titration of dsDNA32 with increasing amounts of ING5 proteins and monitoring complex formation by EMSA. The concentration of DNA was $0.1 \mu\text{M}$ and the protein concentration was (lanes from left to right) 0, 0.05, 0.1, 0.25, 0.5, 0.75, 1 and $2 \mu\text{M}$. Bottom left panel: titration of ssDNA32 with increasing amounts of ING5 (same concentration as in the top panels). Bottom right panel: quantitative analysis by gel densitometry of the binding of ING5 to fluorescent dsDNA32 measured by EMSA. The symbols and error bars are the mean and standard deviations of three independent experiments.

4. The Nt domain of ING5 forms a symmetric coiled-coil dimer

The N-t 1-105 domain of ING5 (hereafter named Nt₁₋₁₀₅ or Nt domain) presents a helical pattern in the circular dichroism spectrum that is more pronounced than that of the full length protein (**Figure 1B**), with an ellipticity ratio $[\Theta]_{222\text{nm}}/[\Theta]_{208\text{nm}} > 1$, typical of coiled-coil structures [195]. This secondary and tertiary structure would be consistent with both the asymmetric crystal structure [121] and with the symmetric model based on the homologous ING4 Nt [103].

4.1. Nt domain backbone NMR assignment

In the NMR spectrum of ING5 most of the signals corresponding to the N-t domain can't be observed due to the low solubility of the full length protein. However, the higher solubility of the isolated Nt domain allows the observation of 93 backbone amide signals in the ^1H - ^{15}N correlation spectrum, which could be assigned to specific residues (**Figure 17A**). Although there are 10 residues for which no amide signal was detected (excluding the two extra residues at the N-terminus), the observation of a single set of resonances (only one signal is observed for each residue) indicates that the dimer is symmetric. This result is inconsistent with the crystal structure of ING5 Nt because an asymmetric dimer would result in two sets of signals for each residue (due to their different chemical environment).

Because the crystal of the ING5 Nt domain was obtained at pH 6.5 (lower than the NMR experiments, recorded at pH 7.3) we examined the spectra of the protein behaviour of the protein at different pH values to confirm that no major conformational changes occur at pH 6.5. As can be seen in **Figure 18**, the spectra in the range pH 7.9-6.6 are essentially the same (**Figure 17A**). At lower pH values the protein precipitated and the quality of the spectrum is very poor. However, the absence of two sets of signals at pH 6.6 indicates that there is no asymmetric conformation in solution at this pH.

The symmetric nature of the dimer in solution is, however, consistent with a structure similar to that of the N-terminal domain of ING4, which forms a symmetric dimer with an antiparallel coiled coil fold based on two long helices named $\alpha 2$ (17-52) and $\alpha 3$ (58-103). Chemical shift analysis of the ING5 Nt residues shows two long helices spanning residues 18-54 and 58-100 (**Figure 17B**). Residues at the N-terminus (1-5) and at the C-terminus (100-105) show systematically small heteronuclear $\{^1\text{H}\}$ - ^{15}N NOE values (< 0.62) indicating flexibility on ps to ns times scales (**Figure 17C**). The N-terminal 6 residues of ING5 Nt are flexible and disordered as indicated also by their chemical shifts (**Figure 17 B**), and for several of the following residues the NMR signal could not be observed. Therefore, the presence of the short N-terminal $\alpha 1$ helix between residues 0 and 11 seen in the crystal structure of ING5 Nt could not be confirmed by NMR. The NMR data on ING5 suggests that this segment may be involved in conformational exchange equilibria that cause some of the NMR signals to become non-detectable. The determination of the 3D structure by NMR was precluded (or would be very difficult to achieve) by the low sensitivity of the NMR experiments (because of its size, helical structure and elongated shape). Therefore we tried to confirm by other strategies which was the best representative model of the structure of ING5 Nt in solution.

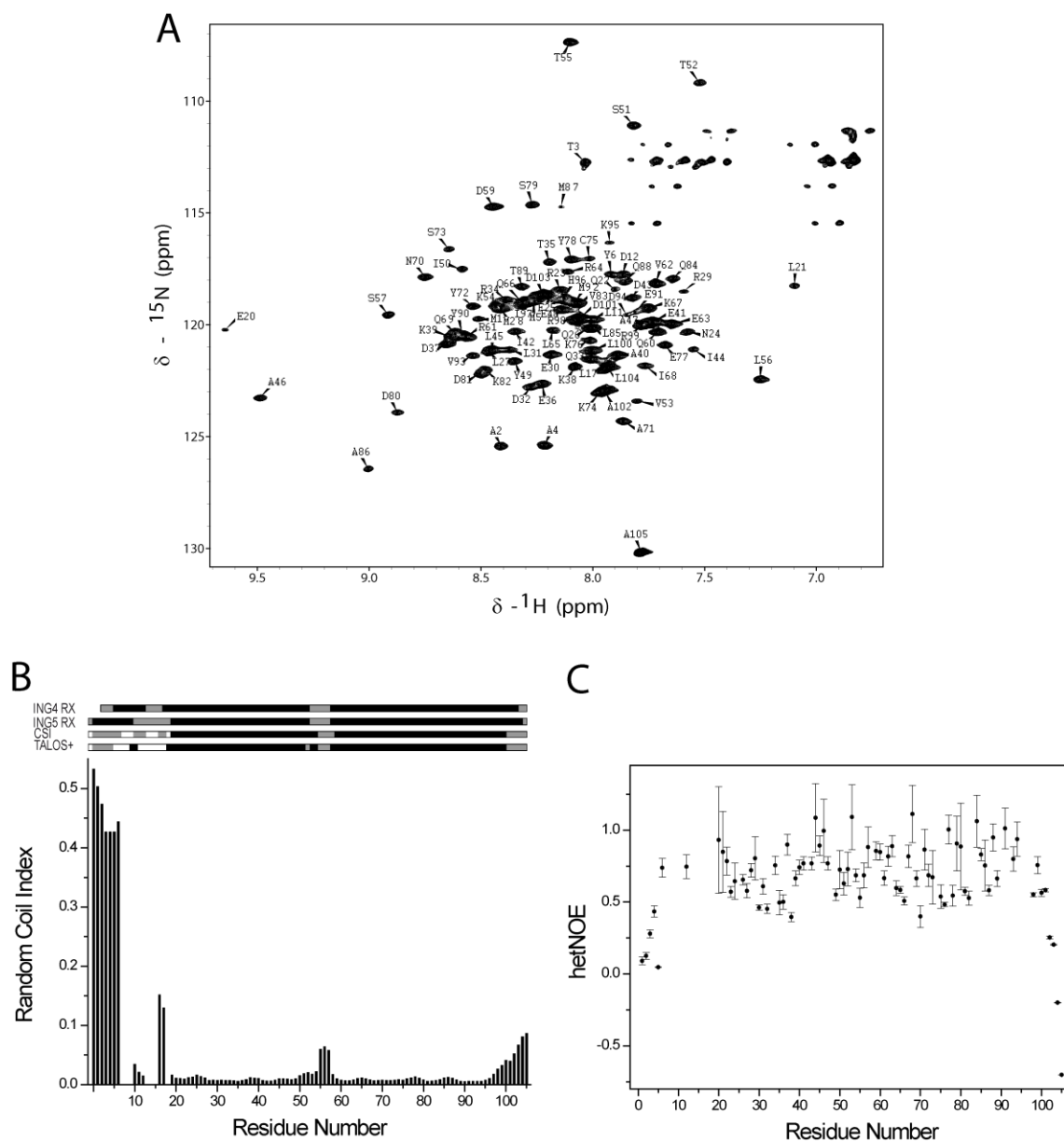


Figure 17. Structural analysis of ING5 Nt in solution by NMR. (A) ^1H - ^{15}N -TROSY NMR spectrum of ING5 Nt at 25 °C and 364 μM protein concentration in 300 μL of 50 mM MOPS pH 7.3, 5% D₂O, 3 μM DSS, 0.01% NaN₃ and 0.5 mg of protease inhibitors. (B) Secondary structure analysis of ING5 Nt from backbone resonances chemical shift analysis and comparison with the crystal structure of both ING5 and ING4 Nt (see below). The black boxes indicate helical regions; the grey boxes coil regions and white boxes residues for which there is no NMR data. The calculated RCI is represented as a red line and the calculated heteronuclear NOE as black dots, with experimental error bars derived from the noise level in the two spectra used to measure the NOE.

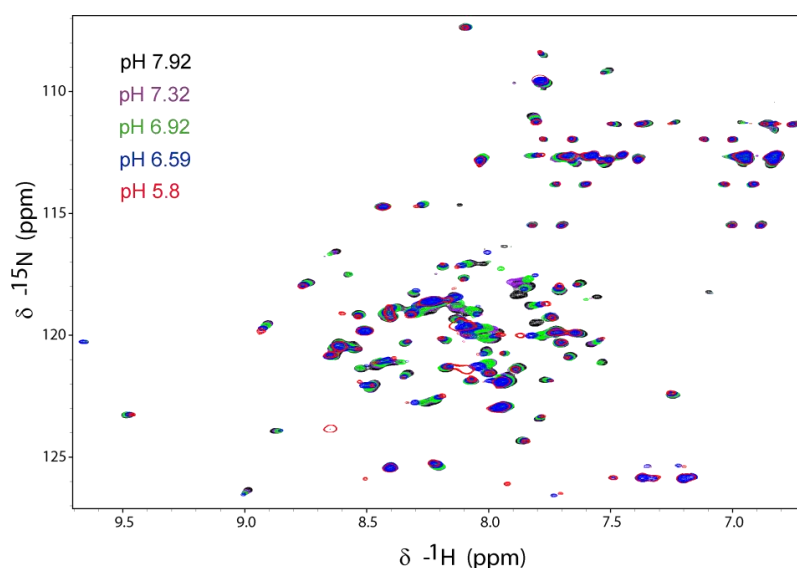
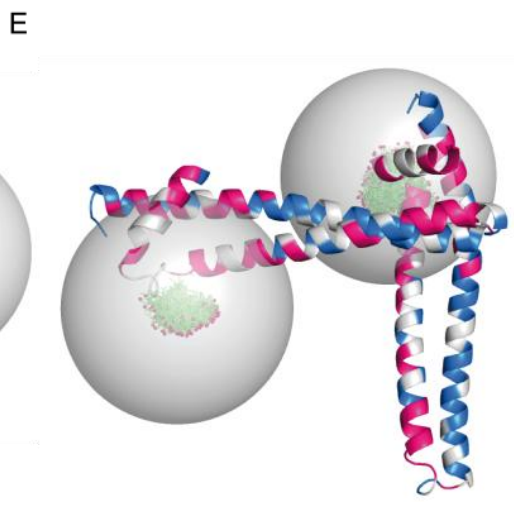
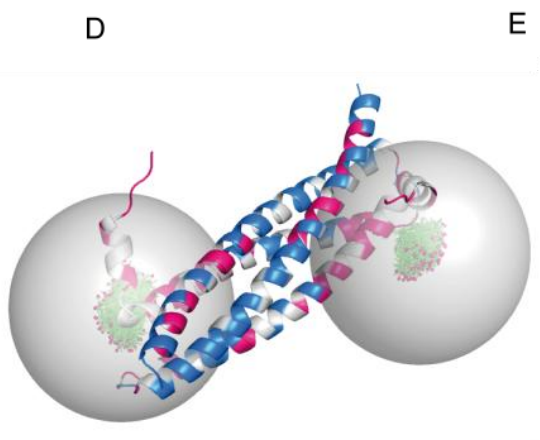
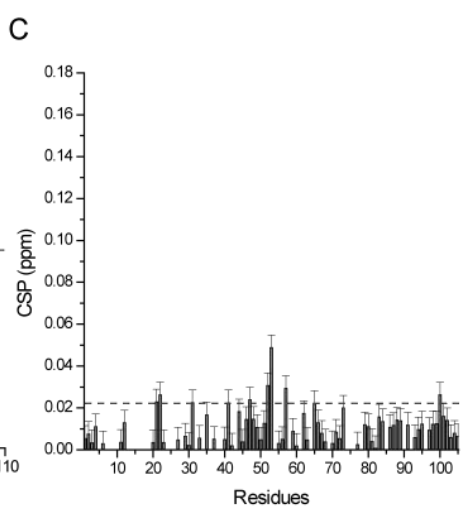
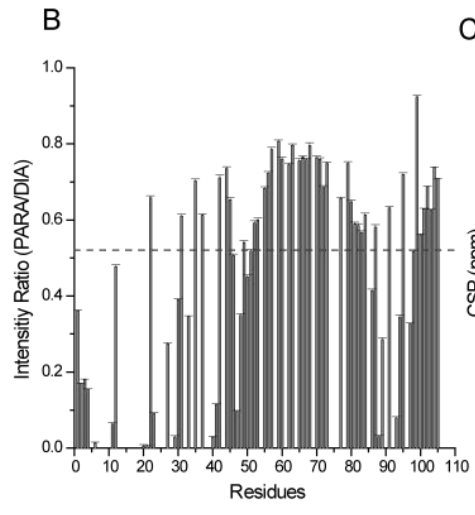
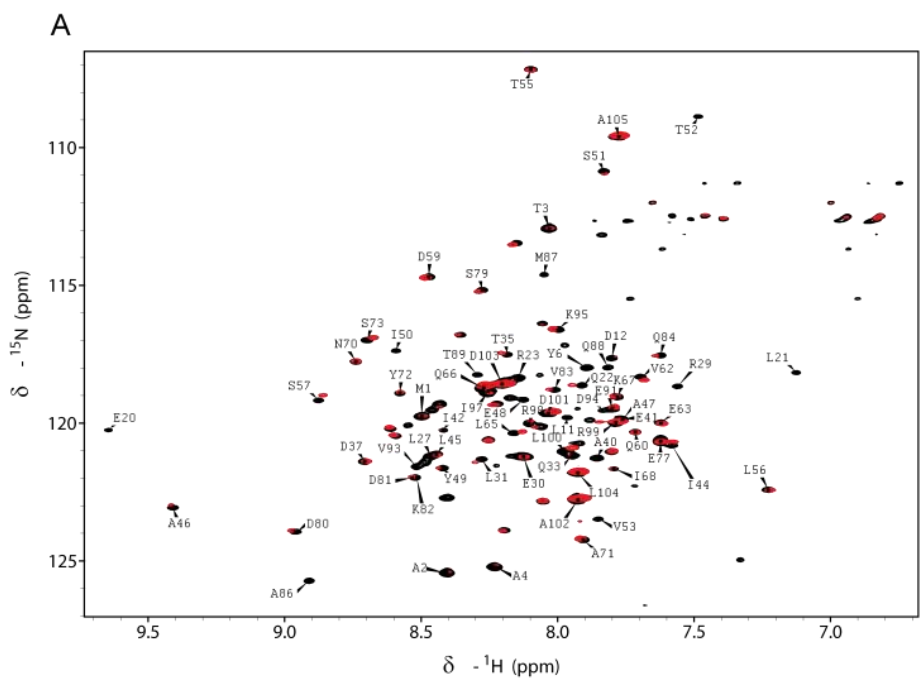


Figure 18. Spectra of ING5 Nt at different pH values. Superposition of five ^1H - ^{15}N -TROSY spectra of ^{15}N ING5 Nt at 180 μM in 400 μL 50 mM MOPS at the indicated pH values, with, 1 mM DTT, 5% D $_2\text{O}$, 3 μM DSS and 0.5 mg of protease inhibitors.

4.2. Paramagnetic relaxation effects

The Nt domain of ING5 possesses two cysteine residues (C19 and C75) that we could label with the paramagnetic tag MTSL to confirm the structural organization of the ING5 dimer in solution by means of NMR paramagnetic relaxation effects (PREs). One of them (C19) is solvent exposed in both the crystal structure (asymmetric dimer) and the model based on the ING4 Nt structure (symmetric dimer). Therefore this is an adequate site to introduce a paramagnetic tag like MTSL and measure paramagnetic relaxation enhancements. The cysteine C75 is, however, partially buried (asymmetric model) or buried (symmetric model), and is not a suitable site for introducing the MTSL probe. Therefore, we generated the C75S mutant and labeled the only remaining cysteine residue (C19) with MTSL. The analysis of these experiments is shown in **Figure 19**. The pattern of signals in the ^1H - ^{15}N -TROSY spectrum in the C75S mutant (**Figure 19A**) is very similar to the wild type and most of the assignments could be transferred by comparison of the two spectra. In the C19-MTSL labeled sample many signals experience large changes in intensity and/or chemical shift perturbation. The ratio of the intensities measured in the paramagnetic (MTSL labeled) and diamagnetic (non-labeled) samples are plotted in **Figure 19B**. The 40% of the residues whose intensities are reduced the most (to less than 53 %) together with the residues showing the largest CSP (larger than the average plus two standard deviations, **Figure 19C**) are mapped on the structure of ING5 Nt in **Figure 19D and 19E**). The pattern of residues whose NMR signals change the most in the MTSL-labeled sample is more consistent with a symmetric (**Figure 19D**) dimer than with an asymmetric one (**Figure 19E**).



4.3. Structural analysis of ING5 Nt three mutants designed to clarify the structural nature of the domain.

The crystallographic structure of the ING5 Nt domain [121] showed a dimer in which each protomer is folded into a helix-loop-helix structure with two long helices plus a short $\alpha 1$ helix at the N-terminus. This secondary structure is similar to that of the crystal structure of ING4 Nt. However, the crystal structure the ING5 Nt domain showed a completely different dimerization interface, being an asymmetric dimer while ING4 forms a symmetric dimer (**Figure 20A**). There is a clear asymmetry between protomers in the dimer and a “closed” and an “opened” conformations can be observed for each one depending on the relative position of helix $\alpha 1$ with respect to the other two (**Figure 20B**). This asymmetry in the ING5 Nt crystal structure is inconsistent with the obtained NMR results and previous results on ING4 N-terminal crystal structure [103].

One possible explanation for the inconsistency is that the presence of two extra residues at the N-terminus favors interactions that, in the crystal conduct to the formation of a dimer of asymmetric dimers (**Figure 20C**). In this arrangement the $\alpha 1$ helices of one dimer pack against the $\alpha 1$ helices of the other dimer, with several polar and apolar interactions stabilizing the assembly, as shown in **Figure 20D**. Indeed, the amine group of the extra residue G-1 of the open protomer of one of the dimers forms a salt bridge with the side chain of residue E8 in the helix $\alpha 1$ in the closed protomer of the other dimer.

At the same time, E8 forms an H-bond with the backbone amide of residue M1 of helix $\alpha 1$ in the open protomer. The E82 residue acts as an N-cap of the first turn of $\alpha 1$ [196]. The same thing happens with residue D101 in $\alpha 3$ of the open protomer, which forms a salt bridge with the amine group of G-1 in the closed protomer of the other dimer and H-bonds with the backbone amides of M1 and A2 of this same closed protomer (**Figure 20E**). This arrangement of a dimer of dimers in the crystal is predicted by the PISA server to be an assembly in solution even more stable than the dimer. However in solution we have determined by SEC-MALS that the protein is a dimer (**Figure 1**).

Figure 19. Paramagnetic relaxation effects on ING5 Nt C75S labeled with MTSL at residue C19. (A) Superposition of ^1H - ^{15}N TROSY spectra of paramagnetic (MTSL-labeled, colored red) and diamagnetic (non-labeled, **black**) ^{15}N ING5 Nt. (B) Intensity analysis of the signals measured in the spectra shown in A. The percentile 40 (separating 40 % of the most perturbed residues) is represented by the dashed line. (C) CSP measured for each residue in the presence of MTSL label. The value for the average plus 2 standard deviations is represented by de dashed line. (D) The ribbon of the structure modeled on the crystal structure of the ING4 homolog or (E) modeled on the crystal structure of the WT is shown in blue. Those residues experiencing the largest PREs or CSP values are in magenta, non perturbed residues in blue, and residues for which no data is available are in white. Ninety possible conformations of the MTSL spin label are represented by green lines (except the oxygen, as a small red sphere) and were calculated with the Pymol plugin MtsslWizard [197]. Large gray translucent spheres with a 10 Å radius, and centered at the cysteine sulfur atom, are shown as a visual guide for the proximity to MTSL.

Another possibility is that the close contact of C19 side chain of one ING5 molecule and the corresponding C19 side chain of a symmetry related molecule stabilizes the asymmetric dimer. In fact, the refinement of the structure provided two alternative orientations for C19 (with 70-30% occupancy) in the major one with the geometry of an intermolecular disulfide bridge (**Figure 21A**). A third possibility could be that the “closed” and the “open” conformations are determined by different conformations of the flexible helix $\alpha 1$ resulting in both kind of protomers being selected in the crystal in the form of an asymmetric dimer. To investigate the source of the inconsistency between the crystal and solution results we examined three mutants: an N-terminal mutant that lacks the two extra residues coming from the cloning procedure (ING5₂₋₁₀₅), the point mutant C19S and a $\Delta\alpha 1$ mutant lacking part of the flexible Nt α -helix.

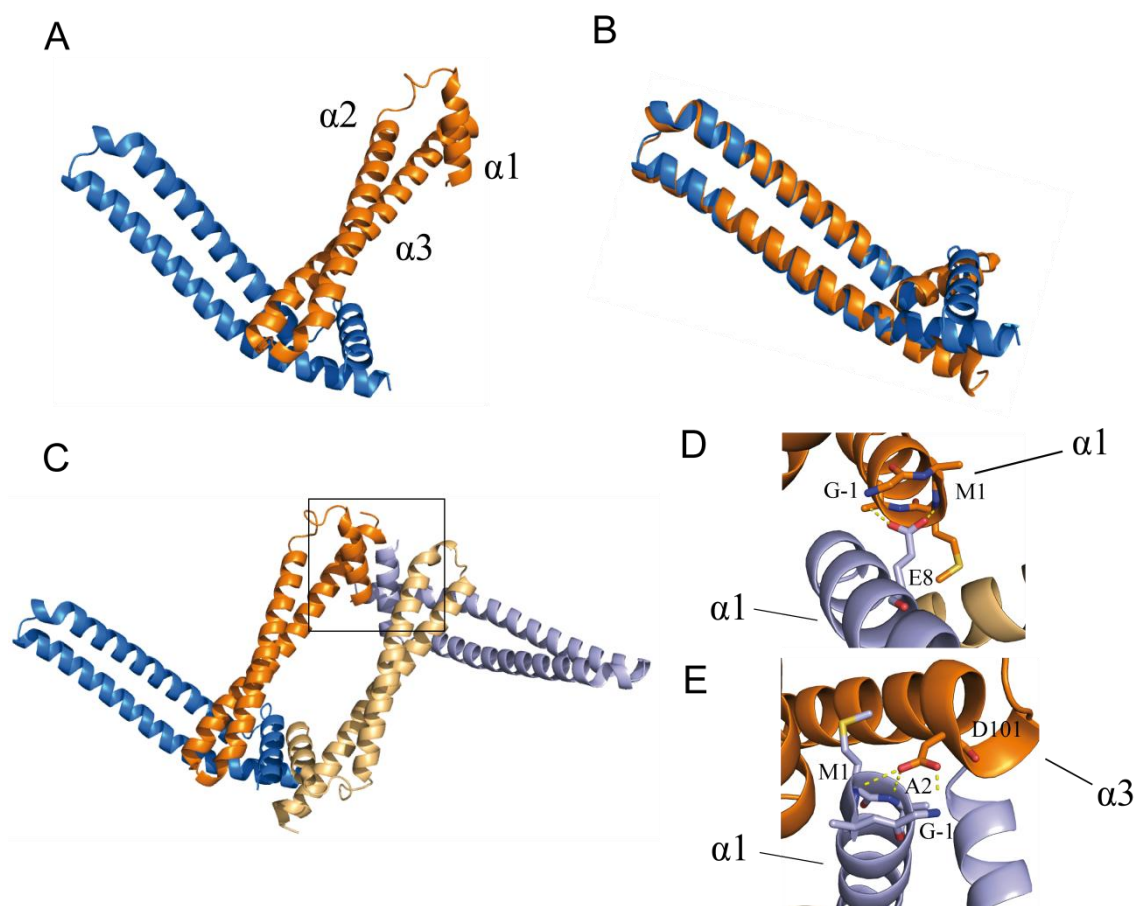


Figure 20. Crystal structure of ING5 Nt. (A) View of the ING5 Nt homodimer in the asymmetric unit of the crystal. The two protomers are shown as ribbons with orange (open) and blue (closed) color. The three helices from each protomer are indicated in the open protomer. (B) Superposition of the two protomers of ING5. (C) Ribbon diagram of the tetrameric assembly in the crystal of ING5 Nt identified by PISA analysis. The tetramer is formed by two asymmetric units, with the chains of the second unit indicated with paler colors. (D) Zoom of the framed area in C showing interactions between residues G-1 and M1 in helix $\alpha 1$ from an open protomer with residue E8 in the helix $\alpha 1$ of a closed protomer of another dimer in the tetrameric assembly. (E) Zoom of the framed area in C showing interactions between residues G-1, M1 and A2 in helix $\alpha 1$ from a closed protomer with residue D101 in the helix $\alpha 3$ of an open protomer of another dimer in the tetrameric assembly.

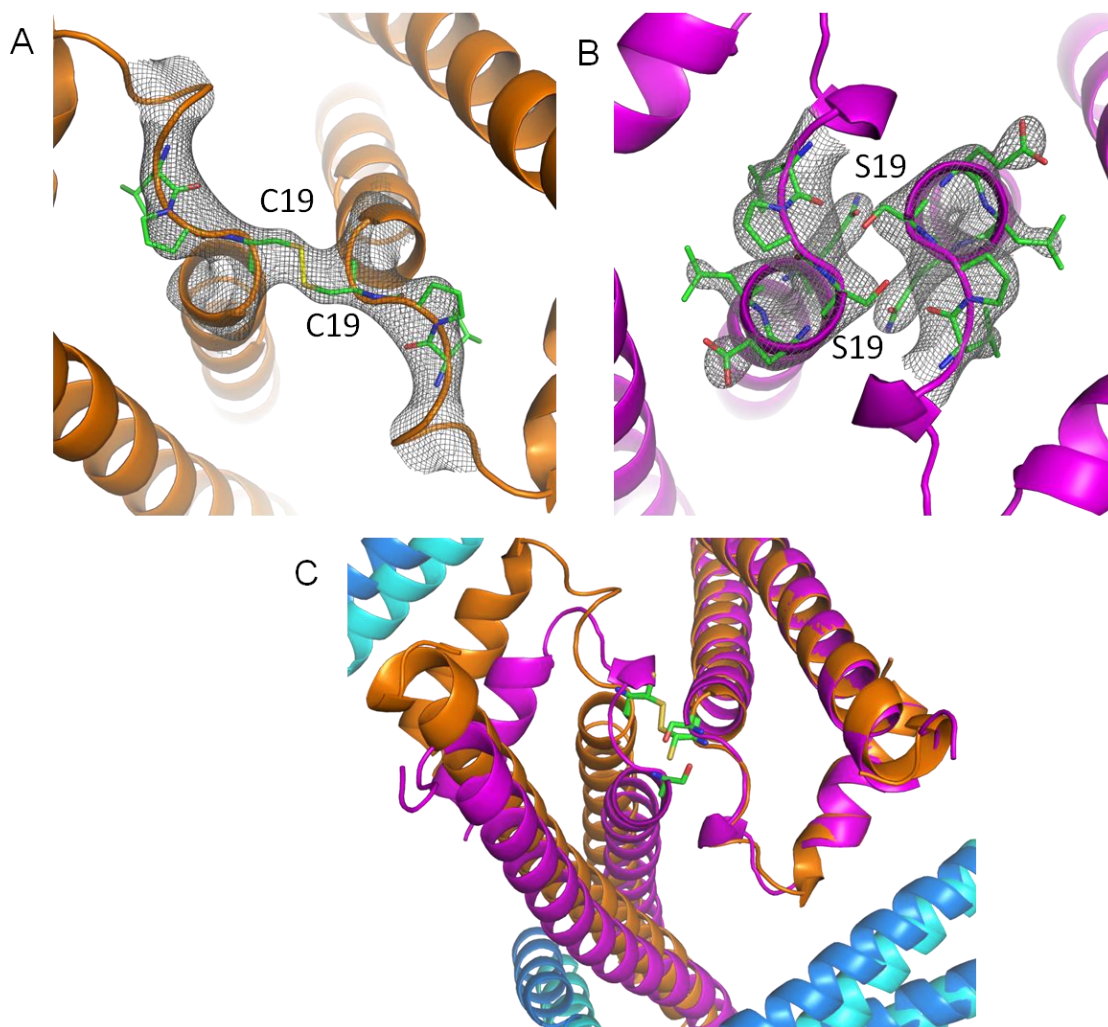


Figure 21. Close view of residue 19 side chain of ING5 Nt and C19S mutant. (A) Partial intermolecular disulfide bridge in ING5 Nt. Density map $2F_o-F_c$ (grey mesh drawn at 1.5σ). (B) The same region as in A corresponding to the C19S mutant. Density map $2F_o-F_c$ (grey mesh drawn at 1.5σ). (C) Superposition of the tetramers of the wild type and the C19S mutant (on one of the open protomers) showing that the assembly is not identical in the two crystals. The side chains of the two C19 involved in the disulfide bridge and the corresponding S19 side chains are shown in sticks. In the case of C19 the two alternative conformations are shown.

The three mutants formed dimers in solution with coiled coil structure, as seen by CD and SEC-MALS (**Figure 22**).

The NMR spectrum of the ING5₂₋₁₀₅ construct is very similar to that of Nt, with most of the differences at signals close to the N-terminal residues (**Figure 23**). This result demonstrates that in solution there is no difference in the structure of the two constructs, both adopting a symmetric coiled coil dimer. However, this protein did not crystallize, indicating that the presence of two extra residues in the N-terminus favored crystallization and, perhaps, the formation of an asymmetric dimer and the tetrameric assembly suggested by PISA in the crystal lattice.

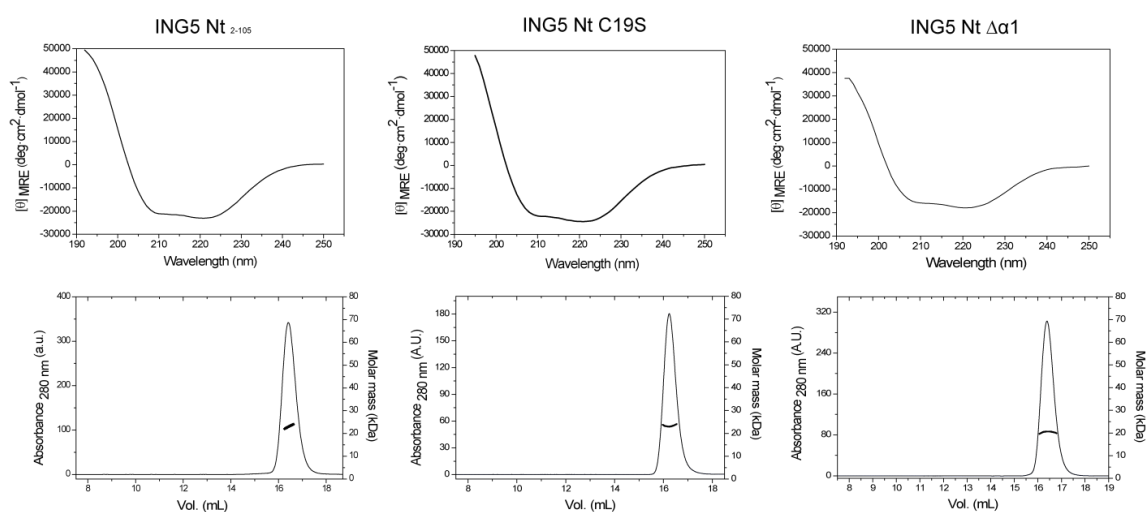


Figure 22. Structural analysis of three ING5 Nt mutants. CD (top panels) and SEC-MALS (bottom panels) analysis of ING5 Nt₂₋₁₀₅, Nt C19S, and $\Delta\alpha 1$ mutant. Data were measured at 25 °C in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT.

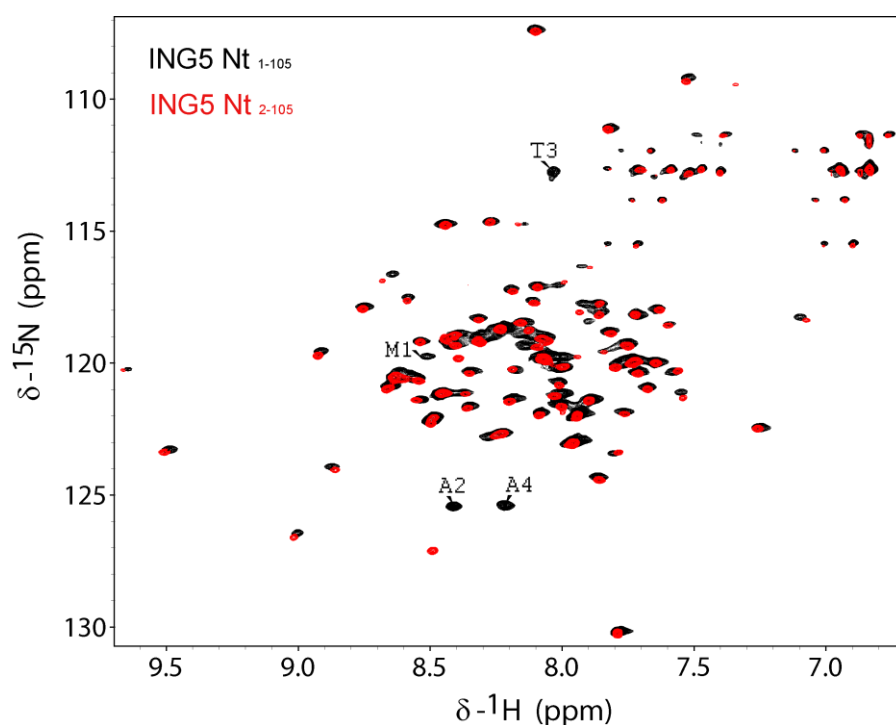


Figure 23. ^1H - ^{15}N -TROSY NMR spectrum of the ING5 Nt₂₋₁₀₅ (red) overlaid on the spectrum of ING5 Nt₁₋₁₀₅ (black). Both spectra were recorded at 25 °C in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT. For the sake of clarity only the N-terminal first four residues are labeled.

Crystallization trials with the $\Delta\alpha 1$ mutant yielded crystals (**Figure 6A** in the materials and methods section) that diffracted at a very low resolution (5.09 Å or lower) so they were not suitable for solving the structure of this mutant (**Figure 6B** in the materials and methods section).

Crystallization trials of ING5 Nt C19S yielded crystals (**Figure 5E-F** in the materials and methods section) that after optimization were suitable for X-ray diffraction and data collection at 3.1 Å resolution (**Figure 6A** in the materials and methods section). The structure was solved by molecular replacement using the crystal structure of ING5 Nt and refined to R_{work} of 21.5 % and R_{free} of 25.4 %. A summary of the crystallographic data

statistics is shown in **Table 1**. Although the C19S mutant formed crystals in slightly different conditions than Nt (pH 7.5 instead of 6.5), the structure is essentially the same as the wild type domain (**Figure 24**).

Table 1. Crystallographic data statistics of ING5 Nt C19S refinement.

ING5 Nt C19S	
Construct	ING5 _{C19S}
Data collection	
Space group	P6 ₅ 22
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	140.94, 140.94, 92.32
α , β , γ (°)	90.0, 90.0, 120.0
Resolution (Å)	92.32 (3.27-3.10)
<i>R</i> _{merge}	0.17 (0.93)
CC _{1/2}	0.98 (0.87)
<i>I</i> / σ (<i>I</i>)	9.7 (3.1)
Completeness (%)	100 (100)
Redundancy	18.3 (18.8)
Refinement	
No. reflections	10275
<i>R</i> _{work} / <i>R</i> _{free}	21.53/25.38
No. atoms	
Protein	1752
Ions/ligands	6*
Water	3
B-factors (Å ²)	
Protein	99.43
ions	81.22
Water	71.80
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.024
Ramachandran	
favoured (%)	96.0
Ramachandran	
outliers (%)	0.0
PDB code	5MTO

Values in parentheses correspond to the highest resolution shell.

* Sulfate and Na ions.

The asymmetric unit of the crystal is formed by two polypeptide chains that form an asymmetric homodimer (**Figure 24A**), together with several ions coming from the precipitant solution. The two protomers in the asymmetric unit form a dimer and correspond in the PDB file to A and B chains, respectively. Each of the two chains forms a helix-loop-helix structure with two long helices (residues 19-52 and 58-101, named $\alpha 2$ and $\alpha 3$, respectively) plus a short $\alpha 1$ helix at the N-terminus (residues 0-14, with residues -2 and -1 being an Gly and Ala preceding M1). The two helices of each protomer form an antiparallel coiled-coil structure, and both helices contain sequence segments with the typical heptad pattern $(abcdefg)_n$, of coiled coils. However, there is a clear asymmetry between the protomers, seen in the superposition based on helices $\alpha 2$ and $\alpha 3$ of both chains (**Figure 24C**). As it happens for the wild type ING5 Nt crystal structure, they differ in the bending of helix $\alpha 3$ and in the relative position of helix $\alpha 1$ with respect to the other two. In the dimer, the region of chain B (close protomer) where the short $\alpha 1$ is located

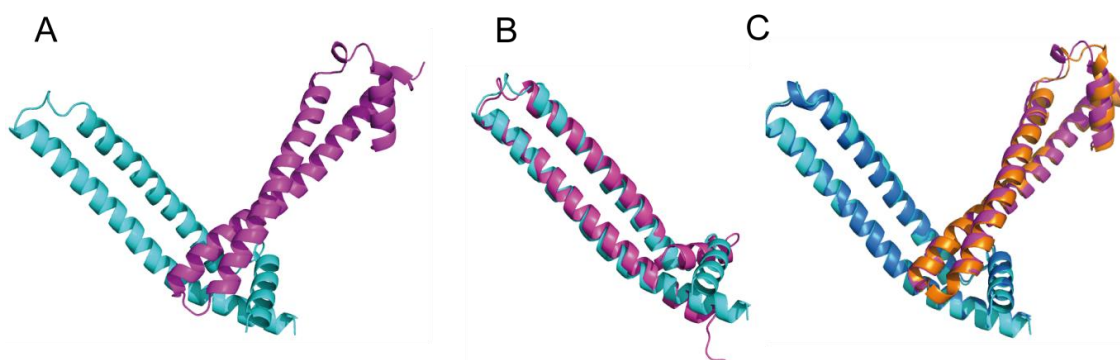


Figure 24. ING5 Nt C19S structure analysis. (A) Structure of the ING5 Nt C19S homodimer in the asymmetric unit of the crystal, with the two protomers shown as ribbons with magenta (open) and cyan (closed protomer) color. (B) Superposition of the two protomers of ING5 C19S. (C) Superposition of the ING5 Nt C19S dimer (magenta and cyan) on the corresponding dimer of the wild type (orange and blue for the open and closed protomers, respectively).

The two protomers share many intramolecular interactions, including hydrophobic, charged, and long range ones, like the interhelical salt bridge D32-K82 (**Figure 25B**). But there are also interactions formed only individually in each of the protomers, like the salt bridge H96-E20 and the H-bond N24-T89, which stabilizes de hydrophobic core of the open protomer (**Figure 25A**). The intermolecular packing at the dimerization interface involves the helix-loop-helix end of the open protomer and the opposite end of the close protomer, stabilized by H-bonds involving side chains on $\alpha 3$ of the open protomer and side chains on $\alpha 2$ and $\alpha 3$ of the closed protomer (**Figure 25B**). This mode of packing results in two different hydrophobic cores: the small three-helical hydrophobic core of the open protomer and the five-helical hydrophobic core formed by the two protomers. These hydrophobic cores are small for a protein of this size, even considering the hydrophobic interface of the coiled coil region, and are the same as in the ING5 crystal structure. As it

happens for ING5 Nt crystal, one possible explanation for this arrangement is the presence of two extra residues at the N-terminus, which favors interactions that, in the crystal, conduct to the formation of a dimer of asymmetric dimers (**Figure 25C**). In this arrangement the $\alpha 1$ helices of one dimer pack against the $\alpha 1$ helices of the other dimer found by PISA, as shown in **Figure 25D**. As it happens in the crystal structure of Nt₁₋₁₀₅, the side chain of extra residue G-1, together with the backbone of residue M1 of helix $\alpha 1$ in the open protomer form H-bonds with the side chain of residue E82 in the helix $\alpha 1$ in the close protomer of the other dimer, acting as N-cap of the last turn of $\alpha 1$ [196]. This arrangement of two dimers in the crystal is predicted by the PISA server to be an assembly in solution even more stable than the dimer. However, in solution we have determined by SEC-MALS that the protein is a dimer (**Figure 22**).

The fact that ING5 and C19S share similar interactions in the dimerization interface, together with the intra and inter coiled-coil contacts indicate that the cause of the ING5 Nt adopting an asymmetric dimer structure in the crystal is not the presence of a disulfide bond linking two molecules in the crystal lattice.

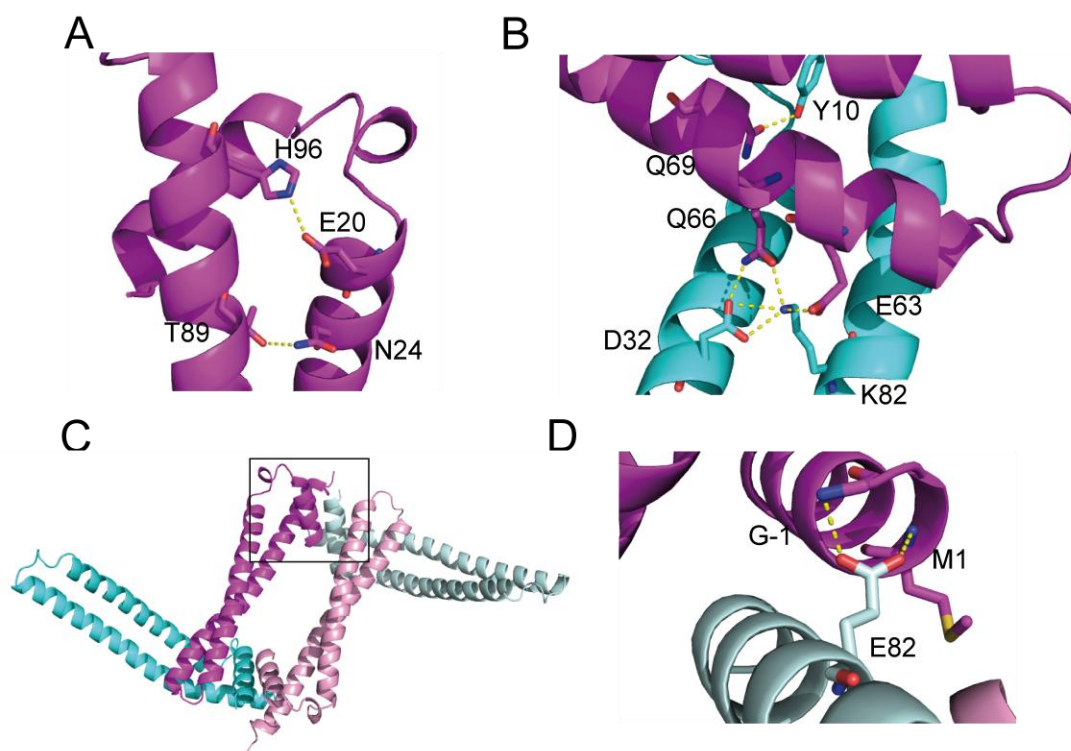


Figure 25. Intramolecular and intermolecular interactions in the ING5 Nt C19S crystal structure. (A) Side chains of residues involved in hydrogen bonds (N24-T89, E20-H96) stabilizing the small hydrophobic core formed by the open protomer. (B) Dimerizing interface and polar contacts between the open and close protomers (Y10-Q69). The side chains of D32 and K82, which are involved in a salt bridge, stabilizing the *coiled-coil* structure in the closed protomer are also shown. These residues are also involved in polar contacts between the open and close protomers (D32-Q66 and K82-E63). (C) Ribbon diagram of the tetrameric assembly in the crystal of ING5 Nt C19S identified by PISA analysis. The tetramer is formed by two asymmetric units, with the chains of the second unit indicated with paler colors. (D) Zoom of the framed area in C. Interactions between the backbone amides of residues G-1 and M1 in helix $\alpha 1$ from an open protomer with the side chain of rE8 in the helix $\alpha 1$ of a close protomer of another dimer in the tetrameric assembly.

The crystallographic models of both ING5 Nt domain and C19S mutant present different symmetry and dimerization interface compared to the crystal structure of ING4. But the data in solution indicate a symmetric structure also for ING5 Nt. We tried to use SAXS to discriminate between both types of conformations in solution. SAXS data analysis shows that ING5 Nt and C19S are folded proteins in solution as we can observe from the pronounced peak in the Kratky plot of both Nt and C19S (**Figure 26**). The distance distribution function, $p(r)$, of Nt and C19S were calculated with a maximum particle dimension of 89 Å (**Figure 26**) and a radius of gyration of 25 Å for both proteins, values similar to those measured for the ING4 Nt [102].

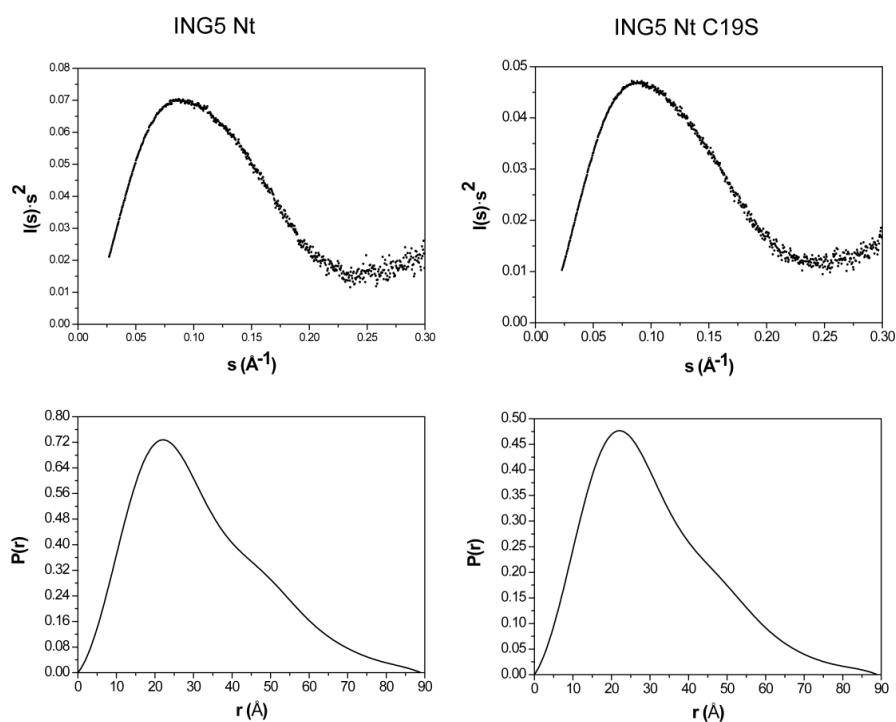


Figure 26. Kratky plots (top panels) and distance distribution functions (bottom panels) of ING5 Nt and its C19S mutant. The data were derived from the SAXS measurements of ING5 Nt and its C19S mutant at 20 °C in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP.

The SAXS data allowed us to generate low resolution *ab initio* structural models for ING5 Nt domain that shows an elongated molecules (**Figure 27A, C**). CRY SOL was used as a tool for SAXS data and structural data comparison. The fitting of the ING5 Nt experimental SAXS data to the calculated scattering (**Figure 27B**) is better for the symmetric model based on ING4 Nt ($\chi^2 = 1.8$) than for the asymmetric crystal structure ($\chi^2 = 9.8$), indicating that in solution the structure of the ING5 dimer is similar to the symmetric model based on ING4 Nt. The SAXS data of the C19S also yields an elongated shape that fits better the symmetric model than the asymmetric one ($\chi^2 = 2.0$ and 12.0, respectively; **Figure 27C, D**). We can conclude from these experiments, together with the CD and NMR data, that ING5 Nt is folded as a helix-loop-helix

structure that dimerizes into a symmetric four-helix coiled-coil, with flexible N-terminal tails that may adopt a transient helical structure.

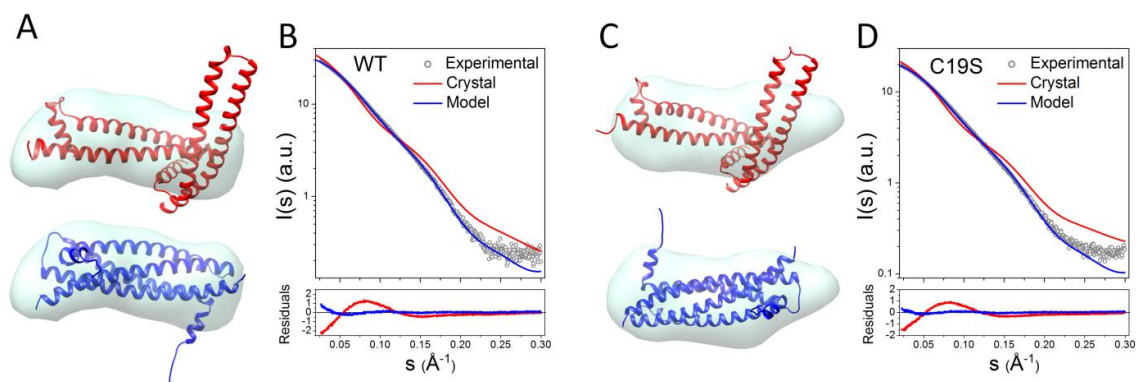


Figure 27. Low resolution structure of ING5 Nt dimer in solution observed by SAXS. (A) Shape reconstruction of the ING5 dimer showing the fit of the data to the asymmetric crystal structure (top panel, structure in red ribbon) or the ING4 homology based symmetric model (bottom panel, model in blue ribbon). (B) Comparison of the experimental scattering (grey circles) with that calculated for the dimer in the crystal (red, $\chi^2 = 9.8$) or the homology model dimer (blue, $\chi^2 = 1.8$). (C) Same as in A for the C19S mutant (D) Same as in C for the C19S mutant ($\chi^2 = 12.0$ and 2.0 for the crystal and the model structures, respectively).

4.4. ING5 Nt domain interaction with JADE1L

JADE 1L is one of the components of the HBO1 complex, and it has been proposed that it is the platform on which the other components assemble [94]. In pull-down experiments it was shown that the conserved Iib domain of JADE 1L interacts directly with ING4, ING5 and with an ING4 construct lacking the PHD [139]. We hypothesized that JADE 1L would interact with the Nt domain of ING4 and ING5. Taking advantage of the assignment of the NMR spectrum of ING5 Nt, we examined the possible binding of ING5 Nt to the Iib domain of JADE 1L by monitoring the perturbations caused on ING5 signals in the presence of the Iib domain of JADE 1L. Since the expression and purification of the JADE domain was unsuccessful we used for this experiment three designed peptides with overlapping sequences corresponding to the Iib domain of JADE 1L (see materials and methods, **Table 2**). The ING5 Nt domain precipitated in the presence of peptide A, indicating an interaction, but it was not possible to ascertain if the interaction was specific.

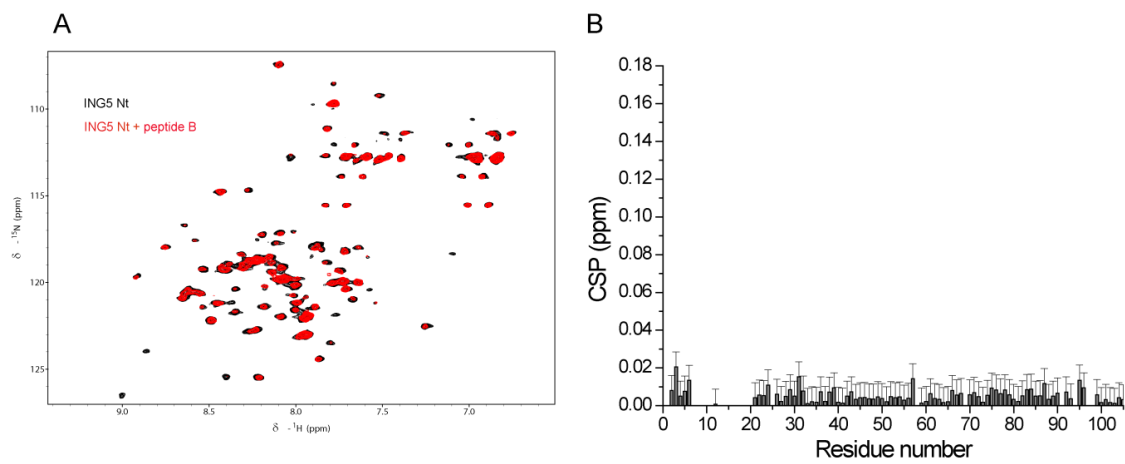


Figure 28. (A) ING5 Nt ^{15}N in the absence (**black**) or in the presence of Jade peptide B (**red**) of a 1:5 molar ratio. Samples were prepared in 50 mM MOPS pH 7.0, 1 mM DTT @ 298K. (B) CSP analysis for each residue in the presence of JADE 1L peptide B.

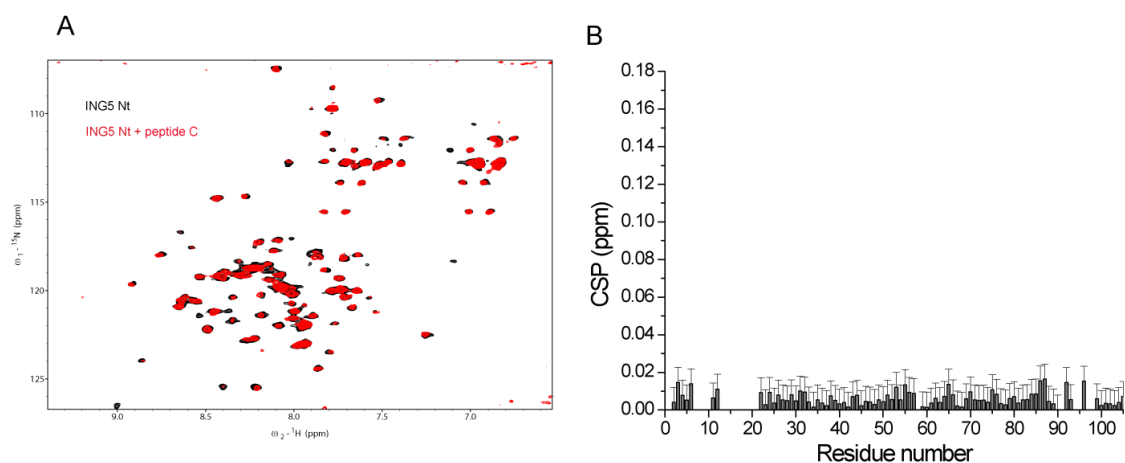


Figure 29. (A) ING5 Nt ^{15}N in the absence (**black**) or in the presence of Jade peptide C (**red**) in a 1:5 molar ratio. Samples were prepared in 50 mM MOPS pH 7.0, 1 mM DTT @ 298K. (B) CSP analysis for each residue in the presence of JADE 1L peptide C.

In **Figures 28** and **29** we can see the results in the presence of peptides B and C, respectively. The CSP measured in the presence of the other two peptides are very small and suggest that there is no specific interaction or that the interaction is extremely weak (**Figure 28B** and **29B**). Therefore, we cannot identify which is the region of ING5 involved in the interaction with Iib domain of JADE 1L.

5. Oligomerization of ING5 inside living cells

5.1. Homodimerization of ING5

To confirm that ING5 forms homodimers not only as a pure protein in solution but also inside the cell, we performed co-immunoprecipitation experiments in human embryonic kidney epithelial 293T (HEK293T) cells transiently expressing ING5 with different N-terminal tags: HA or AU5. We have observed that HA-ING5 co-precipitated with AU5-ING5 by using an antibody anti-AU5 tag (**Figure 30**). These results were reproducible (see **appendix I**) and indicate that at least two ING5 molecules interact inside living cells.

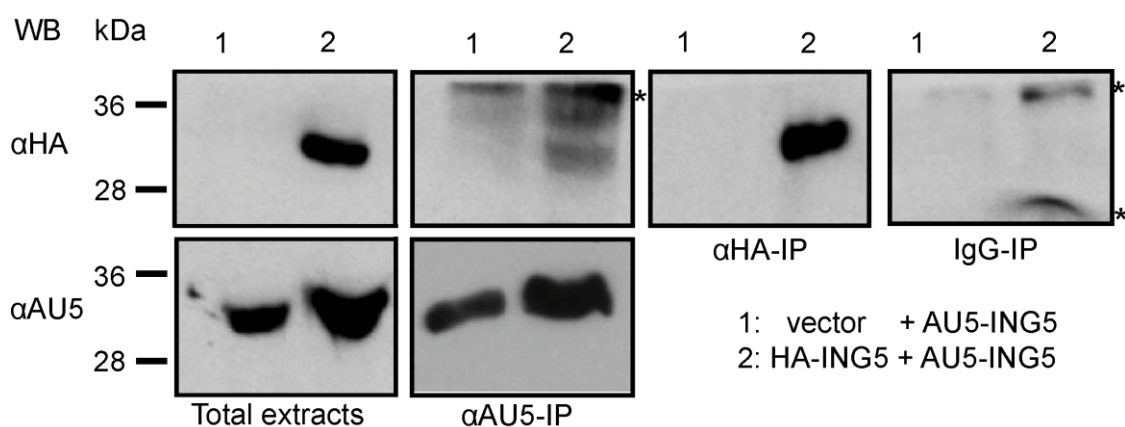


Figure 30. Analysis of ING5 homodimerization inside cells by co-immunoprecipitation. Lysates from cells transiently transfected with vectors expressing AU5-ING4 (left lanes) or both AU5-ING4 and HA-ING5 (right lanes) were immunoprecipitated (IP) with α AU5 or α HA, antibodies against the AU5 tag and HA tag, respectively. The presence of each protein in the immunocomplexes was analyzed by Western blotting (WB) with antibodies against the HA (α HA, upper panels) or AU5 (α AU5, lower panels) tag.

5.2. Heterodimerization of ING5

As it has been discussed in the introduction, the high sequence identity of ING5 and ING4 at their Nt domains (75%), and that both may form part of the HBO1 complex [92], raises the possibility that ING5 and ING4 form heterodimers. Since ING5 is a symmetric dimer in solution, as ING4, some inter helical interactions could be conserved in the ING5/ING4 heterodimer stabilization. An ING5/ING4 heterodimer model based on ING4 crystal structure has an overall distribution of hydrophobic and polar residues [103] and an heptad pattern of the coiled coil that is the same as in the modeled structure of ING5 homodimer and the crystal structure of ING4. Moreover, the intermolecular salt bridge and hydrogen bond between residues at positions 32/39 and 94/69 present in the ING4 homodimer could also be established in the heterodimer [103]. The formation of heterodimers inside cells was observed by co-immunoprecipitation experiments. Extracts

from HEK293T cells transiently expressing both ING5 and ING4 with the two different tags (HA or AU5), were co-immunoprecipitated against the AU5 tag and blotted with the HA tag (or vice versa). With this strategy, co-precipitation of ING5 or ING4 will only occur if they interact. The detection of HA-ING4 by western blotting with anti-HA antibody in immunocomplexes for AU5-tagged ING5 (**Figure 31**) shows that ING5 and ING4 can form heterodimers, or at least that they form part of the same complex inside the cells. These results were reproducible (see **appendix I**).

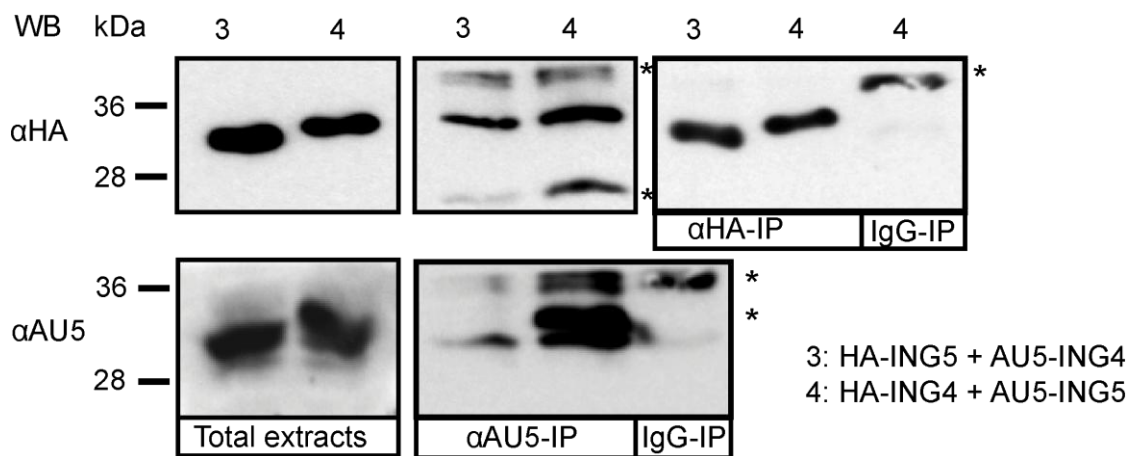


Figure 31. Analysis of ING4/ING5 heterodimerization in cells by co-immunoprecipitation. Lysates from cells transiently transfected with vectors expressing HA-ING5 and AU5-ING4 (lanes with number 3) or HA-ING4 and AU5-ING5 (lanes with number 4) were immunoprecipitated (IP) with α HA or α AU5, antibodies against the HA tag and AU5 tag, respectively. The presence of each protein in the immunocomplexes was analyzed by Western blotting (WB) with antibodies against HA (α HA, upper panels) or against AU5 (α AU5, lower panels) tag.

The formation of heterodimers of the recombinant Nt domains was also observed by ion mobility coupled to electrospray ionization mass spectrometry (ESI-IM-MS) on samples of the pure Nt domains mixed and incubated at 37 °C (**Figure 32** and **Table 3** in the materials and methods section). However, the proportion of the heterodimers is very low relative to the dimers or dissociated monomers, and is comparable to heterotrimers (which were unexpectedly seen in the mass spectrum). This result suggest that isolated Nt domains can form heterodimers, although they are not the most stable species. This is consistent with the unsuccessful attempts of copurification of ING5 heterodimers from mixtures of the two proteins [121].

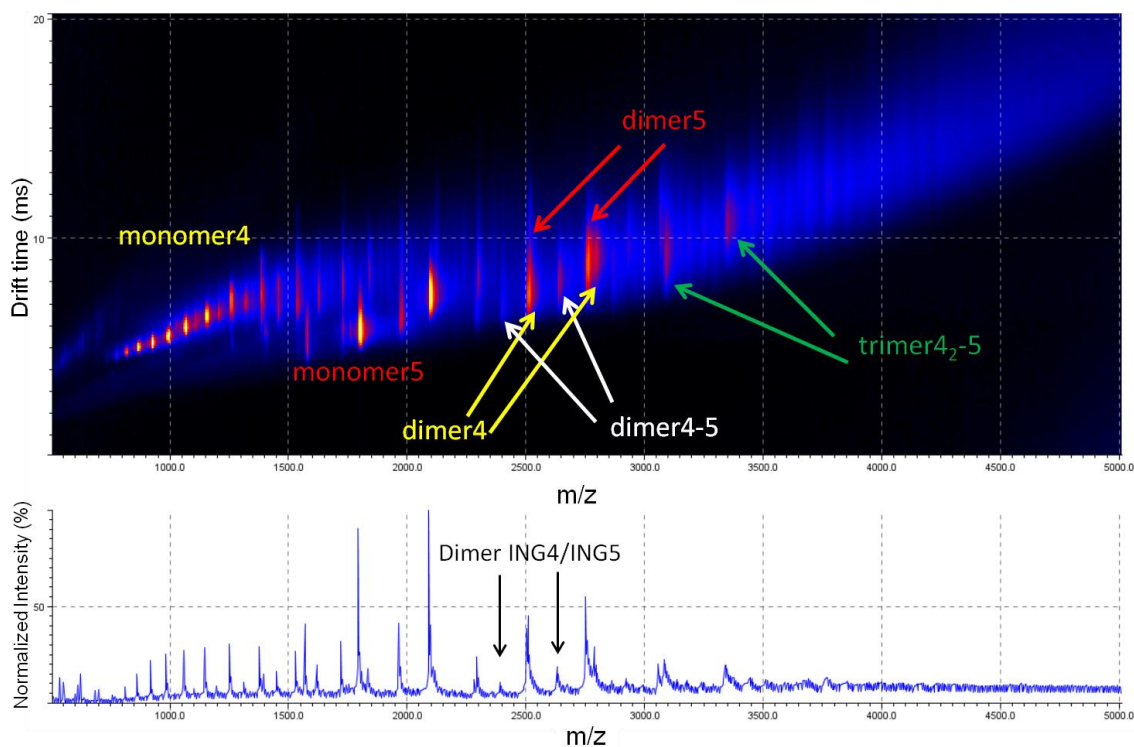


Figure 32. ESI-IM mass spectra analysis of ING5 Nt and ING4 Nt mixture. (Top) Ion mobility drift time *versus* m/z plot of the data acquired for the mixture of ING4 and ING5 Nt domains. The interpretation of the signals in terms of the different monomeric, dimeric, and trimeric forms of the ING4 and ING5 Nt domains (4 and 5, respectively) are indicated with arrows of different colors. (Bottom) ESI-MS spectrum of the mixture of ING4 and ING5 Nt domains. The signals consistent with the ING4/ING5 dimer are indicated.

Altogether these results suggest that the isolated Nt domains can form heterodimers. The heterodimer might be much less stable in solution than the homodimer, but in the cells the heterodimer may be stabilized by other proteins or by post-translational modifications.

6. Structural impact of ING5 N-terminal mutants detected in primary tumors

In the ING5 Nt domain, three tumor specific somatic mutations have been described in oral squamous cell carcinoma [117] (Figure 33A). These mutants, the three of them located in the long helices $\alpha 2$ and $\alpha 3$, are Q33R, I68V and C75R and form dimers in solution (Figure 33B). The CD spectra of all of them are typical of coiled-coil structures, indicating that they adopt helical structures similar to that of the wild type (Figure 33B). However, their stability is markedly different as measured by thermal denaturation (Figure 33C). While the Q33R mutant behaves similarly to the WT, as expected for a solvent exposed polar residue mutated into another polar residue, both I68V and C75R mutants cause a dramatic structural destabilization. The fact that I68V and C75R mutations involve buried or partially buried residues that are changed to a smaller or a

larger and more polar residue, respectively, is consistent with the observed destabilization.

To analyse the effect of the mutations on dimer formation in a more quantitative way, we calculated the dissociation constants of the coiled-coil dimerization at 25°C. Assuming a two state transition as indicated by the single transition observed in the melting curves, (**Figure 34**) it is possible to calculate the K_D at a given temperature from the variation of the mid-point denaturation temperatures with the protein concentration [146, 147]. The calculated K_D for the WT and the three mutants at 25 °C are presented in **Table2**.

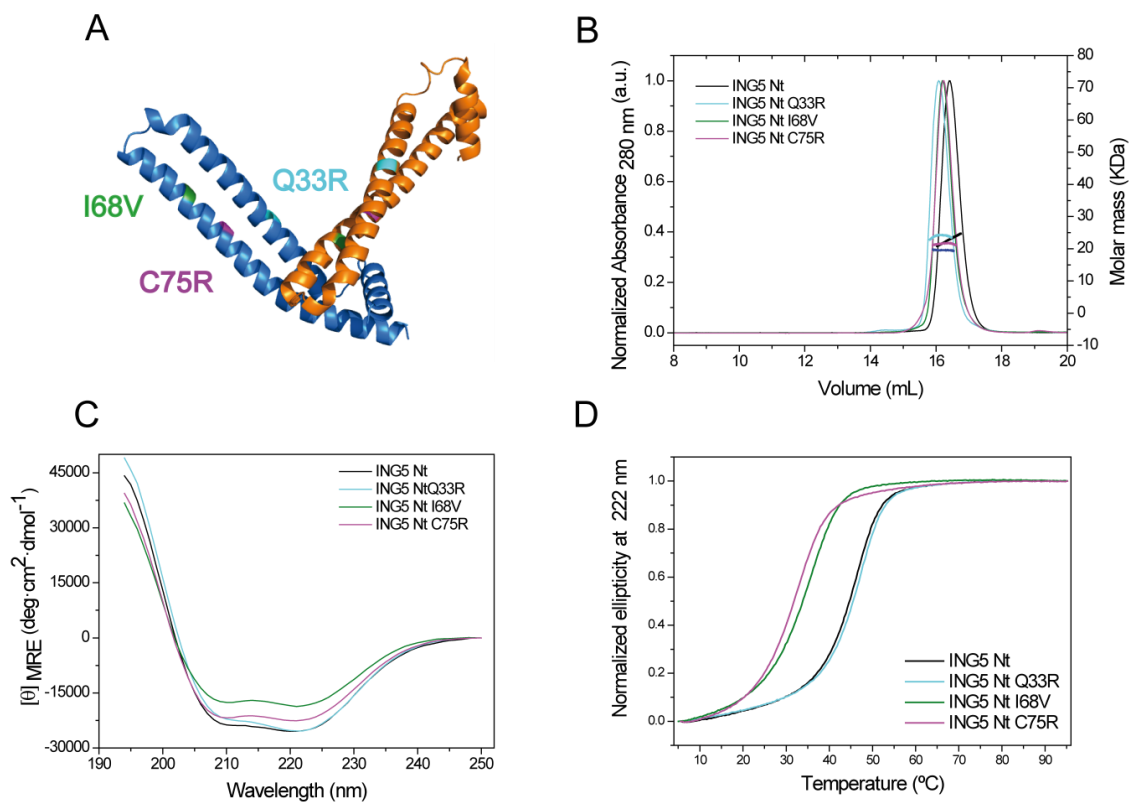


Figure 33. ING5 cancer mutants dimerize in solution through its N-terminal helical domain. (A) View of the ING5 Nt homodimer in the asymmetric unit of the crystal. The two protomers are shown as ribbons with orange (open) and blue (closed) color. The position of the three mutants detected in primary tumors are colored and indicated in the closed protomer. (B) SEC-MALS analysis of ING5Nt and its mutants Q33R, I68V and C75R. All data were obtained in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, at 25 °C. The exclusion volume of the column is 8.7 mL. (C) CD spectra of ING5 molecules at 40 μ M in the same buffer as in B. (D) Thermal denaturation of ING5 molecules in the same buffer as in C.

For ING5 Nt the calculated value of K_D is 4.1 ± 0.7 nM, In the case of Q33R mutant, the introduction of a polar and positively charged residue in a solvent exposed position is consistent with the measured stabilization ($K_D = 0.4 \pm 0.3$ nM). Furthermore, Q33 residue is close to E36 and D37 residues, and the introduced arginine will make helical stabilizing contacts ($i, i+3$ and $i, i+4$ interactions), not only in the symmetric but also in the asymmetric structural model. However, for the I68V and C75R mutants the K_D values are three orders of magnitude larger (1.4 ± 0.2 μ M and 2.7 ± 0.5 μ M, respectively). The strong destabilization caused by these mutations is consistent with both residues being

completely buried in the symmetric model, while they are only partially buried in the asymmetric one.

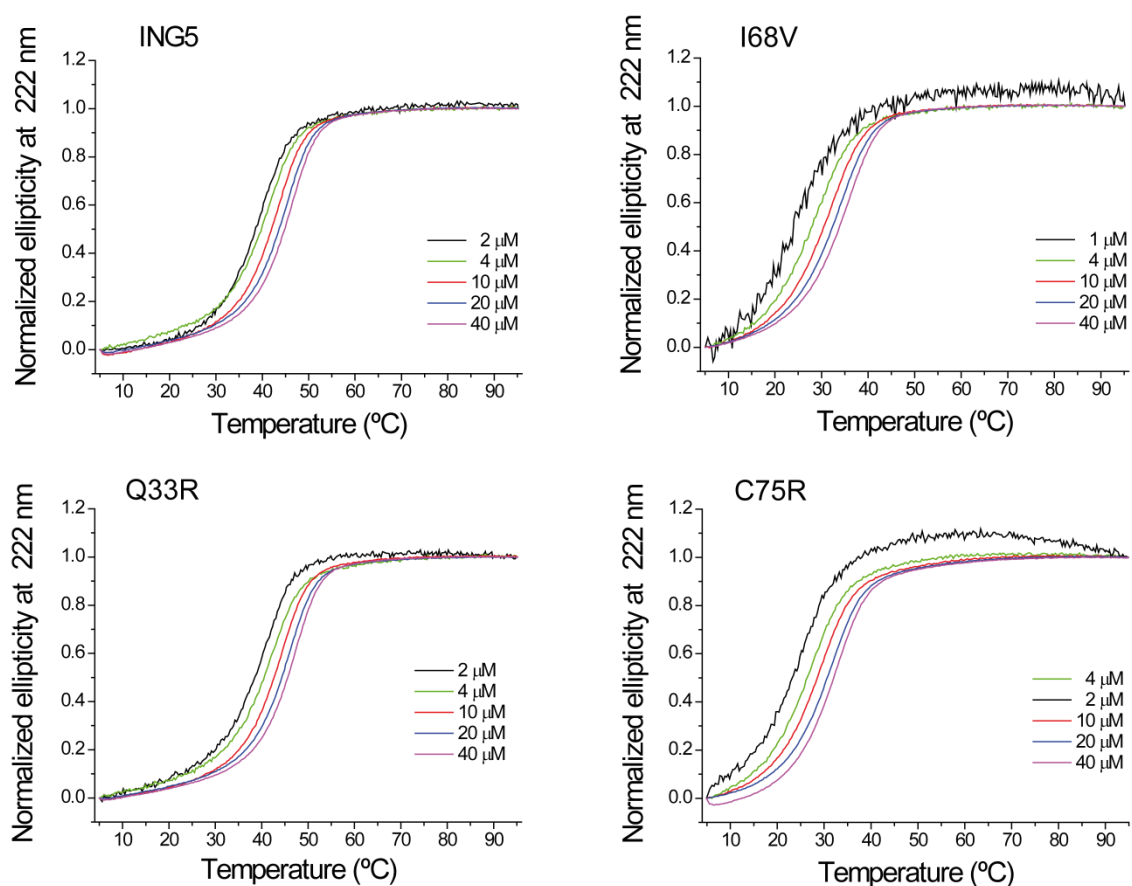


Figure 34. Thermal denaturation curves for ING5 Nt and its cancer mutants at different concentrations. The normalized ellipticity at 222 nm at each temperature is represented : 40 μM protein in magenta, 20 μM in blue, 10 μM in red, 4 μM in olive, and 2 μM in black (except ING5 Nt I68V, which is 1 μM).

Table 2. Dissociation constant K_D for ING5 Nt and its mutants at 25 °C.

Protein	K_D (M)
ING5	$4.10 \pm 0.69 \cdot 10^{-9}$
Q33R	$0.43 \pm 0.27 \cdot 10^{-9}$
I68V	$1.39 \pm 0.20 \cdot 10^{-6}$
C75R	$2.70 \pm 0.49 \cdot 10^{-6}$

7. Functional impact of ING5 N-terminal mutants

7.1. Cell morphology

To study the functional impact of the expression of ING5 cancer-associated mutants, we first analyzed cell morphology by immunofluorescence. Cells expressing ING5 mutants showed an aberrant morphological phenotype. This can be observed in the immunofluorescence micrographs of **Figure 35**. Cells expressing WT ING5 displayed a cell shape and nuclei/cytoplasm distribution similar to control cells. Cells expressing the Q33R mutant were the most similar to the WT cells, although they were rounder and smaller. However, cells expressing I68V and C75R mutants, the most structurally unstable proteins, displayed features similar to each other but different from control and WT cells: they had large nuclei and were distributed in clusters and/or showed multinuclear cells.

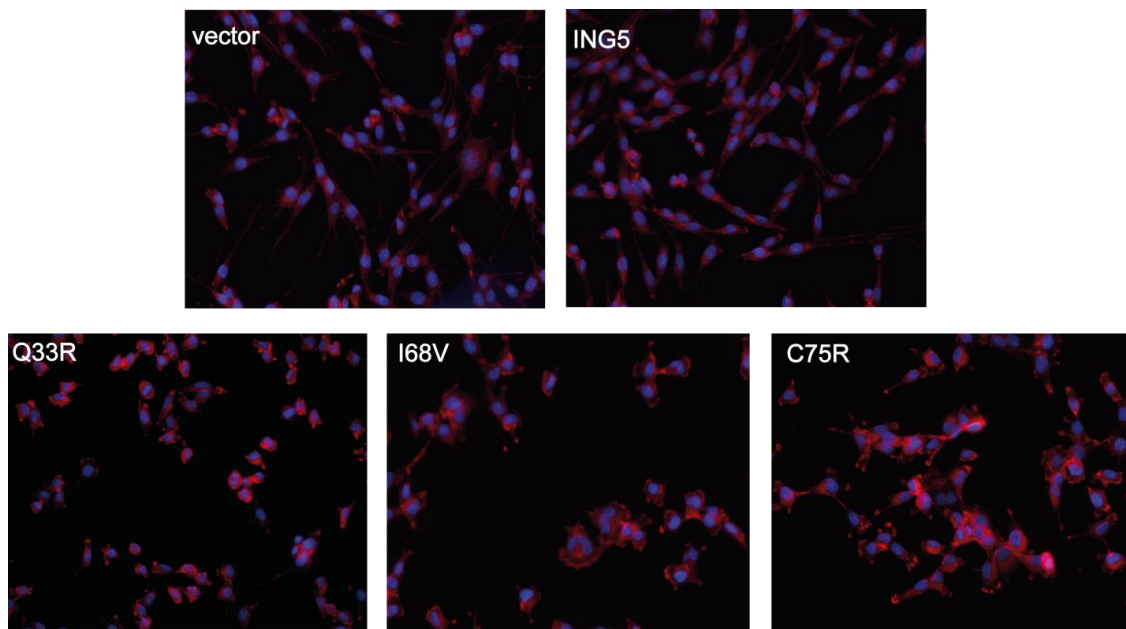


Figure 35. Micrographs of NIH3T3 cells with stable expression of ING5 or its mutant forms. Cells were co-stained with Hoechst 33342 for DNA (blue) and Rhodamine–phalloidin for actin (red) to mark nuclei and cytoplasm, respectively.

7.2. ING5 N-terminal mutants sublocalization in cells

We confirmed by immunofluorescence that all the ING5 protein variants were efficiently expressed (**Figure 36**). ING5 showed a predominantly nuclear localization in speckles, possibly nucleoli, while the mutant forms of ING5 displayed a subcellular distribution slightly different, accumulated on a region of the nuclei close to the cytoplasm.

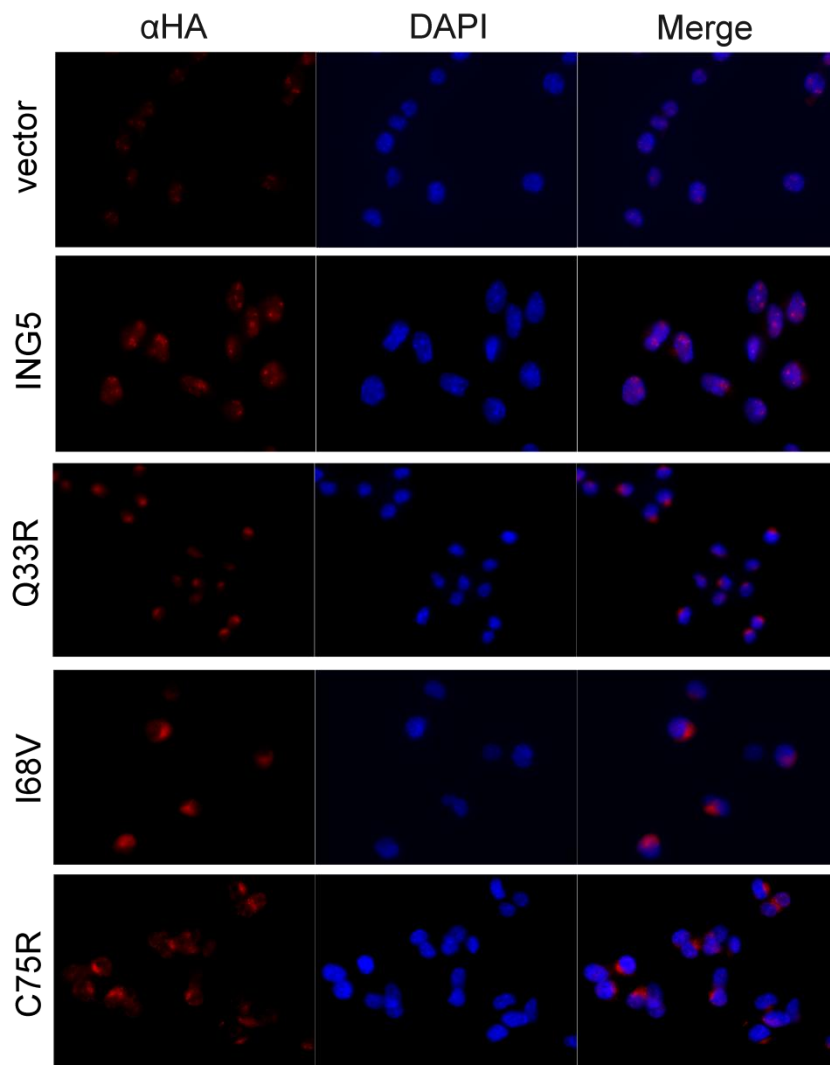


Figure 36. Immunofluorescence analysis of the expression level and subcellular localization of WT and mutant forms of ING5 in stably transfected NIH3T3 cells. An antibody against the HA tag at the N-terminal part of the ING5 proteins was used to visualize ING5 proteins (red). Cells were mounted onto microscope slides using DAPI containing Vectashield to visualize the nuclei (blue).

7.3. Cell proliferation effect of ING5 N-terminal mutants

The functional impact when expressing the ING5 cancer-associated mutants was further examined by proliferation assays in NIH3T3 fibroblasts stably expressing the WT protein and the Q33R, I68V and C75R mutants. The cells were plated and analyzed after 20, 68 and 92 h in culture (**Figure 37**). Cells expressing ING5 displayed a proliferative

behavior that was indistinguishable from control cells (pLPC vector). In contrast, cells expressing the mutant versions of ING5 showed compromised growth. In particular, expression of C75R and I68V mutants markedly reduced cell proliferation, while Q33R mutant exerted a more modest effect.

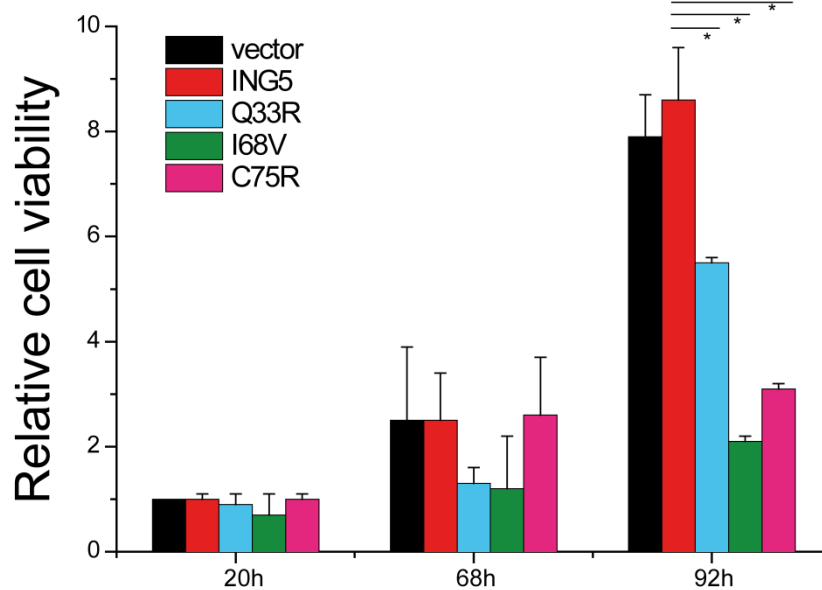



Figure 37. Proliferation assay of NIH3T3 cells expressing ING5 constructs at different times. The data represent the mean \pm S.D. of three independent experiments done in quadruplicates. * $p < 0.05$ by *t*-test.

7.4. Cell cycle profile of ING5 N-terminal mutants expressing cells

The aberrant cell phenotype of the ING5 mutants was further examined by flow cytometry analyzing their cell cycle distribution using stably transfected cells (**Figure 38**). ING5 expressing cells displayed a cell cycle phase distribution that is indistinguishable from the control. However, Q33R mutant expressing cells showed increased S phase arrest (**Figure 38A**, left column), which has been shown in breast cancer to occur under DNA-damage [198]. I68V and C75R mutants produced a significant reduction of cells going through G0/G1 phase due to the increased population of cells showing a G2/M arrest and aneuploidy, even in the absence of external stress conditions (**Figure 38B**, Top). To assess the capacity of mutated ING5 to affect the cell response to DNA damage, cells were treated with doxorubicin, which intercalates with DNA. Induction of DNA damage with a low (100 ng/mL) or a high (400 ng/mL) dose of doxorubicin, resulted in a clear accumulation of WT and Q33R cells in the G2/M phase of the cell cycle and subsequently showed a progressive decline in the G1- and S-phase populations (**Figure 38A**, second and right columns). In contrast, cells expressing C75R and I68V ING5 variants seemed to be unable to arrest in the G2/M check point, leading



to significant accumulation of replicated DNA, as observed in the huge levels of aneuploidy (**Figure 38B**, middle and bottom). These results suggest that those cells expressing Q33R mutant are still able to proceed with their cell cycle and be functional, as ING5, while the expression of C75R and I68V mutants makes the cells unable to arrest the cell cycle under DNA damage leading to aneuploidy. This aberrant behavior could be related to their destabilized structure, which may affect ING5 function.

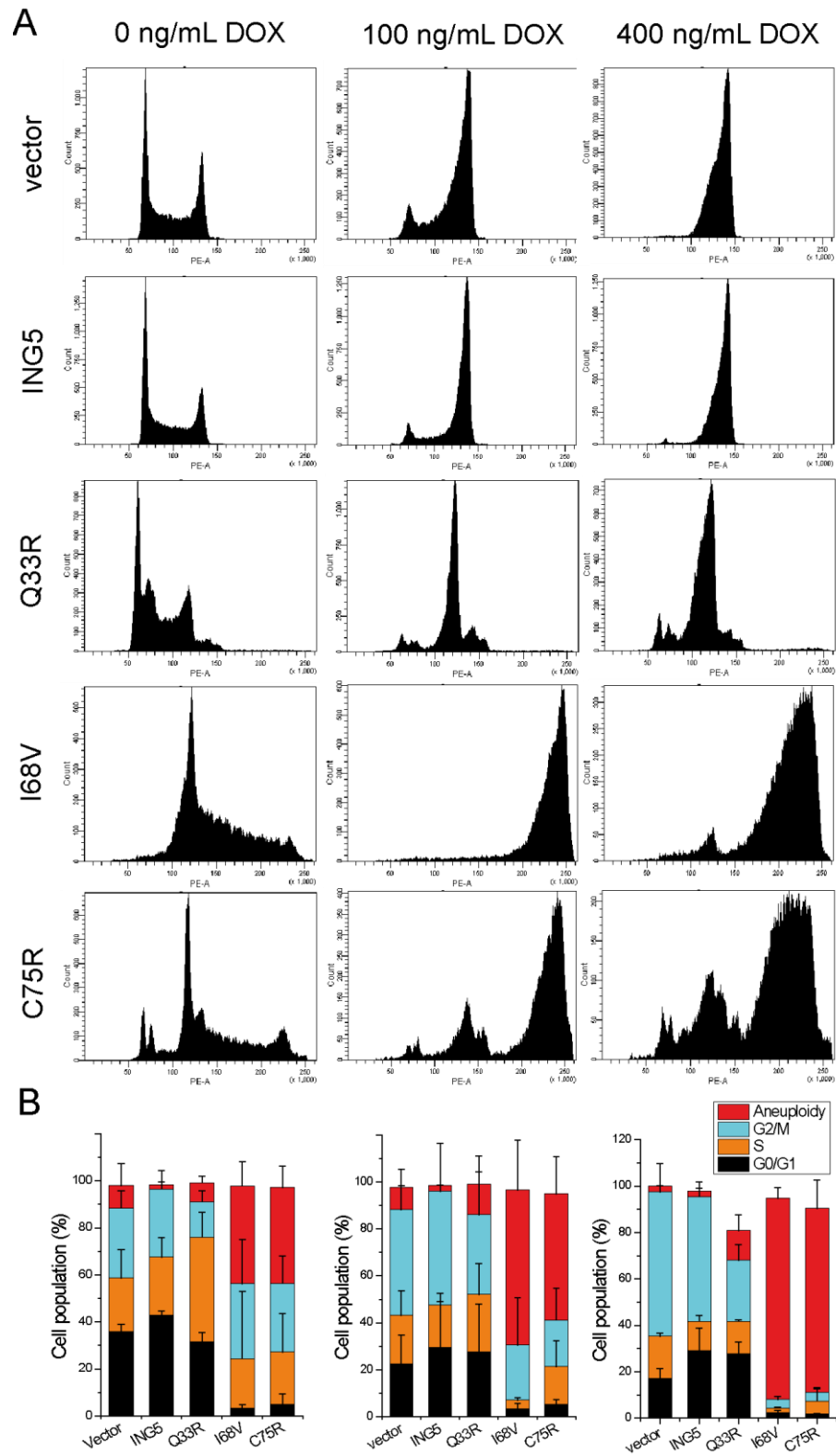
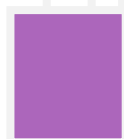
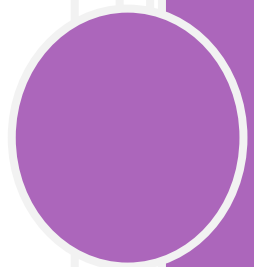


Figure 38. (A) Cell cycle profiles of cells stable expressing ING5 or mutants. (B) Analysis of cell cycle distribution in the presence of ING5 and its mutants.

DISCUSSION



Previous studies on the structural characterization of ING5 have reported on the biochemical and structural features of the PHD as a trimethylated histone binding module, leaving out the N-terminal domain and, the most important, the overall structure of ING5. I started this thesis work soon after our group found that the crystal structure of the N-terminal domain of ING5 showed a dimer with α -helical secondary structure [121], as it had been observed for ING4. However, the crystal structure of the Nt of ING5 revealed a different dimerization interface to that described for the highly homologous ING4, being the former an asymmetric dimer, and the latter a symmetric one.

In this thesis work, we have been able to express and purify the full length version of ING5, which has been very challenging because of the low purification yields and the low solubility of the recombinant protein. Despite of the difficulties, we prepared a good quality sample that allowed us to perform structural analysis by SEC-MALS, circular dichroism and NMR.

The circular dichroism spectrum of ING5 is indicative of a high content of helical structure, similar to the spectrum of its isolated Nt. However, the low absolute ellipticity indicates that a large part of the chain is not helical, which is consistent with the presence of other different structural domains in the molecule. ING5 shows cooperative thermal denaturations with a midpoint melting temperature of 42°C, which is similar to that of its isolated N-terminal domain (45°C). However, the calculated T_m are smaller compared to those of the corresponding ING4 molecules (56°C) [102]. There are differences in the experimental conditions, in which the thermal denaturations of both proteins were measured (Tris pH 8.0 for ING5 molecules and sodium phosphate pH 6.5 for ING4 molecules). The different experimental conditions might possibly be the reason for the differences in T_m ; however, there is data in which the midpoint melting temperature of ING5 Nt does not vary with these pH values [121].

We have been able to experimentally determine a molar mass of 57 kDa for ING5 in solution by SEC-MALS, which is consistent with the theoretical calculated mass of a dimer (60 kDa). The dimerization of ING5 has been shown to occur also in cells, as it was proved by co-immunoprecipitation experiments. Thus, ING5 is a dimer in solution and the dimerization site is at the N-terminal domain, with α -helical coiled-coil structure. The NMR spectrum of ING5 shows a single set of dispersed signals that matches that of the PHD domain, indicating that the two PHD finger of the dimer are chemically equivalent. Therefore, ING5 is a bivalent reader of the H3K4me3 mark, as is ING4. We have determined the dissociation constant of ING5 binding to H3K4me3 peptide by NMR, yielding a K_D of $7.3 \pm 2.6 \mu\text{M}$, in the same micromolar range but around 2.5 times smaller than the K_D of the isolated PHD finger bound to the same peptide and measured

by the same method ($K_D = 17.9 \pm 2.5 \mu\text{M}$). This difference in the measured affinity is similar to that observed between ING4 and its isolated PHD (K_D values of $1.3 \pm 1.0 \mu\text{M}$ and $3.9 \pm 1.0 \mu\text{M}$, respectively). Perhaps this difference is due to a loss of avidity in the case of the PHD finger alone, which lacks the bivalent mode of reading H3K4me3 of the full length proteins, as it happens for ING5 and ING4 dimers.

Our structural characterization on the N-terminal domain of ING5 in solution by CD, SEC-MALS, NMR and SAXS unequivocally shows that the N-terminal domain folds into a dimeric, symmetric, coiled-coil structure, probably antiparallel as it was found for ING4. Therefore, the crystal structure of this domain, previously determined [121], is likely a crystallization artefact. It probably happens that under the crystallization conditions (a buffer with pH 6.5 and a high concentration of poly-alcohols like MPD and PEG) the asymmetric dimer conformation (or the symmetric tetramer identified by PISA analysis) exists in solution and forms crystals while the symmetric dimer does not crystallize. The reasons for this arrangement in the crystal could be either the presence of two extra residues at the N-terminus, or a close contact between the C19 side chain of one ING5 molecule and the corresponding C19 side chain of a symmetry related molecule, forming a disulfide bridge that stabilizes the asymmetric dimer. However, structural analysis of the Nt₂₋₁₀₅ construct by NMR and the C19S mutant by crystallography, confirms that the crystal structure artefact is not a consequence of the presence of extra residues at the N-terminus or the intermolecular disulfide bridge formation between two molecules in the crystal lattice.

The ING5 homodimer is an elongated protein complex similar to ING4. The PHD fingers are connected to the Nt dimerization domain by the long and disordered NLS segment, and in a head-to-tail dimer, the most probable dimerization form, are directed to opposite sides of the molecule. The two C-terminal PHD fingers are chemically equivalent and autonomous from the rest of the molecule and able to interact with two nearby H3K4me3 from the same nucleosome.

Interestingly, due to the length (85 residues long, $\sim 255\text{\AA}$) and flexibility of the NLS region, one ING5 dimer would be able to bind histone H3 tails not only in the same, but in two different nucleosomes at the same time. This hypothesis has been already suggested for ING4 [103]. The structural information available for polynucleosomal states, as the tetranucleosome structure [199], sheds a light on the association between nucleosomes and their relative orientation, and provides a reliable model to test this hypothesis. An estimation of the separation between the N-terminal tails of histone H3 were obtained from the tetranucleosome model and the distance between two tails in two consecutive nucleosomes was calculated to be in the range of 79–112 \AA [103], which is

smaller than the length of the NLS region in extended conformation (~255 Å). Thus, as well as ING4, the ING5 dimer could bind the two histone H3 tails of the same or adjacent consecutive or non-consecutive nucleosomes, if they are close enough (**Figure 16** of the introduction).

In the context of chromatin, ING5, as well as the other INGs, will bind the nucleosome by the recognition of the H3K4me3 mark and recruit HATs or HDACs to these sites. These protein-protein interactions will promote epigenetic changes by the HAT or HDAC complexes that associate with INGs, but it remains unclear the mechanism by which modifications of histones within the same or nearby nucleosomes are favored. The possibility that ING proteins contribute to nucleosome binding through direct protein binding to DNA had not been explored, but in this thesis work, for the first time, we have analyzed the function of the central intrinsically disordered region (IDR) of ING4, which was previously identified as a nuclear localization sequence (NLS) containing nucleolar-targeting signals. Other intrinsically disordered proteins (IDP) or proteins containing IDRs [200] have been described to interact directly with DNA with relatively low affinity but collectively contributing to avidity for the DNA. We have shown that the NLS region of ING4 directly binds preferentially dsDNA substrates of different nucleotide sequences with an affinity in the low micromolar range [193]. The measured affinity is three orders of magnitude lower than the low nanomolar range measured for transcription factors such as p53 [201] or c-myc [202]. This is consistent with ING4 being a reader of H3K4me3 and not a transcription factor recognizing a specific DNA sequence. Our results can be probably extrapolated to ING5 and the other ING proteins since the alignment of their NLS region (IDP) sequences shows several clusters of positively charged residues, maybe ING3 being an exception because it has much less positively charged residues in this region compared to the others [72]. In the case of ING5 we have found that it also binds preferentially dsDNA, but with lower affinity than ING4.

The results presented in this thesis reveal a novel role for the central disordered NLS region of ING proteins as a modulator of the interaction with the nucleosome. Moreover, the bivalent recognition of two H3K4me3 by two PHD fingers and the bivalent binding to two DNA regions by ING4 or ING5 dimers will result in cooperativity and a strong nucleosome binding [203] (**Figure 1**). This will have an impact on the recruitment of the HAT complexes, so that the HBO1 or MOZ/MORF complexes will be more efficiently recruited to the chromatin sites with the H3K4me3 mark. Although clusters of positively charged residues are conserved among the NLS regions of ING proteins, they vary in length and sequence, which could lead to subtle differences in the selectivity or avidity of the HATs and HDAC complexes recruited by INGs. Recently, a new model of

scaffolding in HBO1 acetyltransferase complex has been proposed in which BRPF or JADE scaffold subunits interchange for different histone acetylation specificity [94]. This observation, together with the possible differences in the DNA binding of the different ING proteins, suggests that the NLS-DNA interaction may have a biological function in the ING family of proteins further from the role of nuclear localization and strengthening the nucleosome binding. The NLS regions of the ING proteins could provide flexibility to allow for multiple protein interactions in HAT and HDAC complexes.

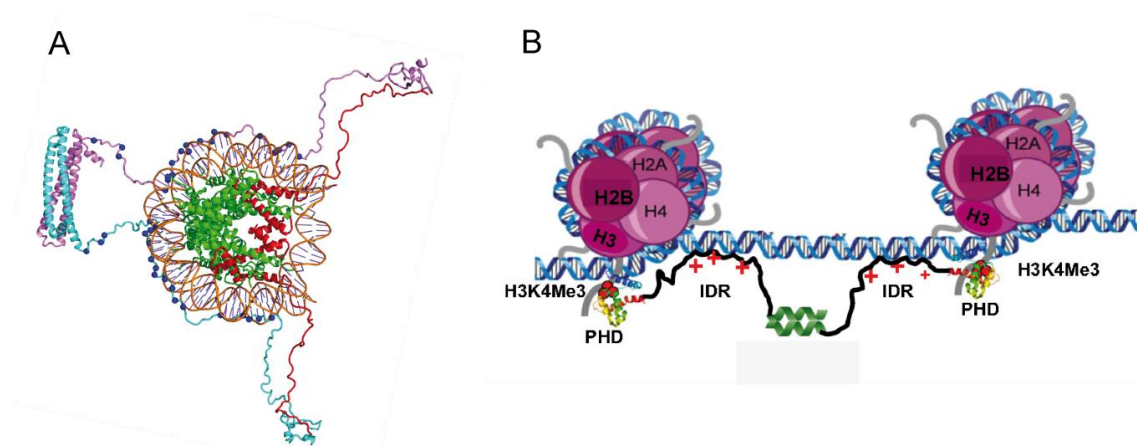


Figure 1. Model of ING4 dimer binding to chromatin. The bivalent recognition of two H3K4me3 by two PHD fingers and the bivalent binding to two dsDNA regions by two NLS regions of ING4 result in cooperativity and a strong binding either to the same (A) or different nucleosomes (B). Figure B has been modified from *Tallen and Riabowol 2017* [204].

The alignment of the ING protein sequences with the structure of the N-terminal domain of ING4 showed a high level of homology that suggested that the conservation of many of the residues is due to the structural constraints [205]. Interestingly, the regions of highest homology keep the same or similar amino acid residue in positions *a*, *d* and *g*, essential for the formation of the heptad pattern typical of coiled-coil structures in the leucine-zipper-like domain. ING4 and ING5 have been shown to be part of the same chromatin remodeling complex, the HBO1 HAT [92]. Taking into account the high sequence homology with the N-terminal domain of ING4 (75% identity), we could hypothesize the formation of a heterodimer between ING4 and ING5. The possibility of heterodimer formation was already evaluated by molecular modelling with the prediction that ING4/ING5 heterodimers may be as stable as the corresponding homodimers [103]. This prediction was based on a homology model of the ING5 dimer showing a distribution of hydrophobic and polar residues that matched, to a large extent, the intermolecular contacts in the crystal structure of ING4 [205]. Indeed, our results, for the first time, show by immunoprecipitation that ING5 and ING4 can form heterodimers, or at least that they form part of the same complex inside the cells. As it has been for some transcription factors, heterodimerization could have a regulatory role [130]. Moreover, it

could provide a different way of scaffolding leading to an exchange of recruited subunits that might result in specificity changes regarding histone acetylation. The observation, for the first time, that ING4 and ING5 can form heterodimers, indicates that the function of these proteins is more complicated than thought at first sight. It would be interesting to further study how the members of the ING family of epigenetic readers impose order on the local epigenetic status within different nucleosomes by the interplay of different members of the family and their conserved regions.

The analysis of three tumor specific somatic mutations described in oral squamous cell carcinoma [117] for ING5 Nt domain (Q33R, I68V and C75R), revealed that they form dimers in solution and showed typical CD spectra of coiled-coil structures (**Figure 2A**). Therefore these mutants adopt similar structures to that of the wild type; however, their thermal stability is markedly different, and the effect of each mutation is more consistent with an Nt domain forming a symmetric dimer than an asymmetric one, in agreement with the structural information described above. The Q33R mutant is more stable than the wild type (with a measured K_D of 0.4 ± 0.3 nM and 4.1 ± 0.7 nM, respectively), which is consistent with the fact that the Q33 residue is close to E36 and D37 residues, and the introduced arginine will make helical stabilizing contacts (i+3 and i+4 interactions) (**Figure 2B**). This will occur not only in the symmetric but also in the asymmetric structural model. However, the most unstable mutations, I68V and C75R, involve a change to a smaller or a larger and more polar residue (**Figure 2B**), respectively. The calculated dissociation constants for these mutants ($K_D = 1.4 \pm 0.2$ μ M and 2.7 ± 0.5 μ M, respectively) are three orders of magnitude higher than the one of ING5 ($K_D = 4.1 \pm 0.7$ nM). These two positions are buried in the symmetric dimer, but only partially buried in the asymmetric one; therefore, the strong destabilization is more consistent with the symmetric than with the asymmetric dimer.

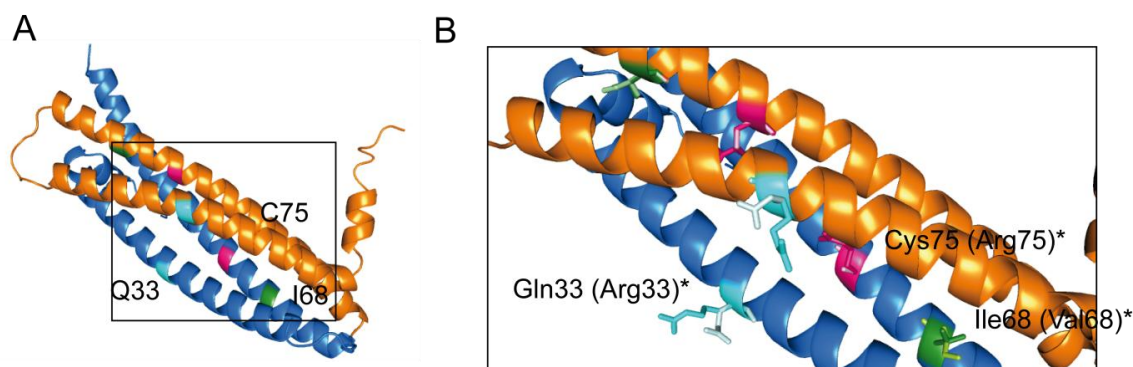



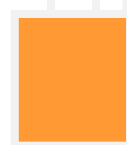
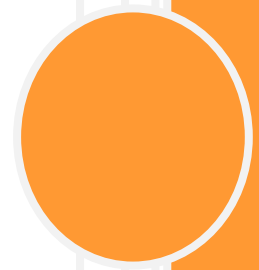
Figure 2. Mutated residues described in cancer for ING5. **A.** General view of the position of the mutations in the symmetric model of the N-terminal domain of ING5. **B.** Mutations (dark colors) and wild type residues (paler colors) displayed as sticks (Q33R in cyan, I68V in green and C75R in magenta).

In the context of the cell, the expression of these mutants leads to a different cell phenotype. Cells expressing Q33R mutant were the most similar ones to the WT expressing cells, although a bit rounder and smaller. However, Q33R displayed a subcellular distribution similar to I68V and C75R, accumulated on a region of the nuclei close to the cytoplasm. Cells expressing Q33R mutant presented an increased proportion of cells in S phase arrest, which has been shown in cancer to occur under DNA-damage or replicative stress [198]. Indeed, there is experimental data indicating that ING5 is essential for DNA replication, and in the absence of ING5, MCF7 cells showed a high proportion of cells in S-phase arrest [92]. In our model, NIH3T3 cells, the expression of Q33R mutant might interfere with the WT in complexes that are unable to form necessary interactions and mimic what it happens in the absence of ING5. Cells expressing this mutant might be unable to overcome cell cycle checkpoints, arresting cells in S-phase. However, the effect of this mutation on ING5 stability is not as drastic as the other two, as expected for a solvent exposed polar residue mutated into another polar residue, and cells expressing this mutant present only a modest effect in cell proliferation. This is consistent with Q33R mutant being similar to ING5 that has been shown to have little effect in proliferation when expressed in other cell types [115 1091]. On the contrary, cells expressing I68V and C75R mutants, the most unstable ones, were morphologically similar to each other but very different to ING5 expressing ones: they had large nuclei and were distributed in clusters and/or showed multinuclear cells. Indeed, they displayed a subcellular distribution that was slightly different to that of ING5, which is probably related with the fact that the cells expressing these mutants display an aberrant cell cycle profile with an increased cell population arrested at G2/M. Moreover, these cells appear to have lost control on their cell cycle, accumulating high amounts of aberrantly duplicated DNA. Dysregulation of cell cycle checkpoints can increase susceptibility to mutations, genomic instability and tumorigenesis [206, 207]. There is strong evidence in the literature for a high frequency of aneuploidy in cancer [208] and this can be related to what we saw in our model, in which cells expressing I68V and C75R mutants seem to be unable to overcome cell cycle checkpoints, accumulating DNA aberrantly and showing aneuploidy. In the presence of genotoxic insult (doxorubicin), this situation aggravates in the case of I68V and C75R mutants, displaying huge levels of aneuploidy. However, this is not the case for Q33R mutant, whose cell cycle profile is similar to ING5 and control cells.



Our results suggest that the expression of ING5 mutants described in oral squamous cell carcinoma [117], in NIH3T3, cells have drastic effects in the cell cycle in the case of I68V and C75R, which correlates with tumorigenic activity. However, there are still some questions to address regarding the proliferation of the cells expressing these mutants. Their cell cycle is completely out of control which is the typical scenario in cancer cells that are able to skip cell cycle checkpoints and keep proliferating, but these mutants showed a markedly reduced cell proliferation more typical of tumor suppressive activity. There is not many published data about ING5 functions but there are conflicting views of ING5 as a tumor suppressor or oncogene, which are context specific. In fact, there are experimental evidences that suggest that ING family of proteins may play dual roles, as tumor suppressors or oncogenes, under different cellular conditions [83]. In particular, ING5 showed oncogenic activity in MEFs cultured in 0.1% FCS medium [209]. Therefore, further functional experiments in cancer cell lines should be done to better understand the effects of ING5 in DNA replication and in cell cycle, and how it contribute to its tumor suppressor or oncogenic functions depending on the cell context.

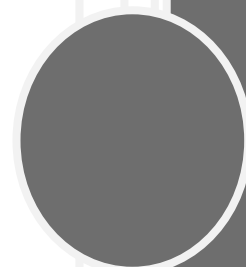
CONCLUSIONS



CONCLUSIONS

- The ING5 protein forms an elongated dimer in solution with three structurally autonomous domains: the N-terminal dimerization domain, a central disordered region and a C-terminal PHD finger. This structural organization of ING5 is similar to that of its homolog ING4.
- ING5 forms homodimers inside living cells and can also form heterodimers with ING4.
- ING5 binds the histone H3 peptide trimethylated at lysine 4 with a K_D of 7.3 ± 2.6 μM at 25 °C. Because of its dimeric nature ING5 is a bivalent reader of the H3K4me3 mark.
- The PHD finger alone binds H3K4me3 with a slightly less affinity, $K_D = 17.9 \pm 2.5$ μM at 25°C, perhaps due to avidity loss in the isolated PHD.
- The central NLS region of ING4 binds preferentially double stranded DNA with micromolar affinity. The NLS region of ING5 also binds preferentially double stranded DNA but with less affinity than ING4.
- The N-terminal domain of ING5 forms in solution a symmetric *coiled-coil* dimer similar to that of ING4, in contrast with the asymmetric dimer in the crystal, which is likely an artefact of crystallization.
- The crystal structure artefact is not a consequence of the presence of extra residues at the N-terminus or the intermolecular disulfide bridge formation between two molecules in the crystal lattice.
- The thermal stability of ING5 is similar to that of its N-terminal domain, and smaller than the corresponding ING4 molecules.
- The dissociation constant of the N-terminal dimeric domain of ING5 is 4.1 ± 0.7 nM at 25 °C, which is around three orders of magnitude smaller than the value measured for mutants I68V and C75R, and one order of magnitude larger than for mutant Q33R.
- The ING5 Q33R mutant has little impact on NIH3T3 cell proliferation, but the mutation causes S phase arrest of the cell cycle.
- The I68V and C75R mutations reduce cell proliferation and induce an aberrant behaviour in the cell cycle, with many cells showing aneuploidy.

REFERENCES



REFERENCES

- [1] Fischle W, Wang Y, Allis CD. Histone and chromatin cross-talk. *Curr Opin Cell Biol.* 2003;15:172-83.
- [2] Santos-Rosa H, Caldas C. Chromatin modifier enzymes, the histone code and cancer. *Eur J Cancer.* 2005;41:2381-402.
- [3] Kan PY, Caterino TL, Hayes JJ. The H4 tail domain participates in intra- and internucleosome interactions with protein and DNA during folding and oligomerization of nucleosome arrays. *Mol Cell Biol.* 2009;29:538-46.
- [4] Groth A, Rocha W, Verreault A, Almouzni G. Chromatin challenges during DNA replication and repair. *Cell.* 2007;128:721-33.
- [5] Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell.* 2007;128:707-19.
- [6] Ordu O, Lusser A, Dekker NH. Recent insights from in vitro single-molecule studies into nucleosome structure and dynamics. *Biophysical Reviews.* 2016;8:33-49.
- [7] Figueiredo LM, Cross GAM, Janzen CJ. Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat Rev Micro.* 2009;7:504-13.
- [8] Widom J. Role of DNA sequence in nucleosome stability and dynamics. *Q Rev Biophys.* 2001;34:269-324.
- [9] Nightingale KP, O'Neill LP, Turner BM. Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. *Curr Opin Genet Dev.* 2006;16:125-36.
- [10] Bowman GD, Poirier MG. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chemical Reviews.* 2015;115:2274-95.
- [11] Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. *Annu Rev Biochem.* 2009;78:273-304.
- [12] Sheinin MY, Li M, Soltani M, Luger K, Wang MD. Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. *Nat Commun.* 2013;4:2579.
- [13] Andrews AJ, Luger K. Nucleosome structure(s) and stability: variations on a theme. *Annu Rev Biophys.* 2011;40:99-117.
- [14] Luger K, Dechassa ML, Tremethick DJ. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol.* 2012;13:436-47.
- [15] Talbert PB, Henikoff S. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol.* 2010;11:264-75.
- [16] Wang GG, Allis CD, Chi P. Chromatin remodeling and cancer, Part I: Covalent histone modifications. *Trends Mol Med.* 2007;13:363-72.
- [17] Xu YM, Du JY, Lau AT. Posttranslational modifications of human histone H3: an update. *Proteomics.* 2014;14:2047-60.
- [18] Huynh JL, Casaccia P. Epigenetic mechanisms in multiple sclerosis: implications for pathogenesis and treatment. *Lancet neurology.* 2013;12:195-206.
- [19] Latham JA, Dent SY. Cross-regulation of histone modifications. *Nat Struct Mol Biol.* 2007;14:1017-24.
- [20] Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* 2001;15:2343-60.
- [21] Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol.* 2005;6:838-49.
- [22] Bannister AJ, Kouzarides T. Reversing histone methylation. *Nature.* 2005;436:1103-6.
- [23] Bedford MT, Richard S. Arginine methylation an emerging regulator of protein function. *Mol Cell.* 2005;18:263-72.
- [24] Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol.* 2014;15:703-8.

- [25] Casadio F, Lu X, Pollock SB, LeRoy G, Garcia BA, Muir TW, et al. H3R42me2a is a histone modification with positive transcriptional effects. *Proc Natl Acad Sci U S A*. 2013;110:14894-9.
- [26] Patel DJ. A Structural Perspective on Readout of Epigenetic Histone and DNA Methylation Marks. *Cold Spring Harb Perspect Biol*. 2016;8:a018754.
- [27] Kuo M-H, Allis CD. Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays*. 1998;20:615-26.
- [28] Struhl K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev*. 1998;12:599-606.
- [29] Bode AM, Dong Z. Inducible covalent posttranslational modification of histone H3. *Sci STKE*. 2005;2005:re4.
- [30] Nowak SJ, Corces VG. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet*. 2004;20:214-20.
- [31] Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001;293:1074-80.
- [32] Hyland EM, Cosgrove MS, Molina H, Wang D, Pandey A, Cottee RJ, et al. Insights into the role of histone H3 and histone H4 core modifiable residues in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 2005;25:10060-70.
- [33] Watanabe S, Resch M, Lilyestrom W, Clark N, Hansen JC, Peterson C, et al. Structural characterization of H3K56Q nucleosomes and nucleosomal arrays. *Biochim Biophys Acta*. 2010;1799:480-6.
- [34] Ye J, Ai X, Eugeni EE, Zhang L, Carpenter LR, Jelinek MA, et al. Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly. *Mol Cell*. 2005;18:123-30.
- [35] Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol*. 2007;14:1025-40.
- [36] Falkenberg KJ, Johnstone RW. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat Rev Drug Discov*. 2014;13:673-91.
- [37] Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128:693-705.
- [38] Khorasanizadeh S. The nucleosome: from genomic organization to genomic regulation. *Cell*. 2004;116:259-72.
- [39] Filippakopoulos P, Knapp S. The bromodomain interaction module. *FEBS Lett*. 2012;586:2692-704.
- [40] Eissenberg JC. Structural biology of the chromodomain: form and function. *Gene*. 2012;496:69-78.
- [41] Flanagan JF, Mi LZ, Chruszcz M, Cymborowski M, Clines KL, Kim Y, et al. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*. 2005;438:1181-5.
- [42] Sims RJ, Chen C-F, Santos-Rosa H, Kouzarides T, Patel SS, Reinberg* D. Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *The Journal of biological chemistry*. 2005;280:41789-92.
- [43] Blus BJ, Wiggins K, Khorasanizadeh S. Epigenetic virtues of chromodomains. *Crit Rev Biochem Mol Biol*. 2011;46:507-26.
- [44] Capili AD, Schultz DC, Rauscher IF, Borden KL. Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. *Embo J*. 2001;20:165-77.
- [45] Bienz M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem Sci*. 2006;31:35-40.
- [46] Champagne KS, Kutateladze TG. Structural insight into histone recognition by the ING PHD fingers. *Curr Drug Targets*. 2009;10:432-41.
- [47] Shindo H, Suzuki R, Tsuchiya W, Taichi M, Nishiuchi Y, Yamazaki T. PHD finger of the SUMO ligase Siz/PIAS family in rice reveals specific binding for methylated histone H3 at lysine 4 and arginine 2. *FEBS Lett*. 2012;586:1783-9.

- [48] Zeng L, Zhang Q, Li S, Plotnikov AN, Walsh MJ, Zhou MM. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature*. 2010;466:258-62.
- [49] Leung CC, Glover JN. BRCT domains: easy as one, two, three. *Cell Cycle*. 2011;10:2461-70.
- [50] Wu Q, Jubb H, Blundell TL. Phosphopeptide interactions with BRCA1 BRCT domains: More than just a motif. *Prog Biophys Mol Biol*. 2015;117:143-8.
- [51] Clapperton JA, Manke IA, Lowery DM, Ho T, Haire LF, Yaffe MB, et al. Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer. *Nat Struct Mol Biol*. 2004;11:512-8.
- [52] Chen C, Nott TJ, Jin J, Pawson T. Deciphering arginine methylation: Tudor tells the tale. *Nat Rev Mol Cell Biol*. 2011;12:629-42.
- [53] Cote J, Richard S. Tudor domains bind symmetrical dimethylated arginines. *J Biol Chem*. 2005;280:28476-83.
- [54] Lee J, Thompson JR, Botuyan MV, Mer G. Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat Struct Mol Biol*. 2008;15:109-11.
- [55] Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100:57-70.
- [56] Ponder BA. Cancer genetics. *Nature*. 2001;411:336-41.
- [57] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646-74.
- [58] Adamson ED. Oncogenes in development. *Development*. 1987;99:449-71.
- [59] Weinstein IB, Joe AK. Mechanisms of disease: Oncogene addiction--a rationale for molecular targeting in cancer therapy. *Nat Clin Pract Oncol*. 2006;3:448-57.
- [60] Ythier D, Larrieu D, Brambilla C, Brambilla E, Pedeux R. The new tumor suppressor genes ING: genomic structure and status in cancer. *Int J Cancer*. 2008;123:1483-90.
- [61] Coles AH, Jones SN. The ING gene family in the regulation of cell growth and tumorigenesis. *J Cell Physiol*. 2009;218:45-57.
- [62] Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet*. 2005;37:391-400.
- [63] Stucki M, Clapperton JA, Mohammad D, Yaffe MB, Smerdon SJ, Jackson SP. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell*. 2005;123:1213-26.
- [64] Hirota T, Lipp JJ, Toh BH, Peters JM. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature*. 2005;438:1176-80.
- [65] Ahn SH, Cheung WL, Hsu JY, Diaz RL, Smith MM, Allis CD. Sterile 20 kinase phosphorylates histone H2B at serine 10 during hydrogen peroxide-induced apoptosis in *S. cerevisiae*. *Cell*. 2005;120:25-36.
- [66] Marks P, Rifkind RA, Richon VM, Breslow R, Miller T, Kelly WK. Histone deacetylases and cancer: causes and therapies. *Nat Rev Cancer*. 2001;1:194-202.
- [67] Carlson S, Glass KC. The MOZ histone acetyltransferase in epigenetic signaling and disease. *J Cell Physiol*. 2014;229:1571-4.
- [68] Russell M, Berardi P, Gong W, Riabowol K. Grow-ING, Age-ING and Die-ING: ING proteins link cancer, senescence and apoptosis. *Exp Cell Res*. 2006;312:951-61.
- [69] Piche B, Li G. Inhibitor of growth tumor suppressors in cancer progression. *Cell Mol Life Sci*. 2010;67:1987-99.
- [70] Garkavtsev I, Kazarov A, Gudkov A, Riabowol K. Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation. *Nat Genet*. 1996;14:415-20.
- [71] Gong W, Suzuki K, Russell M, Riabowol K. Function of the ING family of PHD proteins in cancer. *Int J Biochem Cell Biol*. 2005;37:1054-65.
- [72] He GH, Helbing CC, Wagner MJ, Sensen CW, Riabowol K. Phylogenetic analysis of the ING family of PHD finger proteins. *Mol Biol Evol*. 2005;22:104-16.

- [73] Scott M, Boisvert FM, Vieyra D, Johnston RN, Bazett-Jones DP, Riabowol K. UV induces nucleolar translocation of ING1 through two distinct nucleolar targeting sequences. *Nucleic Acids Res.* 2001;29:2052-8.
- [74] Tallen G, Riabowol K. Keep-ING balance: tumor suppression by epigenetic regulation. *FEBS Lett.* 2014;588:2728-42.
- [75] Soliman MA, Riabowol K. After a decade of study-ING, a PHD for a versatile family of proteins. *Trends Biochem Sci.* 2007;32:509-19.
- [76] Scott M, Bonnefin P, Vieyra D, Boisvert FM, Young D, Bazett-Jones DP, et al. UV-induced binding of ING1 to PCNA regulates the induction of apoptosis. *J Cell Sci.* 2001;114:3455-62.
- [77] Han X, Feng X, Rattner JB, Smith H, Bose P, Suzuki K, et al. Tethering by lamin A stabilizes and targets the ING1 tumour suppressor. *Nat Cell Biol.* 2008;10:1333-40.
- [78] Peña PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, et al. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature.* 2006;442:100-3.
- [79] Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, et al. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature.* 2006;442:96-9.
- [80] Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, Lugovskoy AA, et al. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell.* 2003;114:99-111.
- [81] Kaadige MR, Ayer DE. The polybasic region that follows the plant homeodomain zinc finger 1 of Pfl is necessary and sufficient for specific phosphoinositide binding. *J Biol Chem.* 2006;281:28831-6.
- [82] Feng X, Hara Y, Riabowol K. Different HATS of the ING1 gene family. *Trends Cell Biol.* 2002;12:532-8.
- [83] Unoki M, Kumamoto K, Takenoshita S, Harris CC. Reviewing the current classification of inhibitor of growth family proteins. *Cancer Sci.* 2009;100:1173-9.
- [84] Peña PV, Hom RA, Hung T, Lin H, Kuo AJ, Wong RP, et al. Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. *J Mol Biol.* 2008;380:303-12.
- [85] Palacios A, Munoz IG, Pantoja-Uceda D, Marcaida MJ, Torres D, Martin-Garcia JM, et al. Molecular basis of histone H3K4me3 recognition by ING4. *J Biol Chem.* 2008;283:15956-64.
- [86] Champagne KS, Saksouk N, Pena PV, Johnson K, Ullah M, Yang XJ, et al. The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. *Proteins.* 2008;72:1371-6.
- [87] Marks PA. The clinical development of histone deacetylase inhibitors as targeted anticancer drugs. *Expert opinion on investigational drugs.* 2010;19:1049-66.
- [88] Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science.* 2009;325:834-40.
- [89] Zhang H, Shang YP, Chen HY, Li J. Histone deacetylases function as novel potential therapeutic targets for cancer. *Hepatol Res.* 2016.
- [90] Parthun MR, Widom J, Gottschling DE. The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. *Cell.* 1996;87:85-94.
- [91] Wapenaar H, Dekker FJ. Histone acetyltransferases: challenges in targeting bi-substrate enzymes. *Clin Epigenetics.* 2016;8:59.
- [92] Doyon Y, Cayrou C, Ullah M, Landry AJ, Cote V, Selleck W, et al. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol Cell.* 2006;21:51-64.
- [93] Kuzmichev A, Zhang Y, Erdjument-Bromage H, Tempst P, Reinberg D. Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1). *Mol Cell Biol.* 2002;22:835-48.

- [94] Lalonde ME, Avvakumov N, Glass KC, Joncas FH, Saksouk N, Holliday M, et al. Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. *Genes Dev.* 2013;27:2009-24.
- [95] Iizuka M, Sarmiento OF, Sekiya T, Scrabble H, Allis CD, Smith MM. Hbo1 Links p53-dependent stress signaling to DNA replication licensing. *Mol Cell Biol.* 2008;28:140-53.
- [96] Hung T, Binda O, Champagne KS, Kuo AJ, Johnson K, Chang HY, et al. ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. *Mol Cell.* 2009;33:248-56.
- [97] Avvakumov N, Cote J. The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene.* 2007;26:5395-407.
- [98] Contzler R, Regamey A, Favre B, Roger T, Hohl D, Huber M. Histone acetyltransferase HBO1 inhibits NF-kappaB activity by coactivator sequestration. *Biochem Biophys Res Commun.* 2006;350:208-13.
- [99] Cui S, Gao Y, Zhang K, Chen J, Wang R, Chen L. The Emerging Role of Inhibitor of Growth 4 as a Tumor Suppressor in Multiple Human Cancers. *Cellular Physiology and Biochemistry.* 2015;36:409-22.
- [100] Shiseki M, Nagashima M, Pedoux RM, Kitahama-Shiseki M, Miura K, Okamura S, et al. p29ING4 and p28ING5 bind to p53 and p300, and enhance p53 activity. *Cancer Res.* 2003;63:2373-8.
- [101] Palacios A, Garcia P, Padro D, Lopez-Hernandez E, Martin I, Blanco FJ. Solution structure and NMR characterization of the binding to methylated histone tails of the plant homeodomain finger of the tumour suppressor ING4. *FEBS Lett.* 2006;580:6903-8.
- [102] Palacios A, Moreno A, Oliveira BL, Rivera T, Prieto J, Garcia P, et al. The dimeric structure and the bivalent recognition of H3K4me3 by the tumor suppressor ING4 suggests a mechanism for enhanced targeting of the HBO1 complex to chromatin. *J Mol Biol.* 2010;396:1117-27.
- [103] Culurgioni S, Munoz IG, Moreno A, Palacios A, Villate M, Palmero I, et al. The crystal structure of the inhibitor of growth 4 (ING4) dimerization domain reveals the functional organization of the ING family of chromatin binding proteins. *J Biol Chem.* 2012.
- [104] Gunduz M, Nagatsuka H, Demircan K, Gunduz E, Cengiz B, Ouchida M, et al. Frequent deletion and down-regulation of ING4, a candidate tumor suppressor gene at 12p13, in head and neck squamous cell carcinomas. *Gene.* 2005;356:109-17.
- [105] Li M, Jin Y, Sun WJ, Yu Y, Bai J, Tong DD, et al. Reduced expression and novel splice variants of ING4 in human gastric adenocarcinoma. *J Pathol.* 2009;219:87-95.
- [106] Tapia C, Zlobec I, Schneider S, Kilic E, Guth U, Bubendorf L, et al. Deletion of the inhibitor of growth 4 (ING4) tumor suppressor gene is prevalent in human epidermal growth factor 2 (HER2)-positive breast cancer. *Hum Pathol.* 2011;42:983-90.
- [107] Wang QS, Li M, Zhang LY, Jin Y, Tong DD, Yu Y, et al. Down-regulation of ING4 is associated with initiation and progression of lung cancer. *Histopathology.* 2010;57:271-81.
- [108] Moreno A, Palacios A, Orgaz JL, Jimenez B, Blanco FJ, Palmero I. Functional impact of cancer-associated mutations in the tumor suppressor protein ING4. *Carcinogenesis.* 2010;31:1932-8.
- [109] Li XH, Kikuchi K, Zheng Y, Noguchi A, Takahashi H, Nishida T, et al. Downregulation and translocation of nuclear ING4 is correlated with tumorigenesis and progression of head and neck squamous cell carcinoma. *Oral Oncol.* 2011;47:217-23.
- [110] Raho G, Miranda C, Tamborini E, Pierotti MA, Greco A. Detection of novel mRNA splice variants of human ING4 tumor suppressor gene. *Oncogene.* 2007;26:5247-57.
- [111] Garkavtsev I, Kozin SV, Chernova O, Xu L, Winkler F, Brown E, et al. The candidate tumour suppressor protein ING4 regulates brain tumour growth and angiogenesis. *Nature.* 2004;428:328-32.

- [112] Ozer A, Wu LC, Bruick RK. The candidate tumor suppressor ING4 represses activation of the hypoxia inducible factor (HIF). *Proc Natl Acad Sci U S A*. 2005;102:7481-6.
- [113] Li J, Martinka M, Li G. Role of ING4 in human melanoma cell migration, invasion and patient survival. *Carcinogenesis*. 2008;29:1373-9.
- [114] Walzak AA, Veldhoen N, Feng X, Riabowol K, Helbing CC. Expression profiles of mRNA transcript variants encoding the human inhibitor of growth tumor suppressor gene family in normal and neoplastic tissues. *Exp Cell Res*. 2008;314:273-85.
- [115] Linzen U, Lilischkis R, Pandithage R, Schilling B, Ullius A, Luscher-Firzlaff J, et al. ING5 is phosphorylated by CDK2 and controls cell proliferation independently of p53. *PLoS One*. 2015;10:e0123736.
- [116] Chen W-T, Yang Y-J, Zhang Z-D, An Q, Li N, Liu W, et al. MiR-1307 promotes ovarian cancer cell chemoresistance by targeting the ING5 expression. *Journal of Ovarian Research*. 2017;10:1.
- [117] Cengiz B, Gunduz E, Gunduz M, Beder LB, Tamamura R, Bagci C, et al. Tumor-specific mutation and downregulation of ING5 detected in oral squamous cell carcinoma. *Int J Cancer*. 2010;127:2088-94.
- [118] Zhao S, Yang XF, Shen DF, Gao Y, Shi S, Wu JC, et al. The down-regulated ING5 expression in lung cancer: a potential target of gene therapy. *Oncotarget*. 2016;7:54596-615.
- [119] Zheng HC, Xia P, Xu XY, Takahashi H, Takano Y. The nuclear to cytoplasmic shift of ING5 protein during colorectal carcinogenesis with their distinct links to pathologic behaviors of carcinomas. *Hum Pathol*. 2011;42:424-33.
- [120] Xing YN, Yang X, Xu XY, Zheng Y, Xu HM, Takano Y, et al. The altered expression of ING5 protein is involved in gastric carcinogenesis and subsequent progression. *Hum Pathol*. 2011;42:25-35.
- [121] Rodriguez JA. Caracterización estructural del supresor tumoral ING5. Doctoral Thesis, Universidad del País Vasco/ Euskal Herriko Unibertsitatea. 2014.
- [122] Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA. 3D complex: a structural classification of protein complexes. *PLoS Comput Biol*. 2006;2:155.
- [123] Goodsell DS, Olson AJ. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct*. 2000;29:105-53.
- [124] Levy ED, Boeri Erba E, Robinson CV, Teichmann SA. Assembly reflects evolution of protein complexes. *Nature*. 2008;453:1262-5.
- [125] Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*. 2014;42:W320-W4.
- [126] Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. *Nature*. 2003;423:145-50.
- [127] Arents G, Moudrianakis EN. Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. *Proc Natl Acad Sci U S A*. 1993;90:10489-93.
- [128] Pingoud A, Jeltsch A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res*. 2001;29:3705-27.
- [129] Marianayagam NJ, Sunde M, Matthews JM. The power of two: protein dimerization in biology. *Trends Biochem Sci*. 2004;29:618-25.
- [130] Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci*. 2008;33:220-9.
- [131] Renucci M, Stennicke HR, Scott FL, Liddington RC, Salvesen GS. Dimer formation drives the activation of the cell death protease caspase 9. *Proc Natl Acad Sci U S A*. 2001;98:14250-5.
- [132] Jiang G, den Hertog J, Hunter T. Receptor-like protein tyrosine phosphatase alpha homodimerizes on the cell surface. *Mol Cell Biol*. 2000;20:5917-29.
- [133] Hebert TE, Bouvier M. Structural and functional aspects of G protein-coupled receptor oligomerization. *Biochem Cell Biol*. 1998;76:1-11.

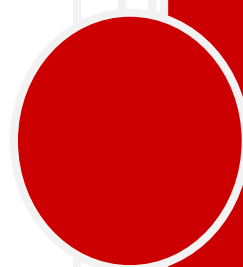
- [134] Culurgioni S. Structural and functional characterization of the dimerization domain of tumor suppressor ING4. Madrid: Autónoma de Madrid; 2010.
- [135] Joerger AC, Rajagopalan S, Natan E, Veprintsev DB, Robinson CV, Fersht AR. Structural evolution of p53, p63, and p73: Implication for heterotetramer formation. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106:17705-10.
- [136] Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res*. 2005;33:W382-8.
- [137] Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. A graphical interface for the FoldX forcefield. *Bioinformatics*. 2011;27:1711-2.
- [138] Rogov VV, Rozenknop A, Rogova NY, Lohr F, Tikole S, Jaravine V, et al. A universal expression tag for structural and functional studies of proteins. *Chembiochem*. 2012;13:959-63.
- [139] Avvakumov N, Lalonde ME, Saksouk N, Paquet E, Glass KC, Landry AJ, et al. Conserved molecular interactions within the HBO1 acetyltransferase complexes regulate cell proliferation. *Mol Cell Biol*. 2011;32:689-703.
- [140] Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003;31:3784-8.
- [141] Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005;41:207-34.
- [142] Tyler RC, Sreenath HK, Singh S, Aceti DJ, Bingman CA, Markley JL, et al. Auto-induction medium for the production of [U-15N]- and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination. *Protein Expr Purif*. 2005;40:268-78.
- [143] Marley J, Lu M, Bracken C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR*. 2001;20:71-5.
- [144] Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*. 2006;1:2876-90.
- [145] Greenfield N, Fasman GD. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*. 1969;8:4108-16.
- [146] Marky LA, Breslauer KJ. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*. 1987;26:1601-20.
- [147] Thomas F, Boyle AL, Burton AJ, Woolfson DN. A Set of de Novo Designed Parallel Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-nanomolar Regime. *Journal of the American Chemical Society*. 2013;135:5161-6.
- [148] Wyatt PJ. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*. 1993;272:1-40.
- [149] Folta-Stogniew E, Williams KR. Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *J Biomol Tech*. 1999;10:51-63.
- [150] Weber PC. [2] Overview of protein crystallization methods. *Methods Enzymol*. 1997;276:13-22.
- [151] Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*. 2007;40:191-285.
- [152] Powell HR. The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr*. 1999;55:1690-5.
- [153] Evans P. Scaling and assessment of data quality. *Acta Crystallogr D Biol* 2006;62:10.
- [154] Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr*. 2010;66:22-5.
- [155] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004;60:2126-32.

- [156] Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*. 2011;67:355-67.
- [157] Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol*. 1963;7:95-9.
- [158] Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010;66:12-21.
- [159] Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372:774-97.
- [160] Alva V, Syamala Devi DP, Sowdhamini R. COILCHECK: an interactive server for the analysis of interface regions in coiled coils. *Protein Pept Lett*. 2008;15:33-8.
- [161] Walshaw J, Woolfson DN. Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol*. 2001;307:1427-50.
- [162] Schrödinger. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. In: 1.5.0.4, editor. 1504 1.5.0.4 ed2010.
- [163] Mertens HDT, Svergun DI. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology*. 2010;172:128-41.
- [164] Pernot P, Round A, Barrett R, De Maria Antolinos A, Gobbo A, Gordon E, et al. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat*. 2013;20:660-4.
- [165] Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI. ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *Journal of Applied Crystallography*. 2007;40:s223-s8.
- [166] Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*. 2003;36:1277-82.
- [167] Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of applied crystallography*. 1992;25:495-503.
- [168] Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J*. 2001;80:2946-53.
- [169] Volkov VV, Svergun DI. Uniqueness of ab-initio shape determination in small-angle scattering. *J Appl Cryst*. 2003;36:860-4.
- [170] Svergun DI, Barberato C, Koch MHJ. CRY SOL a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Cryst*. 1995;28:768-73.
- [171] Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography*. 2001;34:33-41.
- [172] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605-12.
- [173] Sørensen OW. James Keeler. *Understanding NMR Spectroscopy. Magnetic Resonance in Chemistry*. 2006;44:820-.
- [174] Claridge TDW. *High-resolution NMR Techniques in Organic Chemistry*. 1999.
- [175] Harris RK, Becker ED, Cabral De Menezes SM, Granger P, Hoffman RE, Zilm KW, et al. Further conventions for NMR shielding and chemical shifts IUPAC recommendations 2008. *Solid State Nucl Magn Reson*. 2008;33:41-56.
- [176] Bodenhausen G, Ruben DJ. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*. 1980;69:185-9.
- [177] Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*. 2009;44:213-23.
- [178] Berjanskii MV, Wishart DS. A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc*. 2005;127:14970-1.

- [179] Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, et al. ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J Biomol NMR*. 1995;6:135-40.
- [180] Kazimierczuk K, Orekhov V. Non-uniform sampling: post-Fourier era of NMR data collection and processing. *Magnetic Resonance in Chemistry*. 2015;53:921-6.
- [181] Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, et al. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry*. 1994;33:5984-6003.
- [182] Renner C, Schleicher M, Moroder L, Holak TA. Practical aspects of the 2D ¹⁵N-¹H-NOE experiment. *J Biomol NMR*. 2002;23:23-33.
- [183] Jung YS, Zweckstetter M. Mars -- robust automatic backbone assignment of proteins. *J Biomol NMR*. 2004;30:11-23.
- [184] Kazimierczuk K, Orekhov VY. Accelerated NMR spectroscopy by using compressed sensing. *Angew Chem Int Ed Engl*. 2011;50:5556-9.
- [185] Goddar TK, DG. Sparky-NMR assignment and integration software. 2008.
- [186] Bolin KA, Hanson P, Wright SJ, Millhauser GL. An NMR investigation of the conformational effect of nitroxide spin labels on Ala-rich helical peptides. *J Magn Reson*. 1998;131:248-53.
- [187] Bomar MG, Pai MT, Tzeng SR, Li SS, Zhou P. Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase eta. *EMBO Rep*. 2007;8:247-51.
- [188] Berliner LJ, Grunwald J, Hankovszky HO, Hideg K. A novel reversible thiol-specific spin label: papain active site labeling and inhibition. *Anal Biochem*. 1982;119:450-5.
- [189] Lorenzi M, Sylvi L, Gerbaud G, Mileo E, Halgand F, Walburger A, et al. Conformational selection underlies recognition of a molybdoenzyme by its dedicated chaperone. *PLoS One*. 2012;7:e49523.
- [190] Bengoa-Vergniory N, Gorroño-Etxebarria I, González-Salazar I, Kypta RM. A Switch From Canonical to Noncanonical Wnt Signaling Mediates Early Differentiation of Human Neural Stem Cells. *STEM CELLS*. 2014;32:3196-208.
- [191] Pagano JM, Clingman CC, Ryder SP. Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. *RNA*. 2011;17:14-20.
- [192] Uversky VN. Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. *Methods Mol Biol*. 2012;896:179-94.
- [193] Ormazza G, Medagli B, Ibáñez de Opakua A, Rodríguez JA, Merino N, Villate M, et al. The tumor suppressor inhibitor of growth 4 binds double-stranded DNA through its disordered central region. *FEBS Letters*. 2017;591:425-32.
- [194] De Biasio A, de Opakua AI, Mortuza GB, Molina R, Cordeiro TN, Castillo F, et al. Structure of p15PAF-PCNA complex and implications for clamp sliding during DNA replication and repair. *Nature Communications*. 2015;6:6439.
- [195] Dutta K, Alexandrov A, Huang H, Pascal SM. pH-induced folding of an apoptotic coiled coil. *Protein Sci*. 2001;10:2531-40.
- [196] Doig AJ. Recent advances in helix-coil theory. *Biophysical Chemistry*. 2002;101-102:281-93.
- [197] Hagelueken G, Ward R, Naismith JH, Schiemann O. MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL. *Applied Magnetic Resonance*. 2012;42:377-91.
- [198] Zhang W-H, Poh A, Fanous AA, Eastman A. DNA damage-induced S phase arrest in human breast cancer depends on Chk1, but G2 arrest can occur independently of Chk1, Chk2 or MAPKAPK2. *Cell Cycle*. 2008;7:1668-77.
- [199] Schalch T, Duda S, Sargent DF, Richmond TJ. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*. 2005;436:138-41.
- [200] Vuzman D, Levy Y. Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol Biosyst*. 2012;8:47-57.

- [201] Weinberg RL, Veprintsev DB, Bycroft M, Fersht AR. Comparative binding of p53 to its promoter and DNA recognition elements. *J Mol Biol.* 2005;348:589-96.
- [202] Jung KC, Rhee HS, Park CH, Yang CH. Determination of the dissociation constants for recombinant c-Myc, Max, and DNA complexes: the inhibitory effect of linoleic acid on the DNA-binding step. *Biochem Biophys Res Commun.* 2005;334:269-75.
- [203] Ruthenburg AJ, Li H, Patel DJ, Allis CD. Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol.* 2007;8:983-94.
- [204] Tallen G, Riabowol K. DisorderING promotes epigenetic order. *FEBS Lett.* 2017;591:257-9.
- [205] Culurgioni S, Munoz IG, Moreno A, Palacios A, Villate M, Palmero I, et al. Crystal structure of inhibitor of growth 4 (ING4) dimerization domain reveals functional organization of ING family of chromatin-binding proteins. *J Biol Chem.* 2012;287:10876-84.
- [206] Hartwell L. Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell.* 71:543-6.
- [207] Weinert T. DNA Damage and Checkpoint Pathways: Molecular Anatomy and Interactions with Repair. *Cell.* 1998;94:555-8.
- [208] Gordon DJ, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer. *Nat Rev Genet.* 2012;13:189-203.
- [209] Zhang F, Baumer N, Rode M, Ji P, Zhang T, Berdel WE, et al. The inhibitor of growth protein 5 (ING5) depends on INCA1 as a co-factor for its antiproliferative effects. *PLoS One.* 2011;6:10.

APPENDIX



APPENDIX I

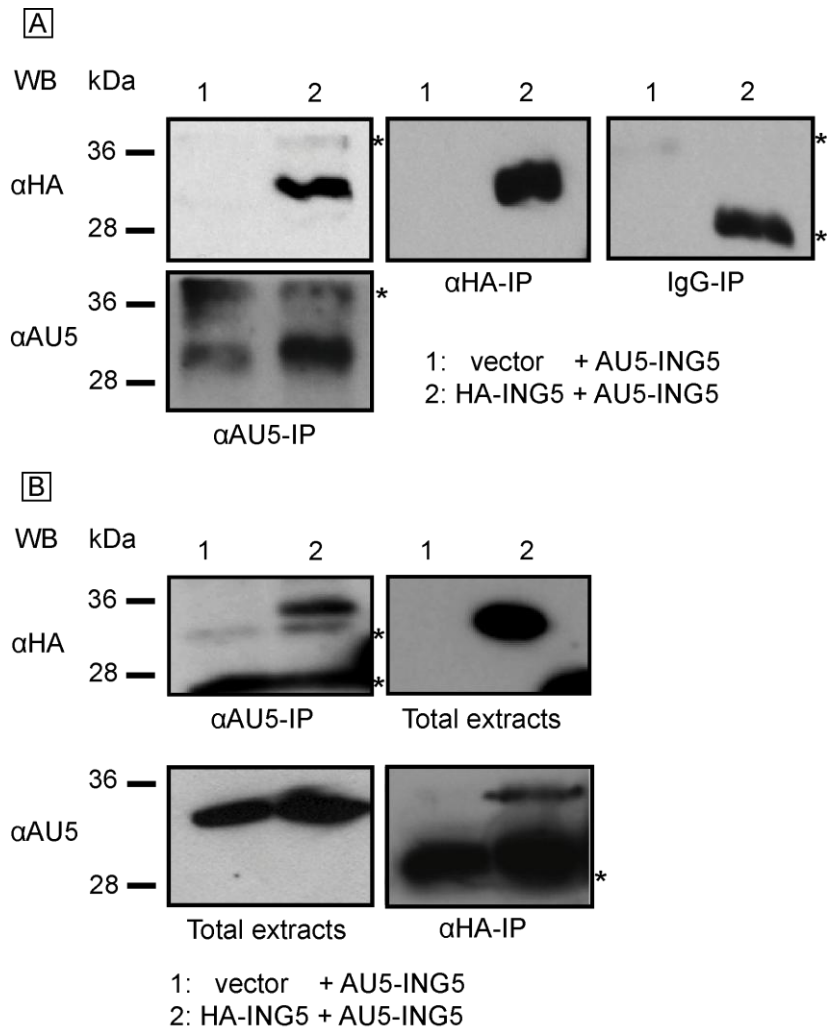


Figure I.1. Analysis of ING5 homodimerization inside cells by co-immunoprecipitation. Lysates from cells transiently transfected with vectors expressing AU5-ING4 (left lanes) or both AU5-ING4 and HA-ING5 (right lanes) were immunoprecipitated (IP) with α AU5 or α HA, antibodies against the AU5 tag and HA tag, respectively. The presence of each protein in the immunocomplexes was analyzed by Western blotting (WB) with antibodies against the HA (α HA, upper panels) or AU5 (α AU5, lower panels) tag. Panels A and B are reproducible experiments from the one showed in results, **figure 30**.

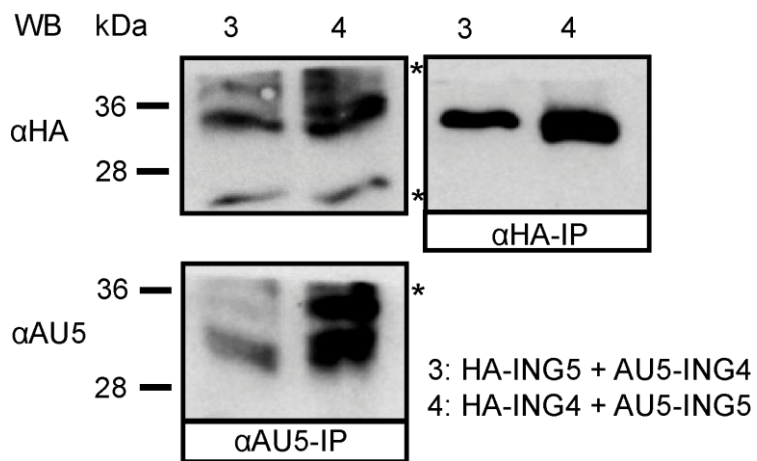


Figure I.2. Analysis of ING4/ING5 heterodimerization in cells by co-immunoprecipitation. Lysates from cells transiently transfected with vectors expressing HA-ING5 and AU5-ING4 (left lanes) or HA-ING4 and AU5-ING5 and (right lanes) were immunoprecipitated (IP) with α HA or α AU5, antibodies against the HA tag and AU5 tag, respectively. The presence of each protein in the immunocomplexes was analyzed by Western blotting (WB) with antibodies against HA (α HA, upper panels) or against AU5 (α AU5, lower panels) tag

APPENDIX II

Table I.1. Chemical shifts of the assigned nuclei of ING5 Nt₁₋₁₀₅ backbone and ¹³Cβ.

Residue nº	AA	¹³ Cα	¹³ Cβ	¹³ CO	¹ Hα	¹ HN	¹⁵ N
-1	GLY	-	-	-	-	-	-
0	ALA	68.748	-	177.87	4.266	-	-
1	MET	55.593	33.014	176.227	4.367	8.507	119.722
2	ALA	52.966	19.366	177.989	4.264	8.406	125.385
3	THR	62.217	69.668	174.464	4.164	8.028	112.689
4	ALA	52.871	19.101	177.529	4.189	8.211	125.382
5	MET	55.989	-	177.802	-	8.27	118.826
6	TYR	57.972	-	-	-	7.914	117.725
7	LEU	-	-	-	-	-	-
8	GLU	-	-	-	-	-	-
9	HIS	-	-	-	-	-	-
10	TYR	-	-	177.411	-	-	-
11	LEU	56.937	-	177.86	-	8.001	119.75
12	ASP	55.879	-	-	-	7.858	117.756
13	SER	-	-	-	-	-	-
14	ILE	-	-	-	-	-	-
15	GLU	-	-	-	-	-	-
16	ASN	53.252	-	174.793	-	-	-
17	LEU	53.216	-	-	-	7.954	121.754
18	PRO	-	-	-	-	-	-
19	CYS	-	-	176.787	-	-	-
20	GLU	59.843	-	178.323	-	9.645	120.197
21	LEU	57.36	-	177.965	-	7.094	118.221
22	GLN	60.049	-	178.866	-	7.91	118.465
23	ARG	58.895	-	178.34	4.025	8.148	118.421
24	ASN	56.255	-	-	4.241	7.572	120.293
25	PHE	59.039	-	178.845	-	8.185	118.6
26	GLN	59.105	-	178.788	-	8.016	120.151
27	LEU	57.768	-	180.021	-	8.464	121.17
28	MET	60.998	-	177.413	-	8.412	119.197
29	ARG	59.547	-	179.777	-	7.586	118.477
30	GLU	59.545	-	179.364	3.984	8.185	121.318
31	LEU	58.325	-	180.42	-	8.376	121.087
32	ASP	57.973	-	178.589	4.388	8.262	122.728
33	GLN	58.89	-	178.133	3.918	8.007	121.535
34	ARG	59.471	-	180.058	3.972	8.39	118.866
35	THR	67.933	-	175.863	-	8.194	117.191
36	GLU	59.625	-	179.949	3.938	8.227	122.615
37	ASP	57.382	40.092	179.269	4.408	8.649	120.809
38	LYS	56.569	-	178.248	-	8.079	121.852
39	LYS	60.519	-	178.751	-	8.635	120.54
40	ALA	55.038	-	180.412	4.185	7.89	121.327
41	GLU	59.454	-	179.027	-	7.732	119.85

42	ILE	66.504	-	176.75	-	8.347	120.289
43	ASP	57.981	-	179.51	4.432	7.823	118.769
44	ILE	64.825	-	0	-	7.546	121.102
45	LEU	57.768	-	180.839	4.052	8.449	121.105
46	ALA	55.685	-	179.027	3.871	9.487	123.241
47	ALA	55.351	-	180.843	-	7.732	119.85
48	GLU	59.687	-	178.866	3.813	8.134	119.326
49	TYR	61.832	-	176.417	-	8.345	121.59
50	ILE	63.793	-	177.418	-	8.581	117.49
51	SER	60.955	63.974	176.083	4.176	7.819	111.06
52	THR	69.86	-	176.927	-	7.519	109.157
53	VAL	65.864	-	175.743	-	7.793	123.365
54	LYS	59.21	-	177.504	-	8.427	119.116
55	THR	61.534	69.86	174.793	4.382	8.1	107.341
56	LEU	54.216	-	177.292	4.484	7.25	122.416
57	SER	56.808	-	-	-	8.915	119.545
58	PRO	66.877	-	178.843	-	-	-
59	ASP	57.053	40.072	179.252	4.334	8.442	114.703
60	GLN	58.402	-	179.226	4.127	7.709	120.311
61	ARG	61.165	-	178.198	-	8.567	120.481
62	VAL	66.664	32.028	178.38	3.581	7.717	118.139
63	GLU	59.67	-	179.488	3.98	7.642	119.926
64	ARG	58.51	-	179.385	3.958	8.108	117.608
65	LEU	58.013	-	179.104	-	8.18	120.193
66	GLN	58.847	28.005	178.694	3.964	8.308	118.996
67	LYS	59.899	-	180.296	3.963	7.751	119.23
68	ILE	66.492	-	177.176	-	7.769	121.821
69	GLN	60.13	-	179.164	3.756	8.62	120.358
70	ASN	55.89	38.154	177.689	4.369	8.751	117.845
71	ALA	55.517	-	180.38	4.22	7.864	124.304
72	TYR	63.695	-	179.758	-	8.537	119.139
73	SER	62.44	-	177.113	-	8.645	116.626
74	LYS	57.082	-	176.917	-	7.964	122.977
75	CYS	64.365	-	177.218	-	8.014	116.998
76	LYS	59.524	-	177.511	4.055	8.013	120.684
77	GLU	59.634	-	179.203	3.992	7.673	120.867
78	TYR	58.093	-	179.008	-	8.092	117.044
79	SER	61.867	-	177.089	-	8.273	114.606
80	ASP	57.565	-	179.585	-	8.872	123.905
81	ASP	57.385	40.086	179.131	4.372	8.504	122.184
82	LYS	60.063	-	177.878	-	8.479	121.999
83	VAL	67.112	-	177.404	-	8.067	119.004
84	GLN	59.03	-	178.914	-	7.634	117.879
85	LEU	57.409	-	180.235	3.982	8.004	120.104
86	ALA	56.015	-	178.744	-	9	126.405
87	MET	57.699	-	179.965	-	8.143	114.696
88	GLN	58.778	-	178.946	-	7.852	117.983
89	THR	67.893	-	176.42	4.103	8.313	118.261
90	TYR	63.138	-	178.755	-	8.582	120.377

91	GLU	59.998	-	178.693	-	7.802	119.391
92	MET	59.177	-	179.456	-	8.08	118.983
93	VAL	67.416	-	176.755	-	8.535	121.346
94	ASP	58.088	-	177.46	-	7.836	119.474
95	LYS	59.465	-	178.718	-	7.925	116.337
96	HIS	59.477	-	177.197	-	8.134	118.782
97	ILE	63.025	-	177.841	-	8.283	119.072
98	ARG	58.135	-	177.751	-	8.088	119.73
99	ARG	57.685	-	177.413	4.12	7.787	120.024
100	LEU	56.102	41.932	178.112	4.141	8.007	121.136
101	ASP	55.018	41.007	176.579	4.47	8.063	119.736
102	ALA	53.052	19.387	177.693	4.197	7.947	122.92
103	ASP	54.744	40.985	176.235	4.525	8.223	118.623
104	LEU	55.03	42.413	176.183	4.286	7.935	121.891
105	ALA	53.933	20.32	182.549	4.034	7.782	109.146

Table I.2. Chemical shifts of the assigned nuclei of the PHD finger of ING5 backbone and $^{13}\text{C}\beta$.

Residue n ^o	AA	$^{13}\text{C}\alpha$	$^{13}\text{C}\beta$	^{13}CO	$^1\text{H}\alpha$	^1HN	^{15}N
175	GLY	-	-	-	-	-	-
176	ALA	-	-	-	-	-	-
177	MET	55.461	-	175.898	4.479	-	-
178	ASP	54.425	41.072	175.676	4.582	8.164	120.95
179	MET	53.236	32.584	174.025	4.787	8.116	121.397
180	PRO	63.004	32.096	176.733	4.462	-	-
181	VAL	62.058	33.169	175.751	4.056	8.225	120.73
182	ASP	51.765	41.703	-	4.909	8.546	126.265
183	PRO	63.667	-	176.865	-	-	-
184	ASN	53.26	39.131	174.964	4.695	8.572	117.39
185	GLU	54.469	30.187	174.042	4.557	7.836	122.642
186	PRO	63.045	32.173	175.326	4.306	-	-
187	THR	59.659	71.323	173.015	4.13	7.55	109.179
188	TYR	57.158	43.624	174.218	-	8.282	118.424
189	CYS	59.431	30.335	175.158	3.676	7.136	115.547
190	LEU	56.834	42.666	177.924	4.342	9.965	126.242
191	CYS	57.902	31.184	175.89	4.624	7.619	114.276
192	HIS	56.986	26.901	173.962	4.372	7.652	119.967
193	GLN	54.253	33.077	176.218	5.073	8.1	117.113
194	VAL	61.597	33.171	176.139	4.563	8.19	113.529
195	SER	59.523	63.352	174.12	4.065	8.888	115.365
196	TYR	57.993	-	174.701	4.634	7.412	120.159
197	GLY	45.725	-	173.265	-	8.673	109.028
198	GLU	56.651	28.413	176.086	-	8.281	122.302
199	MET	53.063	34.577	175.17	5.505	8.322	124.114
200	ILE	58.434	42.532	172.088	4.763	9.69	121.9
201	GLY	43.302	-	172.927	-	8.352	116.26
202	CYS	60.381	32.337	176.956	4.552	8.446	126.286
203	ASP	56.736	40.921	176.611	4.681	8.986	127.028
204	ASN	50.992	38.702	175.066	5.185	8.899	123.368
205	PRO	64.433	32.062	176.774	4.416	-	-
206	ASP	53.367	41.176	175.02	4.79	7.243	115.259
207	CYS	59.335	31.289	176.321	4.052	7.705	126.633
208	PRO	63.908	-	176.737	4.593	-	-
209	ILE	63.162	40.333	177.636	4.145	8.575	124.099
210	GLU	63.147	30.148	176.022	3.743	9.56	120.291
211	TRP	56.122	33.373	174.836	5.362	7.943	120.478
212	PHE	56.205	42.825	176.143	-	9.201	117.238
213	HIS	57.928	30.829	178.16	5	9.41	122.788
214	PHE	60.869	37.854	178.544	3.951	8.44	124.375
215	ALA	55.186	18.445	181.063	4.358	9.006	116.148
216	CYS	62.323	30.645	176.654	4.282	7.437	116.982
217	VAL	60.01	31.235	173.818	4.664	7.38	108.156
218	ASP	55.739	39.427	174.773	4.308	7.656	115.955
219	LEU	53.782	44.581	177.299	4.671	7.886	117.711

220	THR	62.259	70.09	174.446	-	-	-
221	THR	60.158	71.043	172.573	4.55	7.465	114.943
222	LYS	55.275	33.055	174.628	4.016	8.662	126.102
223	PRO	62.9	-	176.289	-	-	-
224	LYS	56.606	32.775	177.345	4.314	8.538	121.442
225	GLY	44.377	-	173.245	-	8.092	110.382
226	LYS	56.277	33.809	176.443	4.294	8.197	120.499
227	TRP	59.038	31.284	172.123	4.344	9.677	126.516
228	PHE	54.096	41.904	173.333	5.071	6.91	123.785
229	CYS	57.139	30.894	172.974	3.536	8.769	125.978
230	PRO	65.514	32.238	179.319	4.108	-	-
231	ARG	58.23	29.805	178.687	4.22	8.105	117.618
232	CYS	64.538	29.361	178.449	3.957	8.868	126.114
233	VAL	65.814	31.683	177.547	3.425	8.554	120.146
234	GLN	57.872	28.609	177.845	4.008	7.486	118.661
235	GLU	57.999	29.938	177.801	4.044	7.905	119.339
236	LYS	57.231	32.185	177.418	4.086	7.814	119.051
237	ARG	56.786	30.67	176.523	-	7.742	119.339
238	LYS	56.734	-	176.449	-	7.847	120.995
239	LYS	56.438	33.052	175.588	4.295	8.115	122.89
240	LYS	57.882	33.679	181.364	4.128	7.939	128.219

APPENDIX III

Table II.1. Calculated dissociation constants (K_D) for ING5 binding H3K4me3 peptide. The maximum chemical shift perturbation for ^1H y ^{15}N (CSP_{max}, is a fitted value as the K_D) is indicated together with the fitting error.

residue	K_D	CSP _{max}	K_D error
182	7.65101	0.14011	1.49454
188	10.45822	0.12878	3.70238
194	5.22382	0.14579	1.87883
209	2.73544	0.11502	0.81456
210	10.77235	0.16492	2.37274
213	7.23825	0.15378	3.39289
215	6.39435	1.01951	0.10985
227	7.55817	2.35624	0.18192

Table II.2. Calculated dissociation constants (K_D) for PHD binding H3K4me3 peptide. The maximum chemical shift perturbation for ^1H y ^{15}N (CSP_{max}, is a fitted value as the K_D) is indicated together with the fitting error.

residue	K_D	CSP _{max}	K_D error
182	15.1875	0.12712	3.14088
188	15.8785	0.15373	1.72860
189	20.1679	0.15004	3.06208
194	19.4489	0.16597	2.28157
195	16.7430	0.12412	2.87740
209	14.5452	0.13729	3.24088
210	16.3068	0.15064	2.80955
211	22.2633	0.19292	3.42671
213	19.5692	0.16511	5.51463
227	18.6875	0.17360	2.64424

The tumor suppressor inhibitor of growth 4 binds double-stranded DNA through its disordered central region

Georgina Ormaza¹, Barbara Medagli², Alain Ibañez de Opakua¹, Jhon A. Rodriguez¹, Nekane Merino¹, Maider Villate¹, Silvia Onesti² and Francisco J. Blanco^{1,3}

1 CIC bioGUNE, Derio, Spain

2 Elettra Sincrotrone Trieste, Italy

3 IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Correspondence

F. J. Blanco, CIC bioGUNE, Parque Tecnológico de Bizkaia 800, 48160 Derio, Spain

Fax: +34 946572502

Tel: +34 946572521

E-mail: fblanco@cicbiogune.es

(Received 26 September 2016, revised 28 November 2016, accepted 30 November 2016, available online 22 December 2016)

doi:10.1002/1873-3468.12514

Edited by Ivan Sadowski

The tumor suppressor inhibitor of growth 4 (ING4) regulates chromatin structure by recruiting the histone acetyl transferase complex HBO1 to sites with histone H3 trimethylated at K4. ING4 dimerizes through its N-terminal domain and recognizes H3K4me3 by the C-terminal plant homeodomain (PHD). The central region of ING4 is disordered and contains the nuclear localization signal. Here, utilizing electrophoresis and nuclear magnetic resonance, we show that ING4 binds double-stranded DNA through its central region with micromolar affinity. Our findings suggest that the cooperativity arising from the presence of two DNA-binding regions in the ING4 dimer, as well as two H3K4me3-binding PHD fingers, may strengthen nucleosome binding and HBO1 complex recruitment.

Keywords: chromatin remodeling; DNA binding; electrophoretic mobility shift assay; inhibitor of growth 4; nuclear magnetic resonance; tumor suppressor

In eukaryotic organisms, the DNA is packed into chromatin, a highly regulated and dynamic structure, with the nucleosome as its fundamental unit. In the nucleosome, super helical turns of DNA wrap a core histone octamer. The structural organization of chromatin is also modulated by nonhistone proteins, and has a functional impact in DNA replication, repair, and transcription [1]. The main chromatin remodeling process is driven by covalent core histone modifications on their N-terminal tails, which can be recognized and modified by specific protein domains with different functions [2]. These modifications regulate the accessibility of the DNA and affect the structure of chromatin in a direct way or, indirectly, by recruiting chromatin remodeling complexes. Inhibitor of growth

4 (ING4) belongs to the ING family of tumor suppressors, composed of five homologous proteins [3], and is involved in the regulation of the transcriptional state of the chromatin by recruiting the histone acetyl transferase (HAT) complex HBO1 to sites with the H3K4me3 mark [4]. This modification is recognized by its conserved C-terminal plant homeodomain (PHD) with micromolar affinity [5–7]. The N-terminal region of ING4 is folded into a coiled-coil domain, forming an antiparallel dimer in solution and in living cells [8,9]. Therefore, ING4 can be described as an elongated dimer with two PHD fingers pointing to opposite directions and tethered by a disordered central region approximately 85 residues long (Fig. 1A). As this central region is rich in basic amino acids and

Abbreviations

CD, circular dichroism; CSP, chemical shift perturbation; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt; DTT, dithiothreitol; EMSA, electrophoretic mobility shift assay; HAT, histone acetyl transferase; HSQC, heteronuclear single quantum coherence; ING, inhibitor of growth; NLS, nuclear localization signal; NMR, nuclear magnetic resonance; PHD, plant homeodomain; SEC-MALS, size exclusion chromatography–multiangle light scattering; UV, ultraviolet; WT, wild-type.

A clone of full-length ING4 with the same Strep-tag inserted after the initial methionine was produced in the same way. The construct for the C-terminal PHD finger of ING4 (residues 188–249 with an extra methionine at the N-terminus), was previously described [6]. The synthetic gene of the ING4 central region deletion mutant (named ING4DNLS and lacking residues 106–187) was purchased from Eurofins Genomics (Ebersberg, Germany), and it was modified with the insertion of a His-tag and a TEV pro-tease site at the N-terminus. All constructs have been codon-optimized for expression in *Escherichia coli*.

Protein expression and purification

The proteins were produced in *E. coli* BL21 (DE3) cells grown in autoinduction medium [10]. Uniformly, ^{15}N -enriched ING4 protein was produced in a modified autoinduction medium [11] and in minimal media as described [12]. Cultures were harvested by centrifugation and resuspended in lysis buffer [20 mM Tris pH 8.0, 1 mM dithiothreitol (DTT)], with the addition of 150 and 300 mM NaCl in the case of Nt and ING4 Δ NLS, respectively, and in the presence of protease inhibitors (one tablet complete EDTA-free per 50 mL). After sonication and ultracentrifugation, proteins were predominantly found in the insoluble (ING4, PHD, and ING4DNLS) or in the soluble (Nt) fraction. Insoluble proteins were solubilized in lysis buffer with 8 M urea and separated by ultracentrifugation at 4 °C and 142 000 g for 3 h. Supernatant was refolded by a 1 : 10–1 : 100 dilution into cold 20 mM Tris pH 8.0, 1 mM DTT, and 50 μM ZnCl₂ (needed for PHD finger folding).

Refolded ING4 was loaded onto a HiLoad 26/10 Q Sepharose anion exchange column equilibrated in 20 mM Tris pH 8.0, 1 mM DTT, and elution was carried out with a 0–0.5 M NaCl gradient in 4.7 column volumes. Eluted fractions were diluted three times in 20 mM Tris pH 8.0, 1 mM DTT, and loaded onto a Hi-Trap SP FF column equilibrated in 20 mM Tris pH 8.0, 50 mM NaCl, and 1 mM DTT for cation exchange chromatography. Elution was done with a 0.05–1 M NaCl gradient in 20 CV. Selected fractions were concentrated and separated by gel filtration in a Superdex 75 26/60 column equilibrated in 20 mM Tris pH 8.0, 300 mM NaCl, and 1 mM DTT. The degree of purity of this sample is higher than 90% as illustrated in Fig. S1. Soluble Strep-ING4 was purified on a Streptactin 5 mL column (IBA-Lifesciences, Goettingen, Germany) equilibrated in 100 mM Tris-HCl, pH 8.0, 150, or 300 mM NaCl, and eluted with 2.5 mM desthiobiotin. Selected fractions were concentrated and loaded onto Superdex 75 16/60 equilibrated with 20 mM Tris-HCl pH 8.0, 300 mM NaCl, and 1 mM DTT. Soluble Strep-ING4 Nt was purified on a Streptactin 5 mL column in the same way as the full-length protein. Refolded PHD was purified as previously described [6]. Refolded His-ING4 Δ NLS was loaded into a

Hi-Trap FF crude 5 mL column equilibrated in 20 mM Tris, pH 8, 300 mM NaCl, 1 mM DTT, and washed with 50 mM imidazol. Elution was done with a 50–300 mM gradient in 20 CV. Selected fractions were diluted 1 : 2 in 20 mM Tris, pH 8.0, and 1 mM DTT, and loaded onto a HiTrap Q HP 5 mL column equilibrated with 20 mM Tris, pH 8, 30 mM NaCl, and 1 mM DTT. Protein was eluted with a 0.05–1 M NaCl gradient in 50 CV. Selected fractions were concentrated and loaded on a Superdex 200 26/60 column equilibrated in 20 mM Tris, pH 8, 300 mM NaCl, and 1 mM DTT. Pure proteins were concentrated by ultrafiltration, flash-frozen in liquid N₂, and stored at 80 °C until used. The identity and purity of each protein was confirmed by MALDI-TOF and SDS/PAGE. Protein concentration was measured by UV absorbance using extinction coefficients calculated using the ExPASy ProtParam tool [13].

Fluorescence electrophoretic mobility shift assay (EMSA)

Oligonucleotides (Table S1) were designed to generate single-strand (ss), double-strand (ds), or primed (p) DNA ligands labeled with [6-Carboxyfluorescein (6-FAM)] at the 5' end. To avoid any quenching effect, the dsDNA has two extra bases at the 5' end. Oligonucleotides were chemically synthesized and HPLC-purified by Thermo Fisher Scientific (Waltham, MA, USA) or Sigma-Aldrich (Saint Louis, MO, USA), and were solubilized in Tris-EDTA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) to a concentration of 100 μM . For the preparation of the different dsDNA, ligands labeled : unlabeled oligonucleotides (in a 1 : 1.2 molar ratio) were mixed and diluted in annealing buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) to the desired final concentration. Annealing was performed by incubation in boiling water for 5 min followed by slow cooling to room temperature. EMSA experiments were performed by incubating increasing concentrations of ING4 or the different domains of ING4 with fluorescent DNA ligands at a final concentration of 0.1 μM in a 15 μL reaction mixture containing 20 mM Tris-HCl pH 8.0, 5 mM MgCl₂, 50 mM NaCl, 2 mM DTT, and 0.1 g L⁻¹ BSA. After incubation for 10 min at room temperature, 5% glycerol was added and the reaction products were separated on a 6% native polyacrylamide gel run at room temperature in cold 0.5X Tris-Borate-EDTA buffer for 50 min at 80 V. Labeled nucleic acid fragments were detected by fluorescence imaging (ImageQuant LAS4000; GE Healthcare, Chicago, IL, USA) and quantification of protein–nucleic acid complexes was performed with IMAGEQUANT TL image analysis soft-ware (GE Healthcare). The apparent equilibrium dissociation constant (K_D) was determined using a Hill equation with a single-site binding model (as implemented in Prism, GRAPHPAD software, La Jolla, CA, USA) from the mean of three independent experiments.

Circular dichroism (CD)

Circular dichroism measurements were performed with a JASCO J-810 spectropolarimeter. The spectrum was recorded on a 35 μM protein sample in PBS (10 mM phosphate, 140 mM chloride, 153 mM sodium ion, and 4.5 mM potassium ion at pH 7.4) using a 0.1 cm path length quartz cuvette at 25 °C. Thermal denaturation from 5 to 95 °C was recorded on a 2 μM protein sample using a stoppered 2 mm path length cuvette by increasing temperature at a rate of 1 °C min^{-1} and measuring the change in ellipticity at 222 nm.

Size exclusion chromatography-multiangle light scattering (SEC-MALS)

Static light scattering experiments were performed at 25 °C using a Superdex 200 10/300 GL column (GE HealthCare) attached in-line to a DAWN-HELEOS light scattering detector and an Optilab rEX differential refractive index detector (Wyatt Technology, Goleta, CA, USA). The column was equilibrated with running buffer (PBS + 0.03% NaN_3 , 0.1 μm filtered) and the SEC-MALS system was calibrated with a sample of BSA at 1 g L^{-1} in the same buffer. Then, a 100 μL protein sample at 35 μM (0.7 g L^{-1}) in PBS was injected into the column at a flow rate of 0.5 mL min^{-1} . Data acquisition and analysis were performed using the ASTRA software (Wyatt Technology). Based on numerous measurements on BSA samples at 1 g L^{-1} under the same or similar conditions, we estimate that the experimental error in the molar mass is around 5%.

NMR spectroscopy

NMR experiments were recorded at 25 °C on a Bruker Avance III spectrometer operating at 18.8 T (800 MHz of ^1H Larmor frequency) equipped with a cryoprobe and z gradients. The spectrum of ING4DNLS was measured on a 27 μM sample in PBS pH 7.4. ING4 binding to dsDNA was investigated by adding increasing amounts of a concentrated stock of the DNA (642 μM in 20 mM MES pH 6.5, 100 mM NaCl, 5 mM MgCl_2 , and 1 mM DTT) to a 25 μM sample of U- ^{15}N -labeled ING4 in 400 μL of the same MES buffer with 5% $^2\text{H}_2\text{O}$ in a 5 mm Shigemi NMR tube. ^1H - ^{15}N -HSQC spectra were recorded (128 indirect points, 2 h total acquisition time) after each addition and the chemical shift of the observed signals was measured. To avoid resonance shifts due to different ionic strength or pH in the DNA and protein samples, the oligonucleotides were column-desalted into the MES buffer, annealed, and concentrated by ultrafiltration (3 kDa cut off), and the protein was dialyzed in the same batch of buffer used to prepare the DNA. The oligonucleotides used for the NMR experiment did not have any 6-FAM fluorescent probe. TOPSpin (Bruker) and Sparky (University of California) were used

for NMR data processing and spectral analysis, respectively. Chemical shifts were measured relative to internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) for ^1H and calculated for ^{15}N [14]. Dissociation constants (K_D) were determined by the combined ^1H and ^{15}N chemical shift perturbations (CSP) fitting as described [6] using Prism (GRAPHPAD software).

Results

We have previously studied the structure of human ING4 protein purified from inclusion bodies without any affinity purification tag [8]. When a construct with a Strep-tag at the N-terminus was used, a small amount of soluble full-length human ING4 could be isolated from lysed bacterial cells, but the protein copurified with DNA, as seen by ultraviolet absorbance and by native PAGE stained for protein or DNA (data not shown). This prompted us to identify the ING4 domains involved in the interaction as well as the binding determinants regarding structure and length of the DNA molecule. Because pure untagged ING4 was more soluble than pure Strep-ING4 (and could be obtained with higher yields), untagged refolded ING4 was used throughout this study.

ING4 weakly binds to an 18 bp DNA duplex (dsDNA18) and the binding is enhanced when the longer dsDNA32 substrate is used (Fig. 2). No binding can be detected for the isolated N-terminal or PHD domains. These results point to the central region as the binding site for the DNA, and this is confirmed by the deletion mutant ING4DNLS not binding to any of the two dsDNAs (Fig. 2). This mutant lacks residues 106–187, which encompass most of the central NLS region connecting the two folded domains of ING4. Biophysical characterization of the mutant shows a symmetric dimer with predominantly helical coiled-coil structure and C-terminal PHD fingers, as seen by SEC-MALS, CD and NMR (Fig. 3).

As ING4 is a dimer with two DNA-binding sites, an avidity effect may be expected in the presence of the long dsDNA32 that can reach both sites independently, in such a way that binding to the first site increases the likelihood for the second interaction to occur. The ING4 dimer is head-to-tail, with the two NLS DNA-binding regions pointing to opposite sides of the dimerization domain. The length of this domain is approximately 70 Å, whereas the length of an 18 bp dsDNA is only about 60 Å. The dimeric nature of ING4 makes it also possible that the long dsDNA32 binds more than one ING4 dimer. This may explain the second band shift seen at high protein concentrations with dsDNA32 in Fig. 2. The multiple modalities

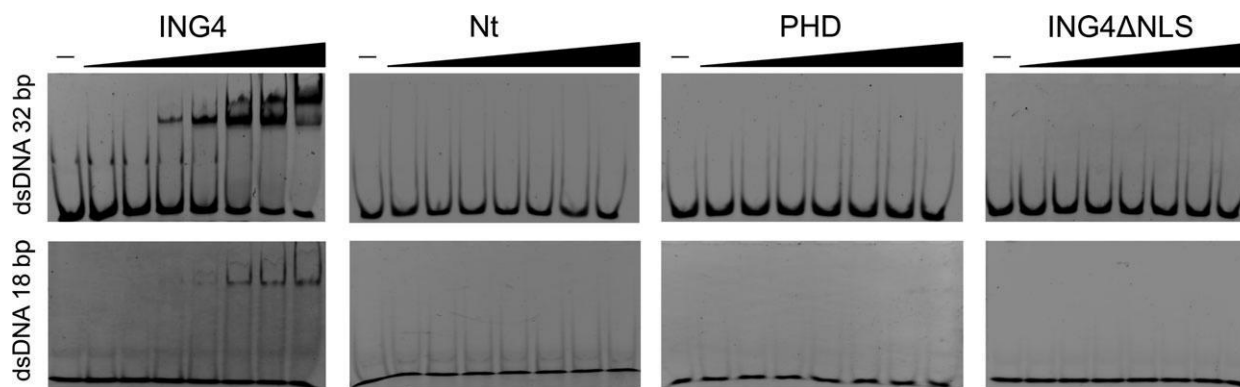


Fig. 2. Titration of dsDNA32 (top) and dsDNA18 (bottom) with increasing amounts of ING4 proteins and monitoring complex formation by EMSA. The concentration of DNA was 0.1 μM and the protein concentration was (lanes from left to right) 0, 0.05, 0.1, 0.25, 0.5, 0.75, 1, and 2 μM.

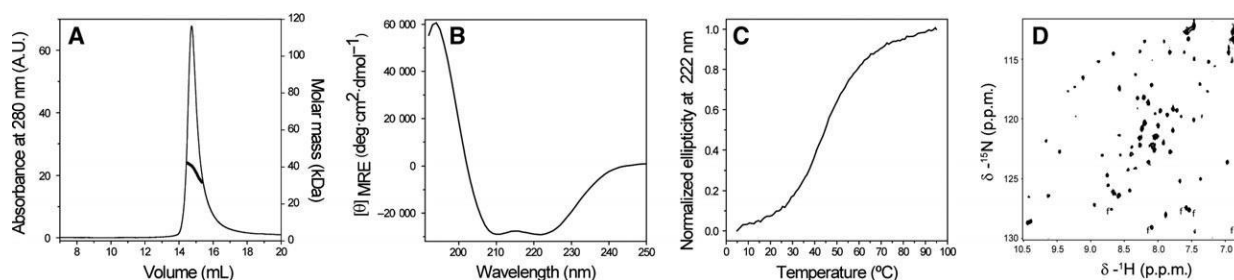


Fig. 3. Biophysical characterization of mutant ING4DNLS. (A) SEC-MALS analysis of the dimerization. The molar mass (thick line) at the center of the chromatography peak (thin line) is 40.7 kDa. The value calculated from the amino acid composition of the monomer is 19.5 kDa, indicating that the protein forms a dimer. (B) Secondary structure by far-UV circular dichroism. The ratio of the ellipticity at 222 and 208 nm is larger than 1, suggesting that it contains a large proportion of coiled-coil structure. (C) Tertiary structure analysis by thermal denaturation followed by the change in the CD signal at 222 nm, showing a midpoint melting temperature of 45 °C. (D) ¹H-¹⁵N HSQC NMR spectrum showing dispersed (PHD finger) and nondispersed (other parts of the protein) signals in the proton dimension. The signals labeled with 'f' are folded in the ¹⁵N dimension and their real chemical shift is 18.5 ppm smaller.

of binding make it difficult to quantify and interpret the affinity of the interaction. An attempt to do so using a Hill equation [15] yields an apparent dissociation constant of 0.6–0.03 μM (Fig. S2). ING4 also binds to single stranded DNA molecules (Fig. 4), but more weakly than it binds the corresponding duplexes (Fig. 2). ING4 also binds to a primed DNA (Fig. 4), but not as strongly as to dsDNA32. All these results indicate that ING4 has a preference for double-strand DNA binding.

Because of the flexible nature of the central region, ING4 yields an NMR spectrum where certain structural features can be recognized in spite of its large size (55 kDa dimer). In particular, the chemically equivalent C-terminal PHD finger gives a pattern of well resolved signals in the ¹H-¹⁵N HSQC NMR spectrum while the central region and the N-terminal domain mainly contribute signals that are non-dispersed in the proton dimension [8]. Most of the signals

observed at 25 °C belong either to the small PHD finger, which has been specifically assigned [6], or to the central disordered region, because the size and shape of the dimeric N-terminal domain make its NMR signals mostly unobservable unless temperature is increased [8]. Therefore, we can use NMR to observe the interaction with the dsDNA32 directly in solution (Fig. 5), a complementary experiment to EMSA. As shown in Fig. 5, the backbone amide signals that experience CSP are in the central region of the proton frequency dimension. In this region of the NMR spectrum of ING4, there are five signals that are absent in the spectrum of ING4DNLS and that have ¹⁵N chemical shifts typical of glycine residues. These signals are highlighted with a dashed rectangle in the spectra of Fig. 5, and very likely correspond to the 5 glycine residues present in the NLS region (Fig. 1B). Three of them experience relatively large shifts and the other two shift very little. This is consistent with three

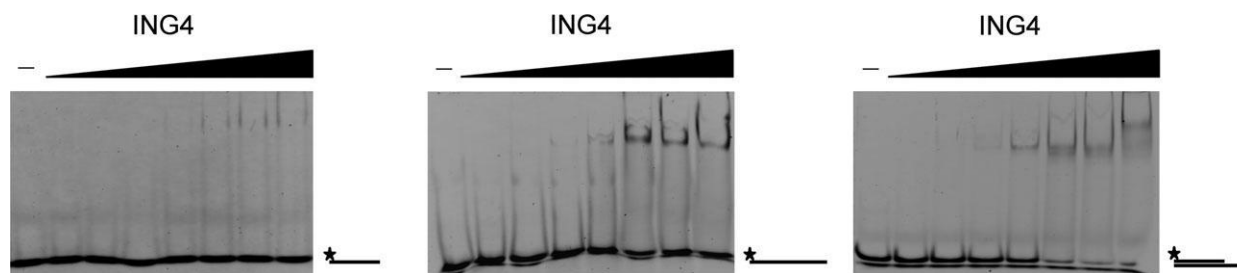


Fig. 4. Titration of fluorescent DNA molecules with increasing amounts of ING4 and monitoring complex formation by EMSA. From left to right: 20 base-long ssDNA, 34 base-long ssDNA, and the corresponding primed DNA. The concentration of DNA was 0.1 μ M in the three experiments, and the protein concentration was (lanes from left to right) 0, 0.05, 0.1, 0.25, 0.5, 0.75, 1, and 2 μ M. A scheme of the structure of the DNA probe used is drawn at the right hand side of the gels, with the asterisk indicating the position of the fluorescent tag.

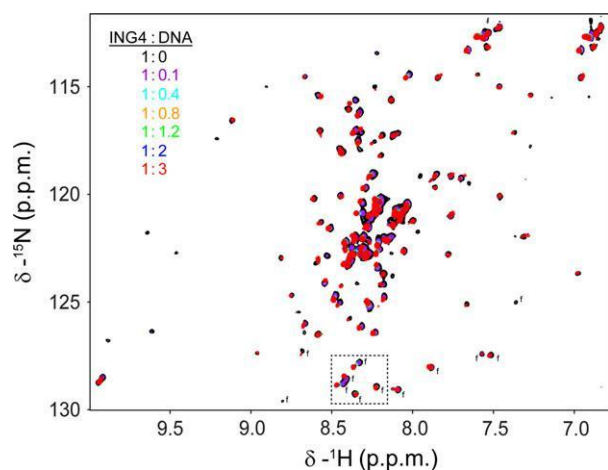


Fig. 5. Overlay of NMR spectra of ING4 in the presence of increasing amounts of dsDNA32 as indicated by the color code (molar ratio ING4 protomer: dsDNA32). The concentration of the sample was 25 μ M and spectra were recorded at 25 °C. The signals labeled with 'f' are folded in the 15 N dimension and their real chemical shift is 18.5 p.p.m. smaller. The dashed rectangle indicates the region expanded in Fig. S3.

of the NLS glycines being located closed to Lys and Arg clusters in the central region of the NLS and two of the glycines being in the C-terminal end of the NLS, with no positively charged residues (Fig. 1B). The measured CSP reach a plateau for all residues at ratios above 1 : 2 (Fig. S3). Fitting the data of the 13 resonances experiencing the largest relative changes to a simple model of one site binding model (one DNA duplex bound to one ING4 protomer) yield apparent dissociation constants in the 0.6–3.8 μ M range. Since the stoichiometry of the complex is unknown, these numbers are only indicative of a low micromolar apparent affinity, similar to the value derived from the EMSA for the same DNA molecule. A further compounding issue is that the first addition of DNA caused a slight protein precipitation in the NMR tube,

which was partially reversed at high ING4 : dsDNA32 ratios. At these high ratios, the CSP change very little, indicating that the protein is saturated with DNA. However, the shifted signals remain sharp and poorly dispersed in the proton dimension, indicating that in the DNA-bound form the backbone of the NLS region remains, at least partially, disordered. This observation is consistent with a weak binding based on electro-static interactions, mediated by the positively charged side chains of the NLS and the phosphate backbone of the DNA, as has been reported for other disordered proteins rich in positively charged residues [16,17]. Therefore, our results indicate that the interaction is primarily electrostatic, favoring long duplexes, and not particular nucleotide sequences. In agreement with this, a 22 bp duplex with an unrelated sequence binds ING4 but more weakly than dsDNA32 (Fig. S4), and a short 10 bp duplex still binds to ING4 but much more weakly (Fig. S5). These observations indicate that the major determinant of DNA binding is not the nucleotide sequence but the length of the duplex, although we cannot exclude that there may be certain sequence preferences.

Discussion

Intrinsically disordered proteins (IDP) and intrinsically disordered regions (collectively referred to as IDP) are abundant in cells, and their frequency rises as the complexity of the organism increases [18]. Many IDP have functional roles [19], frequently related to molecular recognition, and often coupled with disorder-to-order transitions [20]. However, disordered proteins may also remain partly disordered, in the bound state [21]. Many DNA-binding proteins have disordered regions [22] which may interact with the DNA through multiple segments with relatively low DNA affinity but collectively contributing to avidity for the DNA.

ING proteins have been described as tumor suppressors that recognize the histone H3K4me3 modification through their C-terminal PHD fingers [23]. In this way, they bind the nucleosome and recruit HAT or histone deacetylase complexes to chromatin sites enriched in the H3K4me3 mark [4]. The possibility that ING proteins contribute to nucleosome binding through direct protein binding to DNA has not been explored, but our work shows that the central disordered NLS region of ING4 directly binds DNA substrates with different nucleotide sequences, with a preference for double-stranded DNA, and an affinity in the low micromolar range. This affinity is three orders of magnitude lower than the low nanomolar range measured for transcription factors like p53 [24] or c-myc [25], and the same as reported for the lysine specific histone demethylase 1 [26]. Our results are consistent with ING4 being a reader of the H3K4me3 mark and not a transcription factor recognizing a specific DNA sequence. However, two isoforms of ING4 lacking the PHD finger have been described [27], and it is possible that the DNA-binding activity of the NLS region plays a role other than strengthening nucleosome binding.

Our findings on ING4 can be probably extrapolated to other ING proteins, because an alignment of their NLS regions shows several clusters of positively charged residues (Fig. 1B), perhaps with the exception of ING3 whose central region has much less positively charged residues than the others [3,28]. It has been proposed that ING1 binds an AT-motif in the enhancer region of the *a-fetoprotein* promoter, repressing transcription through effects on p53 [29]. This observation is consistent with the ING proteins directly inter-acting with DNA.

Our results point to a novel role of the central disordered NLS region of ING proteins as modulator of the interaction with the nucleosome (and their associated HAT and HDAC complexes). The cooperativity arising from the presence of two H3K4me3 binding PHD fingers and two DNA-binding regions in the ING4 dimer will result in strong nucleosome binding [30] and therefore efficient HBO1 complex recruitment.

Acknowledgements

We thank Alfredo de Biasio for helpful discussions on the experimental design. This work was supported by the Ministerio de Economía y Competitividad grant CTQ2014-56966-R to FJB, BES-2012-052851 FPI contract, and EEBB-I-16-11186 grant to GO, and by the Associazione Italiana per la Ricerca sul Cancro (IG14718) to BM and SO.

Author contributions

FJB designed the study, NM and MV purified the proteins, GO, BM, AIdO performed experiments and analyzed data, SO supervised the study, and FJB wrote the manuscript with contributions from GO, BM and SO.

References

- 1 Fischle W, Wang Y and Allis CD (2003) Histone and chromatin cross-talk. *Curr Opin Cell Biol* 15, 172–183.
- 2 Taverna SD, Li H, Ruthenburg AJ, Allis CD and Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* 14, 1025–1040.
- 3 Russell M, Berardi P, Gong W and Riabowol K (2006) Grow-ING, Age-ING and Die-ING: ING proteins link cancer, senescence and apoptosis. *Exp Cell Res* 312, 951–961.
- 4 Doyon Y, Cayrou C, Ullah M, Landry AJ, Cote V, Selleck W, Lane WS, Tan S, Yang XJ and Cote J (2006) ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol Cell* 21, 51–64.
- 5 Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Pena P, Lan F, Kaadige MR et al. (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 442, 96–99.
- 6 Palacios A, Garcia P, Padro D, Lopez-Hernandez E, Martin I and Blanco FJ (2006) Solution structure and NMR characterization of the binding to methylated histone tails of the plant homeodomain finger of the tumour suppressor ING4. *FEBS Lett* 580, 6903–6908.
- 7 Palacios A, Munoz IG, Pantoja-Uceda D, Marcaida MJ, Torres D, Martin-Garcia JM, Luque I, Montoya G and Blanco FJ (2008) Molecular basis of histone H3K4me3 recognition by ING4. *J Biol Chem* 283, 15956–15964.
- 8 Palacios A, Moreno A, Oliveira BL, Rivera T, Prieto J, Garcia P, Fernandez-Fernandez MR, Bernado P, Palmero I and Blanco FJ (2010) The dimeric structure and the bivalent recognition of H3K4me3 by the tumor suppressor ING4 suggests a mechanism for enhanced targeting of the HBO1 complex to chromatin. *J Mol Biol* 396, 1117–1127.
- 9 Culurgioni S, Munoz IG, Moreno A, Palacios A, Villate M, Palmero I, Montoya G and Blanco FJ (2012) Crystal structure of inhibitor of growth 4 (ING4) dimerization domain reveals functional organization of ING family of chromatin-binding proteins. *J Biol Chem* 287, 10876–10884.
- 10 Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41, 207–234.

- 11 Tyler RC, Sreenath HK, Singh S, Aceti DJ, Bingman CA, Markley JL and Fox BG (2005) Auto-induction medium for the production of [U-15N]- and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination. *Protein Expr Purif* 40, 268–278.
- 12 Marley J, Lu M and Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* 20, 71–75.
- 13 Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD and Bairoch A (2003) ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31, 3784–3788.
- 14 Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL and Sykes BD (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6, 135–140.
- 15 Pagano JM, Clingman CC and Ryder SP (2011) Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. *RNA* 17, 14–20.
- 16 De Biasio A, Ibanez-de Opakua A, Cordeiro TN, Villate M, Merino N, Sibille N, Lelli M, Diercks T, Bernado P and Blanco FJ (2014) p15(PAF) is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. *Biophys J* 106, 865–874.
- 17 De Biasio A, de Opakua AI, Mortuza GB, Molina R, Cordeiro TN, Castillo F, Villate M, Merino N, Delgado S, Gil-Carton D et al. (2015) Structure of p15 (PAF)-PCNA complex and implications for clamp sliding during DNA replication and repair. *Nat Commun* 6, 6439.
- 18 Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF and Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337, 635–645.
- 19 Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM and Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41, 6573–6582.
- 20 Dyson HJ and Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12, 54–60.
- 21 Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC and Asturias FJ (2008) Malleable machines take shape in eukaryotic transcriptional regulation. *Nat Chem Biol* 4, 728–737.
- 22 Vuzman D and Levy Y (2012) Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol BioSyst* 8, 47–57.
- 23 Tallen G and Riabowol K (2014) Keep-ING balance: tumor suppression by epigenetic regulation. *FEBS Lett* 588, 2728–2742.
- 24 Weinberg RL, Veprintsev DB, Bycroft M and Fersht AR (2005) Comparative binding of p53 to its promoter and DNA recognition elements. *J Mol Biol* 348, 589–596.
- 25 Jung KC, Rhee HS, Park CH and Yang CH (2005) Determination of the dissociation constants for recombinant c-Myc, Max, and DNA complexes: the inhibitory effect of linoleic acid on the DNA-binding step. *Biochem Biophys Res Commun* 334, 269–275.
- 26 Hirschi A, Martin WJ, Luka Z, Loukachevitch LV and Reiter NJ (2016) G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme. *RNA* 22, 1250–1260.
- 27 Raho G, Miranda C, Tamborini E, Pierotti MA and Greco A (2007) Detection of novel mRNA splice variants of human ING4 tumor suppressor gene. *Oncogene* 26, 5247–5257.
- 28 He GH, Helbing CC, Wagner MJ, Sensen CW and Riabowol K (2005) Phylogenetic analysis of the ING family of PHD finger proteins. *Mol Biol Evol* 22, 104–116.
- 29 Kataoka H, Bonnefin P, Vieyra D, Feng X, Hara Y, Miura Y, Joh T, Nakabayashi H, Vaziri H, Harris CC et al. (2003) ING1 represses transcription by direct DNA binding and through effects on p53. *Cancer Res* 63, 5785–5792.
- 30 Ruthenburg AJ, Li H, Patel DJ and Allis CD (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* 8, 983–994.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Fig. S1. SDS/PAGE purity analysis of ING4.

Fig. S2. Quantitative analysis by gel densitometry of the binding of ING4 to fluorescent dsDNA32 measured by EMSA.

Fig. S3. (Left) Zoom of the overlaid NMR spectra of ING4 shown in Fig. 5, which contains the backbone amide signals of five residues tentatively assigned to the five glycine residues of the NLS region, three of them experiencing large CSP in the presence of saturating amounts of dsDNA32. (Right) Plot of the CSP of the 13 ING4 backbone amide signals with values larger than 0.015 p.p.m.

Fig. S4. Titration of dsDNA22 with increasing amounts of ING4 protein and monitoring complex formation by EMSA.

Fig. S5. (Left) Zoom of the overlaid NMR spectra of ING4 in the presence of the indicated molar ratios of dsDNA10, with the same backbone amide signals shown in supplementary Fig. 3. (Right) Plot of the CSP of the six ING4 backbone amide signals with values larger than 0.005 p.p.m.

Table S1. Oligonucleotides (5′ to 3′ sequences) used in this study.

Supplementary material

The tumor suppressor ING4 binds double stranded DNA with micromolar affinity through its disordered central region.

Georgina Ormaza¹, Barbara Medagli², Alain Ibáñez de Opakua¹, John A. Rodríguez¹,
Nekane Merino¹, Mainer Villate¹, Silvia Onesti² and Francisco J. Blanco^{1,3}

¹CIC bioGUNE, Derio, Spain; ²Elettra Sincrotrone Trieste, Italy, ³IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.

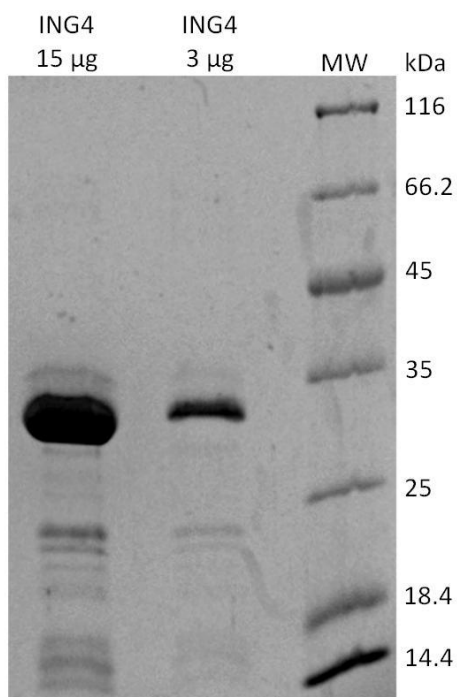
1 Supplementary Table

5 Supplementary figures

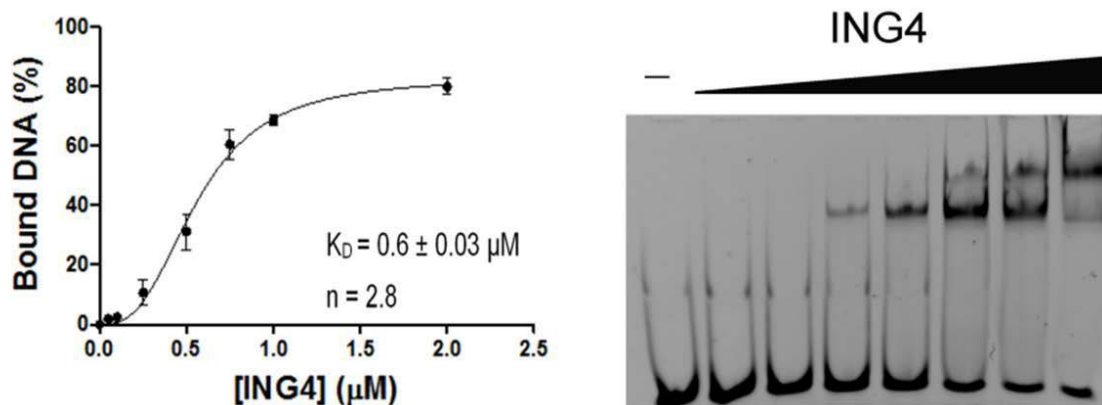
Supplementary table 1. Oligonucleotides (5' to 3' sequences) used in this study.

Name	Sequence
A1	TGGCCTGCAGGCATGCAA
A2	GTCCGTACGTTCTGGCCTGCAGGCATGCAA
A3	GATGAGATTGAGGCTGGCTGGCCTGCAGGCATGCAA
B1	6-FAM-GCTTGCATGCCTGCAGGCCA
B2	6-FAM-GCTTGCATGCCTGCAGGCCAGCCTCAATCTCATC
C1	GAGTGTGGTGTACATGCACTAC
C2	6-FAM-GTAGTGCATGTACACCACACTC
D1	ATACGATGGG
D2	CCCATCGTAT

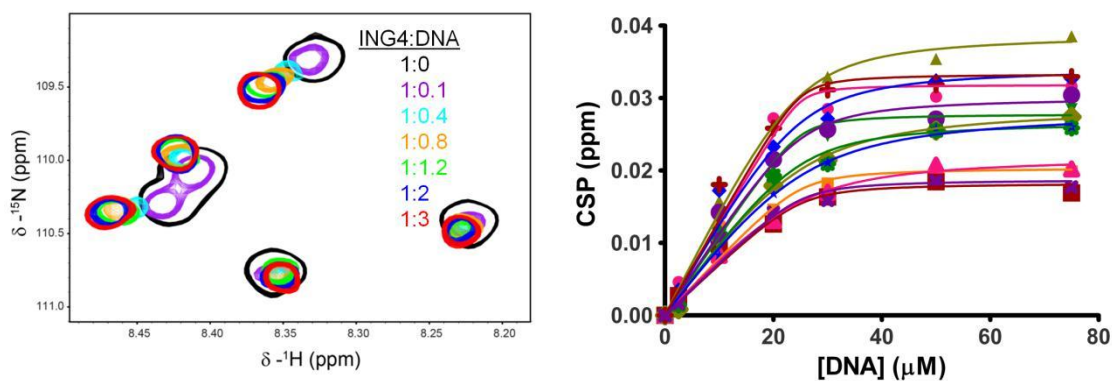
The dsDNA substrates with 18 bp, 32, 22, and 10 bp were obtained by annealing oligonucleotides A1 with B1, A3 with B2, C1 with C2, and D1 with D2, respectively. The pDNA substrate contained an 18 bp long dsDNA region and a 12 base long ssDNA region, and was obtained by annealing oligonucleotides A2 and B1.



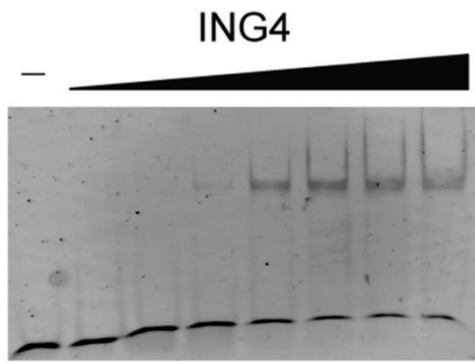
Supplementary figure 1. SDS-PAGE. A 12 % gel was loaded with 15 (left lane) or 3 (right lane) µg of ING4 protein and run at room temperature and 220 volts for 45 min. The gel was stained with coomassie brilliant blue and destained with 30% acetic acid in ethanol. Densitometry analysis of the gel indicates that the major, corresponding to ING4, represents between 90 (left lane) and 94 % (middle lane) of the total protein.



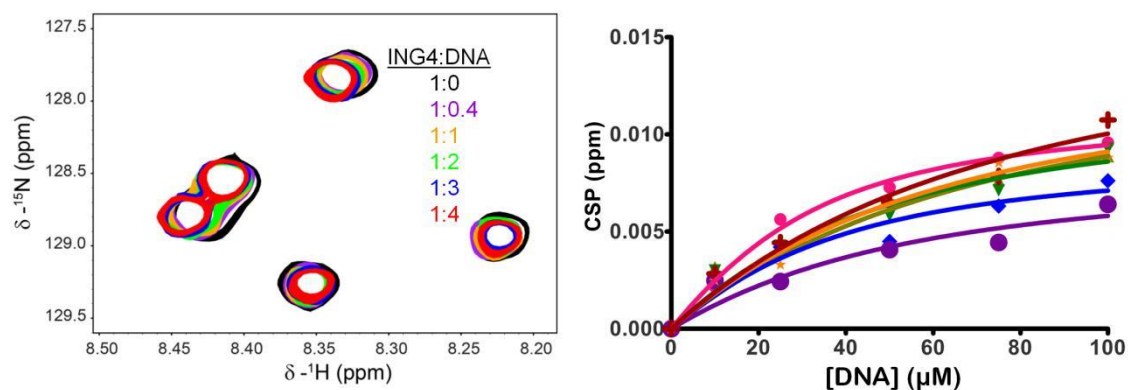
Supplementary figure 2. Quantitative analysis by gel densitometry of the binding of ING4 to fluorescent dsDNA32 measured by EMSA. The symbols and error bars are the mean and standard deviations of three independent experiments. The gel shown on the right corresponds to one of these three experiments. The curves are the fitting to a Hill equation corresponding to a single-site binding model as implemented in Prism (GraphPad software). The apparent equilibrium dissociation constant (K_D) and cooperativity (n) are indicated.



Supplementary figure 3. (Left) Zoom of the overlaid NMR spectra of ING4 shown in figure 5, which contains the backbone amide signals of 5 residues tentatively assigned to the 5 glycine residues of the NLS region, three of them experiencing large CSP in the presence of saturating amounts of dsDNA32. (Right) Plot of the CSP of the thirteen ING4 backbone amide signals with values larger than 0.015 ppm. The lines correspond to the fittings to a one site binding model, as described in the text.



Supplementary figure 4. Titration of dsDNA22 with increasing amounts of ING4 protein and monitoring complex formation by EMSA. The concentration of DNA was 0.1 μM and the protein concentration was (lanes from left to right) 0, 0.05, 0.1, 0.25, 0.5, 0.75, 1 and 2 μM .



Supplementary figure 5. (Left) Zoom of the overlaid NMR spectra of ING4 in the presence of the indicated molar ratios of dsDNA10, with the same backbone amide signals shown in supplementary figure 3. (Right) Plot of the CSP of the six ING4 backbone amide signals with values larger than 0.005 ppm. The lines correspond to the fittings to a one site binding model, and yield dissociation constants in the 17 - 58 μM range.