UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO
Scienze Biochimiche

TESI DI DOTTORATO DI RICERCA
svolta presso Fondazione Centro San Raffaele
Biomolecular NMR Laboratory, Milano, Italia

XXVII ciclo

Structural insights into the interaction between the tandem PHD finger domain P5C5 of NSD1 and the zinc finger motif C2HR of Nizp1

Dott. Andrea Berardi
Matr. R09716

Docente Guida: Prof. Francesco Bonomi
Supervisore: Dott.ssa Giovanna Musco

INDEX

4. Results

5. Discussion

Abstract

Point Mutations or translocation in NSD1 cause the overgrowth disorder Sotos syndrome and acute myeloid leukemia (AML), respectively (Berdasco M. *et al,* 2009, Wang G *et al*, 2003). NSD1 contains several chromatin related domains including a SET domain responsible for histone methyltranferases activity (H3K36 and H4K20), two nuclear receptor-interaction (NID) motifs, five zinc finger domains (PHD1-5), a variant PHD finger (C5HCH), two proline-tryptophan-proline-tryptophan (PWWP1-2) domains (Lucio-Eterovic AK, *et al* , 2011), suggesting a role in chromatin regulation and gene expression. 20 pathological Sotos mutations have been detected on the PHD tandem domain composed by NSD1-PHD5 and NSD1-C5HCH (NSD1-P5C5). The tandem domain is essential for the pathogenesis of acute myeloid leukaemia (AML) caused by the chimeric protein NUP98/NSD1 that forces the abnormal activation of Hox-A and Meis1 genes (Wang at al 2007). The deletion of this tandem domain is sufficient to abolish NUP98/NSD1 interaction with chromatin, preventing both the transcription activation of *HOX* genes and the immortalization of myeloid progenitors. The biological role of NSD1-P5C5 is still unclear. It was proposed that this tandem domain is involved in the recognition of both H3K4me3 and H3K9me3 histone marks, (Pasillas M et al. 2011). However, biophysical experiments in our laboratory did not confirm these results challenging the idea that this tandem domain can really work as epigenetic reader. Previous biochemical studies suggested that NSD1-P5C5 can also work as protein-protein interaction motif, being able to bind to the co-repressor Nizp1 by its C2HR zinc fingers motif (Nizp1-C2HR) thus mediating gene repression (Nielsen AL et al, 2004).

The structural determinants of this interaction are still unknown and have been object of this thesis. In order to get more insights into the physiological and pathological role of NSD1-P5C5, we have solved its (i) solution structure by NMR spectroscopy and (ii) characterized its interaction with Nizp1-C2HR.

NSD1-P5C5 folds as unique functional unit adopting a “face to side orientation”. In

particular NSD1-PHD5 (or NSD1-P5) presents the canonical PHD finger fold, whereas the NSD1-C5HCH (or NSD1-C5) domain displays an atypical topology characterized by the presence of an additional two stranded β-sheet. In order to investigate the impact of Sotos point mutation on NSD1-P5C5 we expressed and purified seven mutants and analyzed them by NMR. The majority of them destabilize the fold, with the exception of the solvent exposed mutation Arg2152Gln and His2205Arg suggesting a functional role for these residues.

We next solved the solution structure of the zinc finger Nizp1-C2HR, an atypical Cys2His2-type zinc finger in which the fourth zinc chelating residue is substituted by an arginine residue. Its fold consists of a short α-helix and of a short two-stranded β-sheet hold together by one zinc ion. Importantly, we showed that three zinc ligands are sufficient to maintain the protein domain fold and functionality.

NMR titrations of ^{15}N labelled NSD1-P5C5 with Nizp1-C2HR and ^{15}N labelled Nizp1-C2HR with NSD1-P5C5 clearly show that the two proteins directly interact. Analysis of the chemical shift displacements upon complex formation allowed to identify the residues of the two protein domains involved in protein-protein interaction. The interaction surface is located on the interface between NSD1-P5 and NSD1-C5 and on the α-helix of Nizp1-C2HR, respectively.

Based on this information using the software HADDOCK we have computed a data driven docking model of the protein complex. In the model Nizp1-C2HR places its α-helix in the groove at the interface between NSD1-P5 and NSD1-C5, creating both hydrophobic and polar intermolecular contacts.

The thermodynamic parameters that govern complex formation were studied by ITC titrations: the binding reaction is entropy-driven, with a stoichiometry of 1:1 and a K_d of 3,80±0,66 μM. In order to solve the structure of the protein complex we performed crystallographic screenings, and we have found preliminary conditions for obtaining single crystals.

In conclusion, the presented results provide novel information on the interaction between a tandem PHD finger domain and zinc finger motif. The results represent, to the best of our knowledge, the first biophysical characterization between two zinc binding domains. Most importantly, these data give the first molecular details of the interaction between NSD1 and Nizp1 and may provide useful insights into the function of NSD1 and its role in pathological conditions both in Sotos Syndrome and AML. Future work will be dedicated to the full three-dimensional characterization of the complex and to the analysis of Sotos mutations on complex formation.

1.Introduction

1.1 Epigenetic

In the eukaryotic organism, all cells inherit the same genetic material, and their ability to maintain specific physical characteristics and biological functions of tissues and organs is due to heritable differences in the packaging of DNA and chromatin (Adwan and Zawia 2013). These differences dictate distinct cellular gene expression programs but do not involve changes in the underlying DNA sequence of the organism (Arrowsmith et al. 2012). Epigenetics (which literally means 'above genetics') underpins the fundamental basis of human physiology and it deals with acquired and heritable modifications on DNA and histone residues that regulate the expression and functions of genes (Patel and Wang 2013).

These modifications include DNA methylation and hydroxymethylation, histone modifications and non-coding RNA regulation. Histone modifications consist of acetylation, methylation, crotonylation, ubiquitination, sumoylation, phosphorylation, hydroxylation and proline isomerization (Davie and Spencer 1999; Tony Kouzarides 2007; Peterson and Laniel 2004). Epigenetic changes are dynamic and they are responsible for orchestrating the myriad transcriptional units, condensed chromatin clusters and other features that distinguish the various cell types and the states of development or disease that share a genome within an individual.

Epigenetics is of therapeutic relevance in multiple diseases such as cancer, inflammation, metabolic disease and neuropsychiatric disorders. The dynamic nature of epigenetics means that it may be possible to alter disease-associated epigenetic states through direct manipulation of the molecular factors involved in this process (Adwan and Zawia 2013; Jakovcevski and Akbarian 2012).

1.2 Chromatin organization

The epigenetic changes alter chromatin organization (Fig 1). The basic unit of chromatin is the nucleosome, which consists of 146 base pairs (bp) of genomic DNA wrapped around an octamer composed of the core histones H2A, H2B, H3 and H4, with the histone tails containing post-translational modifications (PTMs) (Baxevanis and Landsman 1996).

Histone post-translational modification is mediated by two antagonizing sets of enzymatic complexes: the "writer"(such as methyltranferase and acetyltranferase enzyme) and "eraser" (such as histone deacetylases and lysine demethylases), proteins and associated factors that site-specifically attach and remove the modifications, respectively. Usually these modifications occur on

the N-terminal tails of the histones H3 and H4, that protrude out of the nucleosome from the lateral surfaces (Luger et al. 1997). These PTMs compose an "epigenetic code" and serve as scaffolds for reader modules (such as PHD finger, bromodomain, chromo domains etc..), which in turn dictate the ordered recruitment of non-histone proteins and enzymes (Patel e Wang 2013) that alter the physical structure of the chromatin and control the gene expression (Patel e Wang 2013). The misinterpretation of these epigenetic markers lead to inappropriate expression or silencing of gene programs (Bakeret et al. 2008).

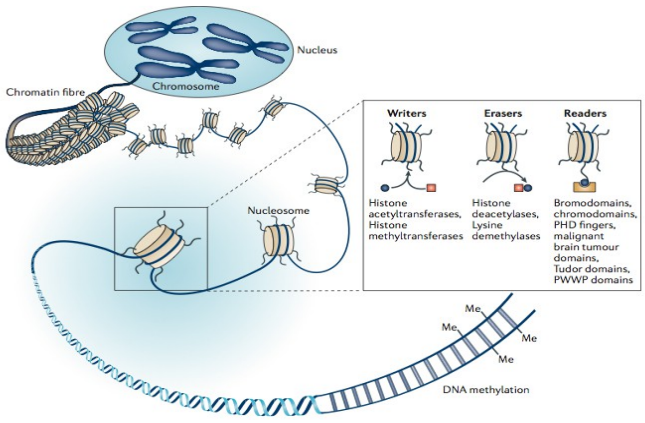


Fig 1. Nucleosome organization (adopted from Arrowsmith et al. 2012). DNA is packaged into chromatin by wrapping around histone proteins (two copies each of histones H2A, H2B, H3 and H4) to form a nucleosome. Nucleosomes are further compacted by additional protein factors to form chromatin, with the degree of compactness dependent on the types of post-translational modification present on the chromatin, with the degree of compactness dependent on the types of post-translational modification present on the histones, especially on their terminal residues, which protrude from the nucleosome particle.

1.3 Chromatin organization in human diseases

Disordered chromatin organization and function have been ascribed key pathogenic roles in several human syndromes. These disorder are associated with the mutations in single genes that encode DNA methyltransferase and histone modifying enzymes or their 'reader' protein (Jakovcevski e Akbarian 2012). In the Table 1 are reported several examples of human disorders correlated with mutation of epigenetic modules.

While physiological role of catalytic enzymes that mediate the addition or removal of histone modifications is well characterized, in the last years several studies have been focused on the reader proteins. Determining the regulatory roles of these motifs in the context of human disease will allow for a more thorough understanding of normal and pathological development, and may provide innovative therapeutic strategies wherein "chromatin readers" stand as potential drug targets.

Recently, the *P*lant *H*omeo*d*omain (PHD) finger is emerging as a versatile reader module that has a crucial role in chromatin organization and in a variety of diseases including immunodeficiency

syndrome, solid and blood cancers, and neurological disorders (Baker et al 2008). In the next chapter we will summarize the principal biological and structural characteristics of these protein domains.

Gene	Function	Syndrome
ATRX (Xq21.1)	Replication-independent nucleosome remodeling and histone H3.3 incorporation	Alpha-thalassaemia, X-linked with mental retardation (ATRX), autism
CREBBP (16p13.3) and EP300 (22q13.2)	Transcriptional co-activator, histone acetyl-transferase	Rubinstein-Taybi syndrome (RSTS) 1 and 2
DNMT1 (19p13.2)	DNA methyltransferase. Disease mutations are associated with hypomethylated repeats and promoters	Hereditary hensory and autonomic neuropathy type 1 with adult-onset dementia (HSAN1E)5, autosomal dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN)
DNMT3B (20q11.21)	DNA methyltransferase. Disease mutations are associated with hypomethylation of pericentric repeats	Immunodeficiency, centromere instability, facial anomalies (ICF1)
ZBTB24 (6q21)	Transcriptional repressor and regulator of DNA methylation at pericentric repeats	Immunodeficiency, centromere Instability, facial anomalies (ICF2) mental retardation syndrome
KDM5C (Xp11.22)	Histone H3-lysine 4 demethylase	X-linked mental retardation, autism
EHMT1 (9q34.3)	Histone H3-lysine 9 methyltransferase	Kleefstra (mental retardation) syndrome109, schizophrenia110, non- specific psychiatric phenotypes and neurodegenerative disease in post- adolescence period
NSD1 (5q35.2-q35.3)	Histone H3-lysine 36 and H4- lysine 20 methyltransferase	Sotos (mental retardation) syndrome
PHF8 (Xp11.22)	Histone H3-lysine 9 demethylase and transcriptional activator	X-linked mental retardation withoutcleft lip and/or palate (Siderius-Hamel)
RSK2 (Xp22.12)	Serine/threonine kinase (of both histones and non-histone proteins)	Coffin-Lowry X-linked mental retardation syndrome
MECP2 (Xq28)	Methyl CpG binding protein	Rett and other neurodevelopmental syndromes, autism

Table 1. Human disorders correlated with mutation of epigenetic modules

1.4 The PHD finger domain

1.4.1 Structural characterization of the PHD finger domain

The PHD finger is a small protein domain of 50–80 amino acid residues of diverse sequences containing a zinc-binding motif that appears in many chromatin-associated proteins (Sanchez and Zhou 2011). Its fold is composed by a two-stranded β-sheet supplemented by one or more helical segments and it is stabilized by a pair of Zn ions coordinated to a Cys4-His-Cys3 segment (Fig 2). Three loops (L1, L2 and L3), variable in length and residue composition, constitute the rest of the domain. The structure of the fold is maintained by the presence of the Zinc ions: mutations targeting the Zinc coordinating residues have detrimental effects on the structure of the domain, and thus on

the correct protein activity (Baker et al. 2008). Conserved hydrophobic residues establish non polar interactions that further stabilize the structure around a well conserved tryptophan residue (Musco and Peterson, 2008).

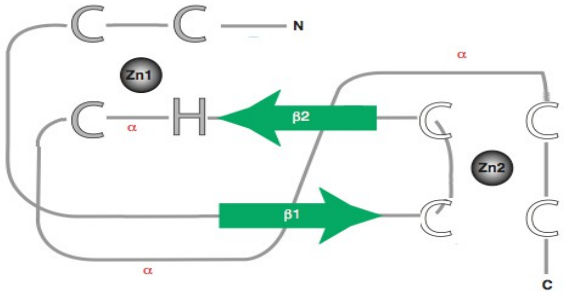


Fig 2. All PHD fingers adopt the same basic topology. Zinc atoms are shown as gray spheres, with Zinc-coordinating residues indicated as large uppercase letters. The two core β-strands are shown in green. Regions that adopt an α-helical conformation in some PHD fingers are indicated with α.

1.4.2 The PHD fingers are epigenetic reader domains

PHD fingers can read the N-terminal tail of histone H3, mainly the methylation state of H3K4 (K4me0 vs K4me3/2), and to a smaller extent the methylation state of H3R2 (R2me0 vs R2me2) and tri-methylated H3K9 and the acetylation state of H3K14 (Sanchez and Zhou 2011; Zeng et al. 2010). Binding of yeast PHD fingers to methylated H3K36 has also been reported (Shi et al. 2007). The reading of H3K4 and H3R2 occurs on a single PHD finger, but at distinct sites. By contrast, the H3K14 reading occurs in conjunction with the reading of H3K4 and H3R2, in a tandem PHD finger of DPF3b (D4, zinc and double PHD fingers, family 3 protein), where the first PHD reads H3K14ac, and the second PHD reads H3K4me0 and H3R2me0 (Zeng et al. 2010).

Common elements are shared by all histone binding PHD fingers for which structural data are available. These domains bind the first six N-terminal residues of H3 (ARTKQT) on the surface of the PHD finger through the formation of an anti-parallel β-strand to the two-strand β- sheet (β1 and β2) of the PHD finger (Fig 5). The backbone amide of A1 is clamped by the backbone carbonyl oxygens of residues in the L3 stretch, that most often contain a Pro-X-Gly-X-Trp motif (where X is any residue). The backbone amide and carbonyl groups of residues R2, T3 and K4 of H3 form hydrogen bonds with the β1 strand backbone. This β-strand conformation orients R2 and K4 into two distinct pockets separated by a highly conserved Trp residue. In general, binding of the H3 tail does not seem to induce significant conformational changes in the PHD finger (Sanchez and Zhou 2011). PHD fingers have been divided into several subgroups, depending on the post-translational modification selectively recognized. In the following, I will briefly describe these classes of PHD finger domains.

1.4.2.1 Reading H3K4me3

Several structures of PHD domains in complex with H3K4me3 are available. The binding, as mentioned above, of H3K4me3 occurs through an "aromatic cage" (Fig 3) where the trimethyl ammonium group is stabilized by van der Waals, cation–π and hydrophobic interactions. In general, residues that are used to form the aromatic cage tend to exist in parts of the structure that are relatively rigid, such as the β-strands, or close to the Zn-coordinating Cys residues (Mellor 2006). The simplest aromatic cage observed so far is that of jumonji, AT-rich interactive domain 1A (JARID1A), which is composed of only two Trp residues (Wang et al. 2009). A slight variation in the H3K4 binding region is observed in the PHD fingers of yeast homolog of mammalian ING1 (Yng1) (Taverna et al. 2006), transcription initiation factor TFIID subunit 3 (TAF3), pygopus homolog 1 (PYGO) (Miller et al. 2013), inhibitor of growth protein 4 (ING4) (Hung et al. 2009) and other PHD fingers of the ING family that have charged (Asp) or hydrophilic (Ser) residues in close proximity to the H3K4 residue. The presence of charge residue shift the affinity in favor of the dimethylated form of K4. However, the biological impact of these small differences in affinity is not clear.

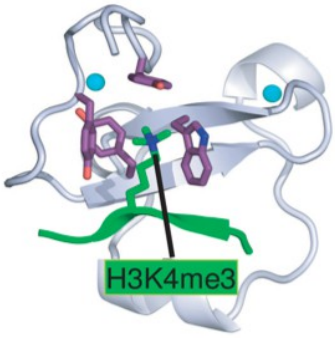


Fig 3. PHD fingers specifically recognize and bind to the methylated-H3K4 (H3K4me3). Adopted from Baker et al 2008.

1.4.2.2 Reading H3K4me0

A second subclass of PHD fingers specifically binds histone H3 tail presenting the unmodified H3K4 state (K4me0) (Fig 4). These class of PHD fingers include the modules of BRAF35–HDAC complex protein (BHC80) (Lan 2007, 80), autoimmune regulator (AIRE) (Chignola et al. 2009), tripartite motif-containing protein 24 (TRIM24) (Hung et al. 2009, 24) and DNA (cytosine-5-methyltransferase 3-like protein, DNMT3L) (Ooi et al. 2007). The binding of K4me0 is achieved by a combination of acidic and hydrophobic residues that replace the aromatic cage. The interaction is abrogated by the methylation of H3K4. Similar to the H3K4me3-binding PHD fingers, the histone H3 tail also forms a b-sheet with the b- strands in the H3K4me0-binding PHD fingers.

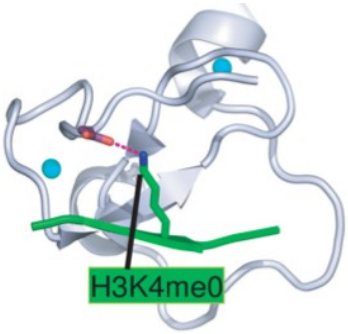


Fig 4. PHD fingers specifically recognize and bind to the unmodified-H3K4 (H3K4me0). Adopted from Baker et al 2008.

To this class belongs the PHD domain that displays a higher affinity to H3K9me3 with respect to H3K9me0. This the case of the second PHD finger of Chromodomain-helicase-DNA binding protein 4 (CHD4), in which the exposed phenylalanine side-chain forms cation-π interactions with the side chain trimethylammonium moiety of H3K9 (Mansfield et al. 2011). The examples proposed show that PHD finger domains can function individually and read a single epigenetic markers. Recently, it has been shown that the association of two PHD finger domains can read specific combination of post-translational marks. In the next paragraph, I will present the cases of tandem PHD fingers that work as unique functional unit recognizing two post-translational markers.

1.4.2.3 Combinatorial reading

The tandem PHD finger of DPF3b was shown to bind H3 unmodified at K4 and acetylated at K14 (H3K14ac) (Zeng et al. 2010). This was the first case of a PHD finger that binds an acetylated lysine, a post-translational modification usually recognized by bromodomains. The NMR structure of the DPF3b-PHD12 bound to H3K14ac shows that the two domains have a typical PHD finger fold and they are oriented one against the other in a face-to-back direction with the H3 peptide binding across the combined domains (see Fig. 5). The first PHD finger (PHD1) in the DPF3b tandem binds H3K14ac through a hydrophobic pocket situated on the opposite site of the central β-sheet. The second PHD finger (PHD2) recognizes the unmodified H3 Arg2 and Lys4 (Fig 5) and the histone assumes the classical β-strand conformation. The PHD2 is similar to the H3K4me0-binding PHD fingers (BHC80, AIRE, DNMT3L): it is characterized by the presence of acidic and hydrophobic residues that form the H3K4me0 binding pocket. In this study, the authors show that the tandem PHD12 finger of DPF3b works as one functional unit and the binding is modulated by two different modifications: it is promoted by acetylation at H3K14 (by PHD1) and inhibited by methylation at H3K4 (by PHD2). Thus, the recognition of acetylated K14 is fundamental for

recruitment of the transcription factor involved in gene activation during the heart and muscle development.

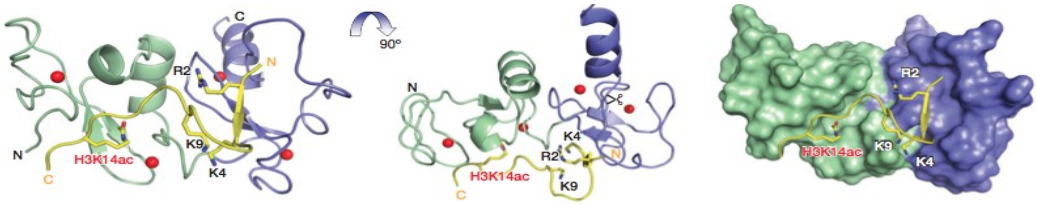


Fig 5. Three-dimensional structure of the tandem PHD finger of human DPF3b. Ribbon diagram depicting the average minimized NMR structure of the PHD12 finger bound to an H3K14ac peptide (yellow) in front and top views. The zinc atoms are highlighted as red spheres (in the first two picture from the left). In the third picture, space-filled PHD12 structure, highlighting the H3K14ac peptide (yellow) bound across the unified structure of PHD1 (green) and PHD2 (blue). (Adopted from Zeng et al. 2010).

The tandem PHD finger of MOZ regulator factor (Dreveny et al. 2014) has a highl homology sequence with the double PHD finger of DPF3b, but it recognizes the acetylated H3K14 in a different manner (Fig 6). Additionally, the MOZ tandem PHD finger is an example of how the chromatin regulator domains can modulate the structure of H3 histone tail and recognize in combinatorial way different post-translational modifications (Fig 7).

The histone tails adopt an unusual α-helical conformation, revealing a unique mode of H3 recognition. In the crystal, the two PHD fingers show a typical globular domain and the histone binding site is localized at the domains interface, stabilized by salt bridges and hydrophobic contacts. The helical structure of the H3 histone tail directs the recruitment of epigenetic marks. The side chain of H3A1 inserts into a hydrophobic pocket formed by P-X-G-X-W motive and the H3 Arg2 engages in a network of hydrogen bonding and electrostatic interactions with two aspartate residues on the second PHD domain. Interestingly, H3 Lys4 engages in direct hydrophobic and hydrogen bonding interactions with residues at the interface, and in this conformation the metylation of the K4 abolishes the binding with H3 histone tail. The K9 is exposed and located on the opposite site with respect to K4, it has no impact on the structure of histone tails and it is accessible for other modification. The acetylated H3K14 side chain is insert in preformed pocket on the first PHD finger surface: this pocket is mainly hydrophobic and exhibits a weak positive electrostatic potential, that is not appropriate for unmodified K14 interaction.

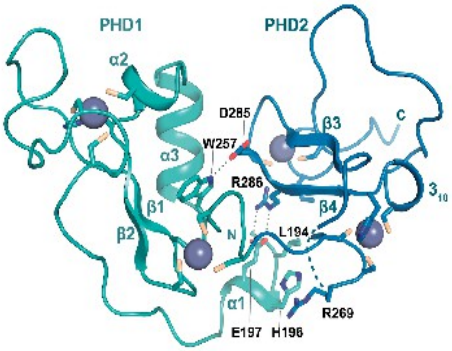


Fig 6. Structure of the MOZ DPF domain. The PHD1 and PHD2 subdomains are colored in turquoise and blue, respectively. Residues engaged in interactions with the N-terminal helix α1 and PHD2 as well as a direct hydrogen bonding interaction between PHD1 and PHD2 are labelled and shown in stick representation. (Adopted from Dreveny et al 2013)

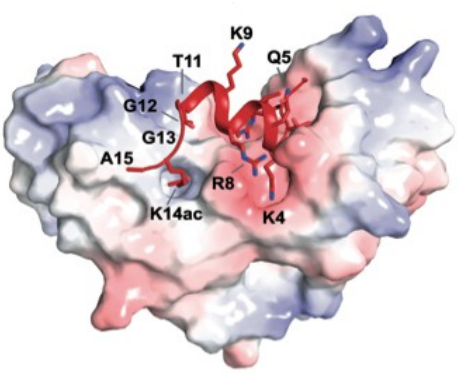


Fig 7. Structural basis of H3K14ac recognition by MOZ. Surface representation of the MOZ DPF domain coloured according to electrostatic potential in complex with H3K14ac in red cartoon representation. (Adopted from Dreveny et al 2013)

PHD fingers are versatile chromatin regulator domain with diversified functions. For example, the KAP1 (KRAB-ZFP–associated protein 1) PHD finger is associated with a bromodomain and it acts as a SUMO E3 ligase (Ivanov et al. 2007). Other PHD fingers were shown to bind simultaneously histone and non-histone partners, such as the Pygo PHD. This domain is a histone H3K4me2/3 reader that simultaneously associates with the histone H3 tail and the HD1 domain of the BCL9/Lgs proteins using a different binding surface. This ternary complex mediates a cross-talk between epigenetic and non epigenetic cellular signalling pathways that is important to transduce Wnt signalling (T. C. R. Miller et al. 2013).

In my PhD thesis I have focused my attention on Nuclear receptor binding SET domain protein 1 (NSD1) (Morishita and di Luccio 2011; Lucio-Eterovic et al. 2010). This protein contains 6 PHD finger that are strictly associated suggesting that they function in cooperative manner and as structural hubs for the recruitment of non-histone protein and epigenetic markers. Deletion, point mutations that involve these PHD fingers cause Sotos syndrome (Pasillas et al. 2011), a rare genetic disorder. When fused to NUP98, NSD1 is involved in acute myeloid leukemia (AML), and its PHD finger domains seem to play a role in aberrant gene expression (Wang et al. 2007). In the next chapter I will describe the biological role of NSD1, and how its PHD finger domains are involved in physiological functions and in pathological condition.

1.5 NSD1

NSD1 is a nuclear protein involved in the regulation of the chromatin structure. It is essential for the development of mouse embryos and disruption of the *Nsd1* gene led to failure in the post-implantation development (Rayasam et al. 2003). It was found for the first time in a series of yeast two-hybrid assays as a nuclear receptor interactor exhibiting properties ascribed to both coactivators and corepressors (Huang et al. 1998). It displays lysine methyltransferases activity by its SET domain (Rayasam et al. 2003) and it contains domains typical of molecules involved in chromatin regulation two proline-tryptophan-tryptophan-proline (PWWP) domains (PWWP1-2), and five classical plant homeodomains (PHDI-V) and a degenerate one (C5HCH) (Morishita and di Luccio 2011) (Fig 8).

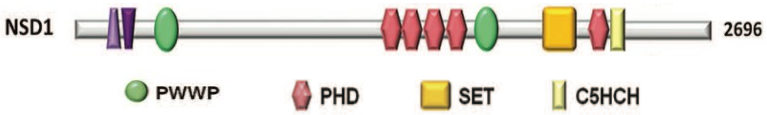


Fig 8. Functional domains contenting in NSD1.

In mouse were two isoforms of NSD1 protein were detected. The first isoform, from the mouse embryo tissues lacks the PWWP1 while the second isoform in the mouse brain tissues contains both the PWWP domains (Huang et al. 1998; Kurotaki et al. 2001). Kurotaki et al. were the first to describe the human homologue of the murine *Nsd1* gene. About NSD1 methyltranferases activity, *in vitro* enzymatic assays showed poor selectivity towards all the core histones (H3, H2A, H2B and H4). When nucleosomes were used as a substrate, the SET domain acted as a selective H3K36 dimethylase. Interestingly, this selectivity could be increased by short DNA fragments in a dose-dependent manner, indicating that the effect of DNA on NSD1 methyltransferase selectivity is independent of the formation of nucleosomes (Li et al. 2009; Qiao et al. 2011). Mass spectrometry experiments show that NSD1 is an interactor of H3K36me3. This epigenetic mark is a signal for gene activation (Vermeulen et al. 2007). Biochemical studies (GST-pull down) suggest that the fifth PHD finger domains recognize both the methylated K4 and K9 on histone tail 3, two epigenetic marks, which usually lead to opposite outcomes, gene activation and gene repression, respectively (Pasillas et al. 2011). Nielsen *et al* identified Nipz1 (NSD1-interacting zinc finger protein 1) as interactor of NSD1. Nizp1 harbours a transcriptional repression domain, for which NSD1 seems to act as a corepressor, and it contains two cluster of Cys-Cys-His-His zinc finger domains that in this conformation display DNA and protein interaction activity. Interestingly, two hybrid assays revel that the tandem domain of NSD1 composed by PHD5 and C5HCH domain recognizes the Cys-Cys-

His-Arg (C2HR) motif of Nizp1, a degenerate copy of the zinc finger motifs.

These features revel a bifunctional role for NSD1 that can positively and negatively regulate transcription depending on the cellular context. In the HCT116 colorectal cancer cells NSD1 binds near various promoter elements and regulates multiple genes involved in various processes (Lucio-Eterovic et al. 2010). In particular, the Chip microarray revels that there are over 300 candidate NSD1 targets in HCT116 cells. This includes genes implicated in cancer: Bone morphogenetic protein 4 (BMP4), Cathepsin D (CTSD), and Kruppel-like factor 6 (KLF6), von Hippel tumor suppressor (VHL), and Kallikreins KLK6 and KLK14. Additionally, several cell cycle-related genes were identified: CDK4, CDC6, CHEK1, CCNE1 (cyclin E), CDKN2A, and the cyclin B interacting protein CCNB1IP1. It was found that NSD1 binds to the promoter regions of 20 different keratin genes, including two (KRT3 and KRT6B) known to be regulated by the p65 subunit of NF-κB. Moreover, it was identified the p65 targets SLC16A1, IGF2BP2 as a possible NSD1-regulated genes. The genes involved in the development and mental retardation candidate gene (ZMYM3) are also potential targets of NSD1. NSD1 regulates target gene expression in a tissue-specific manner. For example the knockdown of NSD1 in MCF-7 breast cancer cells reduce the levels of the KRT10 transcripts, while in HCT116 cells the expression of these gene are independent of NSD1 function. Additionally, in U2OS osteosarcoma cells the ZFP36L1/TPP transcript levels were significantly abrogated, but they were modestly affected in NSD1-deficient HCT116 cells (Lucio-Eterovic et al. 2010). NSD1 binds the promoter region of BMP4 gene and promotes H3K36methylation in this region. H3K36 trimethylation is required for the recruitment of RNA polymerase II to the BMP4 gene promoter and for transcriptional elongation (Berdasco et al. 2009). In pathological conditions, such as the Sotos Syndrome (see chapter 1.41), the bone morphogenesis is strongly perturbed.

1.5.1 NSD1 in human disease: the case of Sotos Syndrome and Acute Myeloid Leukemia (AML)

The physiological role of NSD1 is largely unknown. In this paragraph I will summarize the diseases in which NSD1 is involved.

It has been shown that NSD1 is involved in positive regulation of Hox genes: in fact knockout of *Nsd1* in mice produces developmental defects and precludes expression of HoxA1 and HoxB1 (Rayasam et al. 2003).

The acute myeloid leukemia (AML) disease is characterized by uncontrolled expansion of myeloid-lineage progenitors, HoxA7, HoxA9 and HoxA10 are overexpressed with Meis1, a Hox cofactor gene. About 5% of human AMLs harbour the t(5;11)(q35;p15.5) translocation, a chimeric gene

encoding the FG-repeat domain of NUP98, which binds the histone acetyltransferase CBP/p300, fused to the carboxy-terminal 60% of NSD1, which contains the exons 6-23 that include almost all the functional domains of NSD1 (Wang et al. 2007). According to the mechanism proposed, NSD1 has H3K36 methyltransfesare activity and it regulates normal Hox gene expression, consequently deregulated NUP98–NSD1 activity could link persistent methylation of H3K36 to enforced expression of HoxA genes in AML.

NUP98-NSD1 binds to the Hox-A locus, where it exerts two "writer" activities on histones, i.e. H3K36 methyltransferase and acetylation, mediated by NSD1 SET domain and by the NUP98-recruited CBP/p300, respectively. The combination of these epigenetic marks excluded the H3K27 trimethylation of the locus mediated by EZH2 that it is able to silence the Hox-A locus. In this way only six genes, HoxA5, HoxA7, HoxA9, HoxA10, Meis1 and Rab38, were up-regulated fivefold or more in NUP98–NSD1 progenitors, demonstrating a high targeting specificity by NUP98–NSD1 on the Hox-A locus and on Meis1. NUP98–NSD1 enforces self-renewal and arrests differentiation by activating transcription of Hox-A genes and Meis1. The gene up-regulation coincides with appearance of the H3K4me3 active mark and the depletion in of H3K9me3 repressive mark. Both acetylation and H3K36 methylation were shown to be essential for the disease: the ablation methyl and acetyl tranferase activity prevented the immortalization of the progenitors (Wang et al. 2007). Importantly, the two PHD finger in C-terminal region of NSD1 are important for the binding with the DNA probes 5' of the Hox-A locus. The deletion mutant lacking these two modules was not associated with Hox-A gene actication and stem cell immortalization. Thus, the chromatin binding activity of these domains is crucial for the leukaemogenesis by the chimaeric protein NUP98/NSD1 and According to the proposed model, the recruitment of the NUP98/NSD1 on specific promoter regions is mediated by an unknown adaptor (Wang et al. 2007).

The second pathology in which NSD1 is involved is the Sotos syndrome, an overgrowth disorder. Sotos syndrome is a human developmental and cognitive disorder due to heterozygous mutations in NSD1 (Tatton-Brown et al. 2005; Tatton-Brown and Rahman 2007). This syndrome is a rare genetic disorder (its incidence is approximately 1 in 14,000 births) characterized by excessive physical growth during the first 2 to 3 years of life. The disorder is associated with mental retardation, delayed motor, cognitive, and social development, hypotonia (low muscle tone), and speech impairments. Children with Sotos syndrome tend to be large at birth and are often taller, heavier, and have larger skulls and larger heads (macrocephaly) as compared to healthy children in their age. Although most cases of Sotos syndrome occur sporadically, familial cases have also been

reported. Point mutations, deletions that involve 9 of 13 NSD1 domains, including all 6 PHD domains, cause the Sotos syndrome, indicating that each NSD1 domain performs an essential role (Pasillas, Shah, e Kamps 2011). Sotos syndrome point mutations found in NSD1 PHD fingers are listed in Table 2. Interestingly, numerous mutations associated with Sotos syndrome lie in NSD1-PHD5 and NSD1-C5HCH PHD finger domains.

Domain	Mutation(s)
PHD1	C1570F (2)
PHD2	C1596R (7); C1611S (2); H1616L (3); C1619R(2); L1637P(3)
PHD3	C1640Y (2); R1663C (2); C1674W (3) I1687N (4); P1690L (7)
PHD4	C1713G (7); P1726L (4); A1727P (1) C1733R (7) C1733S (7) L1734P (7)
PHD5	C2124R (2); Y2142N (2); H2143E (3); H2143Y (2) H2143Q (2); C2146R (2); R2152Q (6); C2159Y (2); H2162R (2)
C5HCH	C2164R (2); C2164Y (2); C2167R (1); C2178R (2); F2182I (2); C2183S (2);H2205R (2); C2178Y (7)
SETdomain	R1952W (2); G1955D (2); G1955R (2); R1984Q (1); Y1997C (2); Y1997S (2); R2005Q (2); R2017Q (2); R2017W (2); C2027R (7); G2041D (3)
PWWP2	C1773R (2); C1773S (4); G1792V (2); V1796F (7)

Table 2. Sotos syndrome point mutations in NSD1. References are: 1: Kurotaki et al., 2003; 2: Tatton-Brown et al., 2005; 3: Douglas et al., 2003; 4: Rio et al., 2003; 5: Turkmen et al., 2003; 6: Melchior et al., 2005; 7: Pasillas et al., 2011.

Patients with identical mutations are characterized by extreme variability in terms of learning disability and clinical features, possibly due to phenomena such as stochastic factors and functional polymorphisms of NSD1-interacting genes (Tatton-Brown et al. 2005). Visser R et al studied the gene expression profile of dermal fibroblasts from Sotos syndrome patients suggesting that the haploinsufficiency of NSD1 is associated with the deregulation of the MAPK/ERK signaling pathway, involved in cell differentiation. However, the molecular basis of NSD1 function in Sotos syndrome pathogenesis is still unknown (Visser R et al. 2012).

Rahman et al *2003* noted an increased incidence of neural crest tumors in Sotos cases with confirmed NSD1 aberrations (Rahman N et al. 2005). The hypermethylation of CpG island at the NSD1 promoter leads to inactivation of NSD1 in neuroblastoma (LAI-5S and LAN-I) and glioblastoma (U373-MG) cells lines. In these cell lines, the minimal expression of the NSD1 is associated with reduced levels of trimethylation at residues H4K20 and H3K36. In particular in neuroblastoma cell line, when NSD1 is depleted by means of RNA interference, it is observed an up-regulation of the oncogene Meis1. The quantitative Chip shows that NSD1 localizes at 5' of Meis1 promoter led to transcriptional repression of the gene via its H4K20 methyltransferase activity. Sotos syndrome lymphoblastoid cells , which had an NSD1 genetic defect (R604X,

L779X, C1640Y, R2017W, OGS 55, OGS368, OGS661) expressed high levels of the Meis1 transcript if compared with healthy donors (Berdasco M et al. 2009). The histone modification profile in tumour cells (LAI-5S) and Sotos syndrome cell line in 5' region of Meis1 is very similar.

The epigenetic silencing or genetic loss of NSD1 led to an enrichment in active marks (such as H3K4me3) and a depletion in repressive marks (H4K20me3, H3K9me3, and H3K27me3). In contrast, when the NSD1 CpG island promoter are unmethylated (neuroblastoma cells, SN-N- JD) the recruitment to 5' region of Meis1 is associated with an enrichment in the transcriptional silencing histone marks (H4K20me3, H3K9me3, and H3K27me3) and a depletion in the active marks (H3K4me3). These experimental data suggest that NSD1 is a versatile protein that can act as a corepressor or coactivator depending on the cellular context (Berdasco M et al. 2009).

NSD1 forms a complex with the repressor factor Nizp1, this complex is probably responsible for chromatin compaction and suppression of gene expression. In the next chapter, I will describe the biological role of Nizp1 and how it interacts with NSD1.

1.6 Nizp1 is a zinc finger containing protein that interacts with NSD1

Nizp1 contains 585 amino acids, with a calculated molecular mass of 66.5 kDa (Nielsen et al. 2004) (Fig 9). The N-terminal region of Nizp1 (amino acids 36 to 131) contains a SCAN box, or leucine-rich region (LeR). This motif functions as an oligomerization domain mediating self-association or association with other proteins bearing SCAN domains. A Kruppel-associated box-A (KRAB-A) subdomain was identified in Nizp1 between amino acid residues 219 and 259 that has been shown to act as a potent transcriptional repressor element when it is targeted to DNA. At the C-terminus, Nizp1 contains four classical zinc fingers (C2H2, named briefly ZnF), ZnF1 to ZnF4, that occur in two clusters and are capable of binding to DNA directly. They are preceded by a degenerate copy of this domain which has an Arg residue in place of the last His residue, and it is named C2HR motif. The ZnF domains display DNA and protein interaction properties when they are organized in repeating units.

Fig 9. Functional domains contenting in Nizp1. The C2HR domain is reported in the red square

Two-hybrid assays in yeast and mammalian cells and GST-pull down have shown NSD1 and Nizp1

motifs interact with each other. In particular the tandem domain (PHD5-C5HCH) at C-terminus of NSD1 interacts with the C2HR domain of Nizp1. The other two members of NSD protein family, NSD2 and NSD3 that also contain a PHD5-C5HCH tandem domain, are not able to recognize Nizp1 (Nielsen et al. 2004).

The role of the complex NSD1 and Nizp1 in gene repression has been evaluated by gene reporter assays. In Cos-1 cells, it was assayed the repressive activity of the Nizp1 C2HR motif alone or in combination with overexpressed NSD1. In presence of both the protein the report gene expression is strongly reduced. Moreover, an NSD1 mutant protein, NSD1C2073A/ C2076A, bearing mutations that abolished the interaction with the C2HR motif, did not affect the expression of the target gene. This results indicate the interaction between the two protein is fundamental for the gene repression. In a second set of experiments the role of full-length Nizp1 to mediate the gene repression was evaluated. In HT1080 cells, in presence of full-length Nizp1 and over-expressed NSD1 it is possible to observe the efficient repression of the reporter gene.

However, the C2HR mutations preventing NSD1 interaction severely reduced but did not abolish the Nizp1 repression activity. This data suggests that also the KRAB-A domain of Nizp1 is involved in gene repression. This domain has been shown to silence transcription through the recruitment of the co-repressor protein TIF1β/KAP-1(Nielsen et al. 2004). As mentioned above, NSD1 has an histone methyltranferase (HMTase) domain that specifically methylates histone H3 at lysine 36 (H3K36) and histone H4 at lysine 20 (H4K20) but the functional consequences of these modifications remain to be determined.
Nielsen *et al 2004* propose a model in which Nizp1 is a DNA-binding transcriptional repressor whose function is to target the H3-K36 and H4-K20 HMTase activities of NSD1 to specific promoters. The details of the molecular mechanism that involve the interaction between NSD1 and Nizp1 are largely unknown. As first step towards the molecular characterization in this study we focused our attention on C2HR motif of the Nizp1 and on its interaction with NSD1. This module belongs to the family of ZnF domains in which the final histidine is absent and replaced by arginine residue. In the next chapter we will describe the biological and structural characteristics of the zinc finger modules.

1.6.1 The zinc finger modules in protein interaction

The zinc finger domains has been shown to be very abundant in eukaryotes, about the 3% of human genes encode ZnF-containing proteins (Bouhouche, et al. 2000). Initially discovered in the

transcription factor TFIIIA from Xenopus oocytes (J. Miller, et al.1985), the classical ZnF motif consists of a short sequence (30-40 residues) that uses two or three cysteines and two or one histidines to coordinate a zinc ion (C2H2 or C2HC types, respectively).

The zinc finger domains fold into a characteristic structure consisting of an α-helix and a β-hairpin held together by the zinc ion. The coordination geometry of four coordinated zinc atoms is most commonly tetrahedral (Gamsjaeger et al. 2007) (Fig 10).

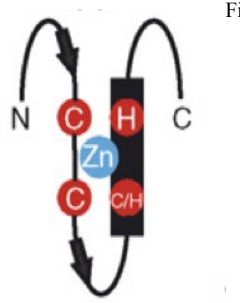


Fig 13. Zinc finger domain topology

In the last 20 years since zinc-binding repeats were first identified in a protein, the DNA-binding properties of zinc-fingers have been explored in exquisite detail (Wolfe SA et al 200), and it is now possible to generate a specific zinc finger proteins that will target specific genomic sites. These zinc finger domains have proved to be common in complex organisms: in humans, >15 000 such domains are predicted to exist in about 1000 different proteins (Rubin GM et al 2000).

It has also emerged that zinc fingers can mediate protein–protein interactions. In the past 3–5 years, detailed structural and functional data have been reported for several zinc finger-protein interactions, showing unequivocally that they can function as protein recognition motifs. In particular, it has been identified a family of classical zinc finger domains, in which the zinc coordinating histidine is not conserved and residues such as aspartate have been shown to function as zinc ligands (Andreini et al 2011). Several studies demonstrate that these domains maintain their protein fold and their ability to recognize DNA promoter region and protein partners (Simpson et al. 2003).

Aim of the work

NSD1 is a multitasking protein whit a dual activity that can act both as a repressor and coactivator, depending on the cellular context (Berdasco et al. 2009; Wang et al. 2007). This versatility is probably linked to the presence of six PHD finger domains that function as structural hubs for multiple binding events recognizing both epigenetic markers and non-histone proteins (Sanchez and Zhou 2011). The emerging standard strategy for characterizing chromatin-binding proteins, such as NSD1, is to establish the relationship between a discrete portion of the large chromatin-binding protein, called module or functional unit, and its substrate (Malecek and Ruthenburg 2012).

In this project we focused our attention on the tandem PHD finger domains in the C-terminal region of NSD1 that is composed by PHD5 and C5HCH finger domains, briefly named as NSD1-P5C5. The tandem domain appears to play a multifaceted role working as both histone reader (H3K4me3 and H3K9me3, Pasillas M et al. 2011) and as interaction domain with the C2HR domain of the transcriptional repressor Nizp1 (Nielsen et al. 2004). The complex formed by NSD1 and Nizp1 is probably responsible for chromatin compaction and suppression of gene expression. It was speculated that the abrogation of NSD1-mediated repression of growth-promoting genes might contribute to pathological conditions (Lucio-Eterovic et al. 2010), such as the Sotos Syndrome an overgrowth disorder that is due to point pathological mutations in NSD1 domains (Tatton-Brown e Rahman 2007). In this context, about 20 mutations target the NSD1-P5C5 suggesting that it has a role in pathology. From the structural point of view, it would be informative to study how the NSD1-P5C5 fold is affected by these mutations. NSD1-P5C5 is essential for the pathogenesis of acute myeloid leukaemia (AML) caused by the chimeric protein NUP98/NSD1 that forces the abnormal activation of Hox-A and Meis1 genes. The deletion of these two domains is sufficient to abolish NUP98/NSD1 interaction with chromatin, preventing both the transcription activation and the immortalization of myeloid progenitors.

Until now, no structural data are available about the interaction of NSD1-P5C5 with non-histone proteins and epigenetic markers. Structural and thermodynamic studies characterizing the molecular details of these interactions would strongly contribute to understand NSD1 function in physiological and pathological conditions and would possibly offer a structural platform to develop synthetic modulators of these interactions.

In this framework, the structure and dynamic properties of this tandem domain have been

characterized by NMR spectroscopy by our group. In this thesis we focused our attention on the role of NSD1-P5C5 as interactor of Nizp1-C2HR with the aim to study by biophysical techniques (NMR spectroscopy, ITC experiments): (1) the structural and functional effect of pathological mutations in NSD1-P5C5 (2) the structure of Nizp1-C2HR, (3) the interaction of NSD1-P5C5 with Nipz1-C2HR domain.

2. Materials and Methods

3.1. General molecular biology techniques

The major molecular biology procedures applied to carry out cloning and site direct-mutagenesis on the murine Nsd1 fifth PHD finger in association with the C5HCH domain (NSD1-P5C5,spacing residues Glu2014-Asp2104, cod.sequence Q96L73) and murine Nizp1 first zinc finger domain (Nizp1-C2HR, spacing residues Glu397-Lys434, cod. sequence Q5SXI5) are described in details in the following.

3.1.1 Standard DNA amplification of the genes of interest

NSD1-P5C5 was cloned into NcoI/KpnI sites of pETM11 expression vector (EMBL)

The plasmids for expression in *E.coli* of Nizp1-C2HR domain were generated using a multi-step PCR-based process. The regions were amplified from the pGEX vector containing the cDNA of the murine Nizp1 protein. According to the sequence the forward and reverse primers were designed and synthesized by PRIMM company, Italy. The used oligonucleotides are reported in Table 1 and the start and the stop codons are bold faced.

Primer Name	Sequence	Length	Restriction Enzyme
Nizp1_f	5'-GGCGCCATGGAGGTG CAAA CCT CAC-3'	25nt	NcoI
Nizp1_r	5'-CAGAAGCCACACAAATAAGCGGCCGCCCG-3'	29nt	NotI

Table 1 - Description of the primers used for the cloning procedure of Nizp1-C2HR

The composition of PCR reaction, in a final solution of 50 µL, is reported in Table2 and the PCR reaction was performed in a thermocycler, according the amplification protocol reported in Table 3.

Reagent	Concetration
DNA template	30ng
5' primer	1mM
3' primer	1mM
dNTPs	0.1mM
DNA polimerase	3 units
Nuclease free water	until 50 µl

Table 2 - Composition of standard PCR reaction mix

Step	Temperature (°C)	Time	Cycle applied
Inizial Denaturation	95	2'	1
Denaturation	95	45"	35
Anneling	55	30"	35
Elongation	72	2'	35
Final Elongation	72	6'	1

Table 3- Standard amplification protocol

The amplified DNA fragments were purified by means of agarose gel (0.8%), digested for 3 houses at 37°C with NotI and KpnI (Table 4) and then purified again on agarose gel.

Reagents	**Concentration or volume**
DNA volume	10-50 ng
NotI	5 units
KpnI	5 units
Digestion buffer	5 µl
Nuclease free water	until 50 µl

Table 4- Composition of standard reaction mix for digestion with restriction enzymes

Ligation between the DNA fragments and the destination vector pETM-41 (the vector provided by EMBL was previously digested with NotI-KpnI and purified), was performed for 2 hours at 22°C (Table 5).

Reagents	**Concentration or volume**
DNA vector	10-20 ng
DNA amplified	2-4ng
Ligase T4 enzyme	200 units
Ligase buffer	2 µl
Nuclease free water	until 20 µl

Table 5 - Composition of standard ligase reaction mix

DNA fragments were mixed at 1:5 ratio of vector and inserts, according to the formula:
ng (digested insert) = 5 x (ng (digested insert) x length (digested insert))/ length (digested vector)

The ligation solutions were transformed in DH5α *E.coli* cells, the constructs were extracted by miniprep and verified by DNA sequencing. The vector pETM-41 containing the correct sequence of the insert was used to transform the BL21 (DE3) competent cells for the protein expression.

3.1.2 Site-directed mutagenesis protocol

The generation of NSD1-P5C5 and Nizp1-C2HR mutations were generated through PCR-based site-directed mutagenesis. For each mutagenesis two synthetic oligonucleotide primers containing the desired mutation were purchased by PRIMM (Milan, Italy) and reported in Table 6 and Table 7.

Primer Name	Sequence	Length
Y30N_f	5'-GCCAGGCTGCCCTAAA GTCAAC CATGCAGACTGTCTGAAT C-3'	19nt
Y30N_r	3'-GATTCAGACAGTCTGCATGGTTGACTTTAGGGCAGCCTGGC-5'	21nt
F70I_f	5'-GTGAGATGTGCCCCAGCTCCATTTGCAAGCAACATCGAGAAGG-3'	21nt
F70I_r	3'-CCCTTCTCGATGTTGCTTGCAAATGGAGCTGGGGCACATCTCAC-5'	21nt
R40Q_f	5'-CTATGAAATAGGAAACCGCAAGGCGGCTTTATACTATGGCG-3	20nt
R40Q_r	5'-CGCCATAGTATAAAGCCGCCTTGCGGTTTCCTATTTCATAG-3'	20nt
H30Y_f	5'-CTATGAAATAGGAAACCGCGCTGCGGCTTTATACTATGGC-3'	20nt
H30Y_r	5'-GCCATAGTATAAAGCCGCAGCGCGGTTTCCTATTTCATAG-3'	20nt
H30R_f	5'-CTATGAAATAGGAAACCGCTATGCGGCTTTATACTATGG-3'	20nt
H30R_r	5'-CCATAGTATAAAGCCGCATAGCGGTTTCCTATTTCATAG-3'	20nt
H50R_f	5'-G AAG TGG GAA TGT CCT TGG CGC CAG TGT GACGTGTGTGGG-3'	21nt
H50R_r	3'-CCCACACACGTCACACTGGCGCCAAGGACATTCCCACTTC-5'	19nt
H93R_f	5'-GCG TCTGTCTTGTACTGAACGGGATTAAGGTACCGGATCC G-3'	19nt
H93R_r	3'-CGGATCCGGTACCTTAATCCCGTTCAGTACAAGACAGAC-5'	19nt

Table 6 - Description of the primers used for site-directed mutagenesis for NSD1-P5C5

Primer Name	Sequence	Length
R34A_f	5'-GG CAC CTG CGC AGC GCC AGG GAG CAG AAG CCA C-3'	14nt
R34A_r	3'-GTGGCTTCTGCTCCCTGGCGCTGCGCAGGTGCC-5'	16nt

Table 7- Description of the primers used for site-directed mutagenesis for Nizp1-C2HR

Mutagenesis reactions were set up for each mutants , as indicated in Table 8

Reagent	Concentration
DNA template	20-50ng
5' primer	1mM
3' primer	1mM
dNTPs	0.2mM
DNA polimerase	3 units
Nuclease free water	until 50 µl

Table 8 - Composition of standard PCR reaction mix

PCR amplification reactions were performed following the protocol reported in Table 9.

Step	Temperature (°C)	Time	Cycle applied
Inizial Denaturation	95	2'	1
Denaturation	95	45"	25
Anneling	55-60	30"	25
Elongation	72 for 2 minutes/kb of plasmide	2'	25

Table 9- Standard amplification protocol

When the reaction was complete, 10 U of the Dpn I (Promega) restriction enzyme were added directly to each amplification, each reaction was incubated at 37 − C for one hour and at 70 − C for 25 minutes. DH5α *E.coli* cells were transformed using 50 µL of each reaction solution. The sequences of the mutants were verified by sequencing finally the correct plasmids were used to transform BL21-(DE3) competent cells to carry out protein expression.

3.2. Expression of recombinant protein in *E. coli*

In the following session are described the protocols used for the hetorologous expression of the recombinant protein. The expression of unlabelled recombinant protein was carried out in LB reach media. NMR experiments required ^{15}N- and ^{15}N-^{13}C- isotopic labeling. Expression of labeled protein samples was achieved by growing and inducing BL21(DE3) *E. coli* cells in ^{15}N or ^{15}N-^{13}C M9 minimal medium (see Table 10 for recipes) instead of LB medium. In M9 minimal medium, ^{15}N-HCl and ^{13}C-glucose are, respectively, the unique source of nitrogen and carbon during the bacterial growth and induction, so that we can achieve uniformly ^{15}N- or ^{15}N-^{13}C-labeled recombinant proteins.

M9 minimal medium (1L) for ^{15}N-labeling	**M9 minimal medium (1L) for ^{13}C-^{15}N-labeling**	**M9 salts pH 7 10X (100 ml)**	**Trace Elements 1000X (250 ml)**
100 ml M9 salts pH 7 10X	100 ml M9 salts pH 7 10X	10 g Na_2HPO_4	12.5 g EDTA pH 7.5
1 ml Trace Elements 1000X	1 ml Trace Elements 1000X	3 g KH_2PO_4	2.10 g $FeCl_3$•$6H_2O$
4 g D-glucose	4 g ^{13}C-D-glucose	0.5 g NaCl	210 mg $ZnCl_2$
1 ml $MgSO_4$ 1M	1 ml $MgSO_4$ 1M		30 mg $CuCl_2$•$2H_2O$
0.3 ml $CaCl_2$ 1M	0.3 ml $CaCl_2$ 1M		25 mg $CoCl_2$•$6H_2O$
3 g $^{15}NH_4Cl$	3 g $^{15}NH_4Cl$		25 mg H_3BO_3
1 ml kanamycin 25 mg/ml	1 ml kanamycin 25 mg/ml		4 mg $MnCl_2$•$6H_2O$
1 ml thiamin 1 mg/ml			
1 ml biotin 1 mg/ml			

Table 10 . Recipes of M9 minimal medium for uniformly ^{15}N- and ^{13}C-labeling

3.2.1 NSD1-P5C5 expression protocol

BL21(DE3) competent cells transformed with pETM-11-P5C5 vector were grown in 1L reach media (LB) or minimal media (added of 0.2mM of ZnCl2) at 37°C for about 2-3 hours. The protein expression was inducted with 1mM of IPTG (isopropyl β-D-1-thiogalactopyranoside) when the OD600 = 0.8-1 and performed for about 20 hours. The domain is expressed with a N-terminal tag of six histidines (6-His tag).

The cellular pellet was re-suspended in 15 ml of 20 mM TRIS-HCl pH 8, containing 150 mM NaCl, 0,2% NP-40, 1 mM TCEP, 0.1mM of ZnCl2 , 20 µg /ml DNAseI, 20 µg/ ml RNAse I and 50 µg/ml lysozime. Cells were disrupted by sonication. Cell debris was separated from the supernatant by centrifugation at 15.000 g for 30 minutes at 4°C. Heterologous expression of the NSD1-P5C5 mutants were carried out using the same protocol described for the wild-type protein.

3.2.2 Nizp1-C2HR expression protocol

Large-scale expression of the recombinant Nizp1-C2HR domain was achieved in 1 l batches of LB medium, starting from a pre-inocule of 20 ml of a 16 hours cell culture. Cells were grown at 37°C, until the OD600 reached 0.8 - 1. Expression was induced by addition of IPTG at final concentration of 1 mM and the temperature was decreased to 28°C after induction. The domain is expressed with a N-terminal tag of six histidines (6-His tag) followed by the MBP protein. Cells were harvested about 16-20 hours after induction by centrifugation at 8.000 *g* for 30 minutes at 4 °C. The cellular pellet was re-suspended in 15 ml of 20 mM TRIS-HCl pH 8, containing 150 mM NaCl, 0,2% NP-40, 1 mM TCEP, 0.1mM of ZnCl2 , 20 µg /ml DNAseI, 20 µg/ ml RNAse I and 50 µg/ml lysozime. Cells were disrupted by sonication. Cell debris was separated from the supernatant by centrifugation at 15.000 g for 30 minutes at 4°C. Heterologous expression of the NSD1-P5C5 mutants were carried out using the same protocol described for the wild-type protein.

3.3. Purification of recombinant proteins in *E. coli*

The protocol for purification of recombinant proteins will be describe in this chapter.

3.3.1 NSD1-P5C5 purification protocol

The recombinant His(6)tagged NSD1-P5C5 was purified by affinity chromatography.

The protein in the soluble fraction was loaded onto a column containing 5 ml of the *Ni-NTA* affinity resin (*Qiagen*) pre-equilibrated with 25ml of the same buffer used for the bacteria lysis, and then washed with 10CV of the buffer. 20 mM TRIS-HCl pH 8, containing 150 mM NaCl, 2 mM of 2-β-mercaptoethanol, 0.1mM of ZnCl2 , 10 mM of imidazole and with 5 CV of buffer 20 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM of 2-β-mercaptoethanol 0.1mM of ZnCl2 and 20 mM of imidazole. The recombinant protein was eluted in the buffer of 20 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM of 2-β-mercaptoethanol 0.1mM of ZnCl2 and 500 mM of imidazole. The His(6)tag was removed by TEV digestion and the protein solution was dialysed over night against the buffer 20 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM of 2-β-mercaptoethanol 0.1mM of ZnCl2 .

The protein in insoluble fraction was re-suspend in the buffer 20 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM of 2-β-mercaptoethanol and 6 M urea. After about 30 minutes of incubation at 37 °C in agitation the protein solution was centrifuged at 15.000 g for 30 minutes at 4 °C. The supernatant obtained was loaded onto a column containing 5 ml of the *Ni-NTA* affinity resin (*Qiagen*) pre-equilibrated with 25ml of the same buffer. The resin was washed two times with 10 CV of the same buffer with increasing amount of imidazole (10mM and 20mM, respectively). The elution was carry out with the buffer 20 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM of 2-β-mercaptoethanol 0.1mM of ZnCl2 and 500 mM of imidazole. In order to refold the protein domain after the elution the protein solution was dialysed over night at 4 °C against the buffer 20 mM TRIS-HCl pH 8, 150 mM NaCl, 5 mM of 2-β-mercaptoethanol 0.2mM of ZnCl2 . The refolded protein was digested with TEV protease. The digested protein solution was loaded onto a 3 ml Ni^{2+}-NTA column (previously equilibrated with the dialysis buffer) to remove 6His-tags and TEV protease (it also has a N-terminal 6His-tag). The flow-through, containing only the recombinant protein, was collected, concentrated and loaded onto a Superdex 75 size exclusion chromatography column (GE Healthcare Life Science) for the latest purification step. The fraction containing NSD1-P5C5 were combined and concentrated using Vivaspin (Sartorius) tubes with a 2 kDa cut-off. Protein purity, as well as the completion of proteolytic digestion, were checked by SDS-PAGE in each purification step.

3.3.2 Nizp1-C2HR purification protocol

His(6)MBP-tagged Nizp1-C2HR was purified using standard Ni-affinity chromatography protocol.

The soluble fraction obtained by cell lysis was loaded onto a column containing 5 ml of the Ni-NTA pre-equilibrated with 25 ml of the same buffer used to re-suspend the cellular pellet and incubate for 30 minutes at 4 °C in agitation. The column was washed with 10 CV of the buffer 20 mM TRIS-HCl pH 8, containing 150 mM NaCl, 2 mM of β-mercaptoethanol 0.1mM of ZnCl2 , 10 mM of imidazole and with 5 CV of buffer 20 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM of β-mercaptoethanol 0.1mM of ZnCl2 and 20 mM of imidazole. The recombinant protein was eluted in the buffer of 20 mM 20 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM of 2-β-mercaptoethanol 0.1mM of ZnCl2 and 500 mM of imidazole. Immediately, after elution the protein solution was added with TEV protease (about 1 mg of TEV for 30 mg of recombinant protein) in order to remove the His(6)MBP-tag. The protein solution was dialyzed over night against the buffer 20 mM TRIS-HCl pH 8, 150 mM NaCl, 2 mM of 2-β-mercaptoethanol 0.1mM of ZnCl2 .

The dialysed protein solution was loaded onto a 3 ml Ni^{2+}-NTA column and the flow-through, containing only the recombinant protein, was collected, concentrated and loaded onto a HiLoad 16/60 Superdex sixe exclusion chromatography column (GE Healthcare Life Science). The fraction containing Nizp1-C2HR were combined and concentrated using Vivaspin (Sartorius) tubes with a 2 kDa cut-off.

3.4. Nizp1-C2HR synthetic peptides.

Synthetic peptides corresponding to Nizp1-C2HR domain have been purchased from CASLO Laboratory (Lyngby, Denmark) as lyophilized acetate salts. They were used for NMR experiments and they were amidated at N-terminus and acetylated at C-terminus. Their purity were certificated of 98-99% (see Table 11 for the list of the synthetic peptides).

Peptide name	Sequence	Lenght
Nizp1-C2HR	KSYVCPNCGKIFRWRVNFIRHLRSRRE	27 residues
Nizp1-C2HR_RH	KSYVCPNCGKIFRWRVNFIRHLRSHRE	27 residues
Nizp1-C2HR_NA	KSYVCPACGKIFRWRVNFIRHLRSHRE	27 residues

Table 11. Nizp1-C2HR synthetic peptides used in binding assays

3.5. Isothermal Titration Calorimetry (ITC)

Isothermal Titration Calorimetry (ITC) is the only technique that can simultaneously determine all binding parameters in a single experiment. The methodology allows determination of the binding affinity (K_a), stoichiometry, enthalpy of the reaction in solution, without the need to use labels. Using the equations 1 and 2 the values of the dissociation constant (K_d, μM), the Gibbs energy (ΔG, kJ mol^{-1}) and the entropy (ΔS, kJ K^{-1}) can be deduced.

$$K_a = 1/K_d, \text{ [Eq.1]}$$

$$\Delta G = \Delta H - T\Delta S, \text{ [Eq.2]}$$

In the instrument there are two cells, made of a highly efficient thermal conducting material, typically gold. Each cell is equipped with a heater. A protecting adiabatic jacket avoids thermal exchange with the outside. The temperature of both cells is measured simultaneously.

One of the cell is filled with buffer and used as a reference cell. Its heater applies a constant power to maintain the temperature stable along the experiment. The other cells contains the sample dissolved in the same buffer, and its heater initially applies the same power. The ligand is loaded into a syringe which sits in a very accurate injection device. The injection device is inserted into the

sample cell containing the protein of interest.

A series of small aliquots of ligand are injected into the protein solution. If there is a binding of the ligand to the protein, heat changes are detected and measured. If the reaction is exothermic, the temperature of the cell increases and its heater reduces the applied power to compensate that with respect to the reference cell. If the reaction is endothermic, the opposite occurs. The instruments records the applied power over time ($\mu J\ sec^{-1}$). At each ligand injection the recorded power reaches a maximum (or a minimum) value and afterwards its initial value is restored, defining a peak. With the consecutive ligand injections, the number of available protein binding site is progressively decreased which is reflected by a progressive decrease of the released (or absorbed) thermal energy. Eventually the protein sample becomes saturated and the peaks display a constant size reflecting the ligand heat dilution see (Fig. 1).

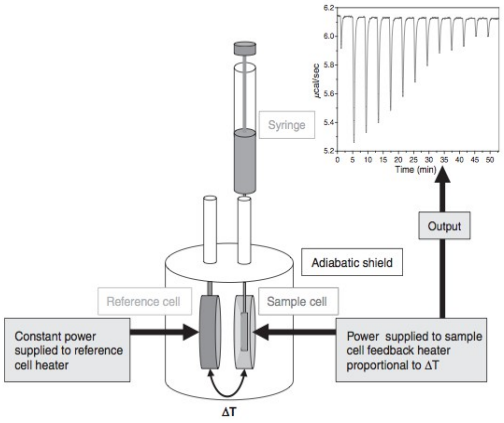


Fig. 1 Representative diagram of a typical power compensation ITC. Major features of this type of instrument such as the reference and sample cells, syringe for adding titrant, and the adiabatic shield are noted in the figure. This diagram shows an oversimplification of how the power applied by the instrument to maintain constant temperature between the reference and sample cells is measured resulting in the instrument signal.

The areas under each peaks are integrated yielding the total exchange heat per injection (kcal per mol of injectant), and data plotted in function of the ligand/protein molar ration at each inijection point. A mathematical model is fitted to the experimental data. For the proteins with a single binding site, a sigmoidal curve is fitted (see Fig. 2). The slope at the liner fraction of the curve given the K_a value. The molar ratio at which the interaction is 50% saturated gives the stoichiometry (n). the difference between the initial and the final total exchange heat gives the enthalpy (ΔH) of the reaction (see Fig. 2).

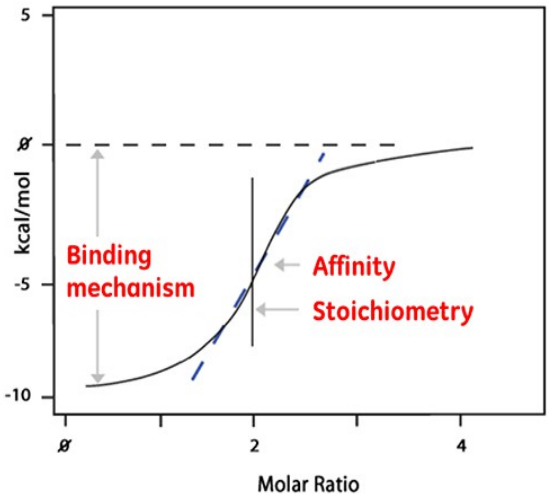


Fig. 2: Data interpretation. Heat peaks after each legand injection are integrated and a curve is fitted to the data. Several thermodynamic parameters can be deduced as well as stoichiometry of the interaction.

We used ITC to test the binding of Nizp1-C2HR domain to NSD1-P5C5.

ITC titration was performed using a VP-ITC isothermal titration calorimeter (MicroCal LLC, Northampton, MA, USA). During the night before titration, recombinant protein and synthetic peptide were dialyzed against the same buffer, containing 20 mM NaH_2PO_4/$Na2HPO_4$ pH 7.2, 150 mM NaCl, 2 mM 2-β-mercaptoethanol and 100 µM $ZnCl_2$. Titration was carried out at 23°C by injecting step by step a 0.2 mM Nizp1-C2HR protein solution into a 15 µM NSD1-P5C5. The titration parameters were stirring of 307 rpm, 28 injections of 10 µl each and 15 minutes of spacing between injections into the ITC sample cell of 1.41 ml. Control experiments were performed under identical conditions to determine the dilution heat of the titrant protein into buffer. The final data were analyzed using the software ORIGIN 7.0 (OriginLab Corp., Northampton, MA, USA), integrals of differential heats of injection were corrected by dilution heats and finally fitted by the proper binding model and nonlinear least-square regression.

3.6. NMR spectroscopy of biomolecules

In these section the basic theoretical and practical concept behind the NMR experiments that are the basis of the results presented in this thesis are summarized.

3.6.1 Basic NMR characterization of a protein: 1D and HSQC

The acquisition of a ^{1}H 1D experiment is usually the first step in the characterization of a protein sample. This experiment does not require the isotopic enrichment of the sample an gives information on the protein fold and stability. In a folded protein proteins sample a large variety of well-defined chemical environments, so signals in the spectrum are dispersed in a wide range of frequencies ranging from 11 to almost 0 ppm. Unfolded proteins yield 1D spectra with much lower

signal dispersion (see Fig. 3).

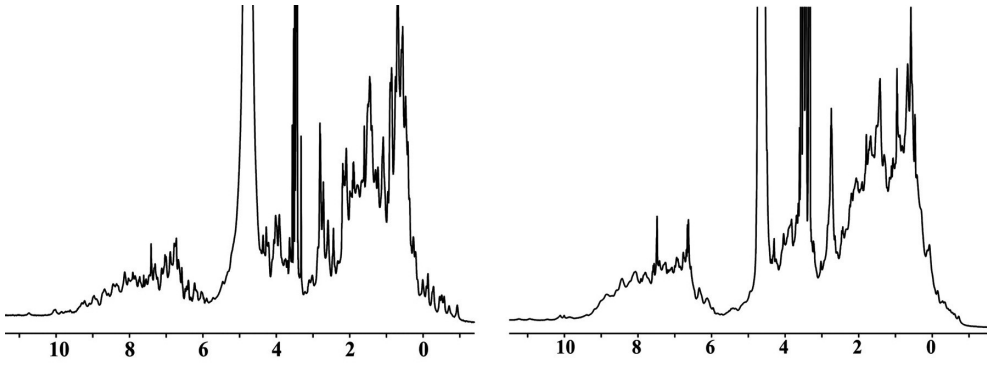


Fig. 3: Example of ^{1}H 1D NMR spectrum. In the left and right panel are reported the typical spectra of folded and unfolded protein, respectively.

Isotopic enrichment of the protein with ^{15}N allows the acquisition of the HSQC (heteronuclear single quantum coherence) experiential. This is a 2D experiment and the magnetization is transferred from hydrogen to attached ^{15}N nuclei via the J-coupling. The chemical shift is evolved on the nitrogen and the magnetisation is then transferred back to the hydrogen for detection. This is the most standard experiment and shows all H-N correlations. Mainly these are the backbone amide groups, but Trp side-chain Nε-Hε groups and Asn/Gln side-chain Nδ-Hδ2/Nε-Hε2 groups are also visible. Consequently, each peaks is considered to roughly represent the chemical environment of each amino acid in the sequence. The HSQC spectrum of a folded protein displays a well-dispersed sharped signals in the 6-11 ppm range in the ^{1}H dimension and in the 140-105 ppm range in the ^{15}N dimension. In Fig. 4 as example the ^{1}H ^{15}N HSQC NMR spectrum of human ubiquitin .

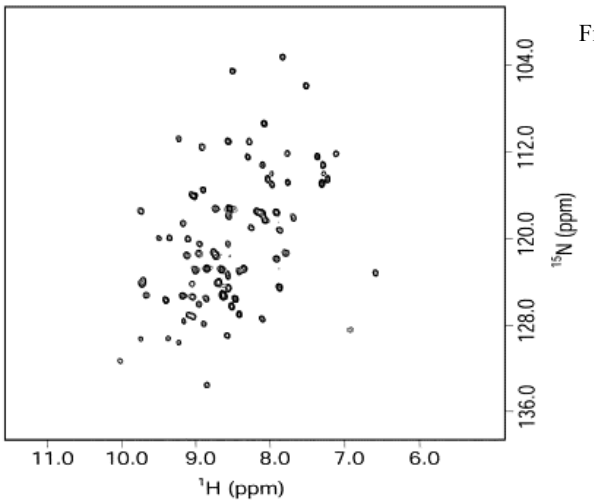


Fig. 4: ^{1}H ^{15}N HSQC NMR spectrum of human ubiquitin.

3.6.2 ^{1}H-^{15}N NMR titrations and chemical shift mapping

NMR chemical-shift mapping method is a highly sensitive tool to prove binding and to map binding sites in protein-ligand interactions (Rajagopal et al., 1997). It requires ^{15}N-labeled protein in order to record 2D ^{1}H-^{15}N HSQC spectra. Titration begins with the recording of a ^{1}H-^{15}N-HSQC spectrum of the free protein, then stepwise additions of the unlabelled titrant are performed and followed by collecting ^{1}H-^{15}N HSQC spectra after each addition. Ligand binding will cause change in the electronic environment of residues directly involved in the interaction and also those suffering structural rearrangement. These changes will be reflected in the spectra as chemical-shift displacements (CSDs). The nature of these perturbations depends on the exchange rate (k_{ex}) between the complex (k_{on}) and the free protein (k_{of}) [Eq.3]:

$$k_{ex} = k_{on} + k_{of} \text{ [Eq.3]}$$

Three main exchange regime can be distinguished depending on the relative value of k_{ex} with respect to the difference of the resonating frequencies in the bound and the free states ($\Delta\omega$):

- Fast exchange: $k_{ex} >> \Delta\omega$. This is the case of interaction that exits only during a fraction of the acquisition time. The observed chemical shift at each titration point is due to the weighted average between the concentrations of the free and bound protein forms, thus it is proportional to the binding K_D (dissociation constant). Displacement will have a maximum when protein saturation is reached (see Fig. 5).
- slow exchange regime: $k_{ex} << \Delta\omega$. In this case, residues affected show reduction of their peak intensities, while new peaks, belonging to the protein in the bound form, appear at different chemical shifts, with an increasing intensity upon ligand addition (see Fig. 5).
- intermediate exchange regime: $k_{ex} \sim \Delta\omega$. In this situation residues affected show enlargement of their peaks during titration, since it is out of the instrument resolution distinguishing between peaks belonging to the free and bound protein forms.

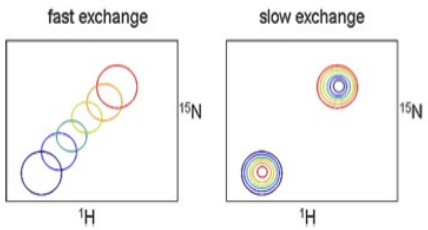


Fig. 5. The dependence of two-dimensional NMR peak shape on exchange rate. (Left) Fast exchange: peaks move smoothly from free (blue) to bound (red). In the limit of very fast exchange, peaks have the same shape throughout. As they move out of this limit, peaks may become broader when in equilibrium between free and bound, and then sharpen up again close to saturation. (Right) Slow exchange: the free peak (blue) decreases in intensity as the bound peak (red) increases. (Adopted from Williamson PM, 2013)

The HSQC titration experiments has been extensively used in this thesis to characterised binding. In a typically experiment, a samples of the titrant species (4-5 mM) and 0.130 mM ^{15}N-labeled recombinant protein were prepared in the same buffer (20 mM NaH_2PO_4/Na_2HPO_4 pH 6.3, 150 mM NaCl, 2 mM DTT, 100 μM $ZnCl_2$, 5% (v/v) D_2O). Titrations were performed at 298K and for each point we registered 1D ^{1}H and 2D ^{1}H-^{15}N-HSQC spectra on a Bruker 600MHz NMR spectrometer equipped with triple resonance cryoprobe and pulse field gradients.

Spectra were processed with Topspin software and analyzed with CCPNmr software looking for changes in chemical shifts. When binding was observed for each residue we were able to measure the ^{1}H-^{15}N chemical shift displacement (CSD) as:

$$CSD = \sqrt{\frac{\Delta\delta^2 HN + \frac{\Delta\delta^2 N}{25}}{2}} \qquad \text{[Eq.4]}$$

where, for each peak, Δδ is the difference of the chemical shift values between the ^{1}H-^{15}N-HSQC spectrum of the free protein and the ^{1}H-^{15}N-HSQC spectrum of each titration point. Residues affected by binding were supposed to be those showing a CSD larger than the average plus one standard deviation, and they were mapped on the three-dimensional protein structure in order to display the interaction surface.

3.6.3 Backbone assignment

The interpretation of the HSQC titrations experiment required the assignment of each resonance in the spectrum to a particular residue in the protein sequence. This is usually archived using a 3D pair of the type of HNCα/ CβCα(CO)NH that allows the assignment of backbone ^{15}N and ^{15}C resonances.

In the following I will briefly describe the 3D NMR experiments used for backbone assignment.
In the HNCα experiment, the magnetisation is transferred from ^{1}H to ^{15}N and then via the N-Cα J-coupling to the ^{13}Cα and then back again to ^{15}N and ^{1}H hydrogen for detection. The chemical shift is evolved for $^{1}H^{N}$ as well as the $^{15}N^{H}$ and 13Cα (see Fig. 6). Since the amide nitrogen is coupled both to the Cα(i) of its own residue and that of the preceding residue Cα(i-1), both these transfers occur and peaks for both Cα are visible in the spectrum. However, the coupling to the directly bonded Cα is stronger and thus these peaks will appear with greater intensity in the spectra. The spectrum produced has 3 dimensions: a proton axis (x), a 15N axis (z) and a 13C axis (y). Each strip is

extracted from the 3D spectrum at a specific nitrogen frequency and at the ^{1}H frequency of its amide proton. The sequential assignment is made by matching the alpha carbon chemical shifts of the residues (i) and (i-1), (see Figure 6).

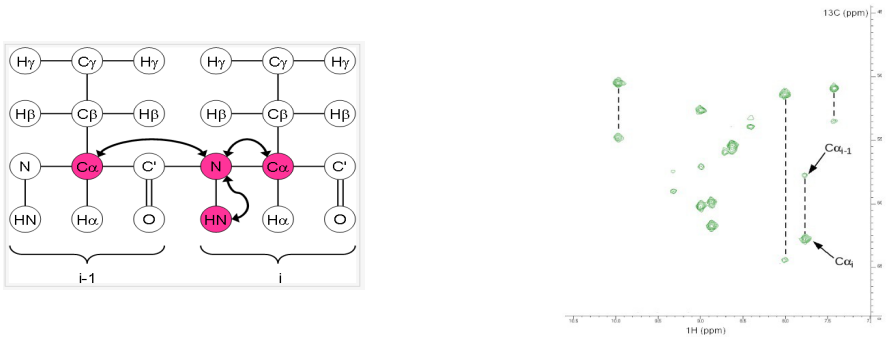


Fig. 6: HNCα experiment. On the left the schematic representation of the magnetization transfers between backbone nuclei, the arrow indicates the magnetization transfer. On the right an example of NH strip: two Cα peaks should be visible, the stronger one belongs to the same residue (Cα(i)) and the weaker one belongs to the preceding residue (Cα(i-1)).

The HNCα experiment is coupled to the CβCα(CO)NH experiment.

In these case the magnetisation is transferred from $^{1}H\alpha$ and $^{1}H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$, respectively, and then from $^{13}C\beta$ to $^{13}C\alpha$. From here it is transferred first to ^{13}CO, then to $^{15}N^{H}$ and then to $^{1}H^{N}$ for detection. The chemical shift is evolved simultaneously on $^{13}C\alpha$ and $^{13}C\beta$, so these appear in one dimension. The chemical shifts evolved in the other two dimensions are ^{15}NH and $^{1}H^{N}$. The chemical shift is not evolved on ^{13}CO (see Fig 7). In this experiment the NH group of each amino acid is coupled to the Cα and Cβ of the preceding residue in the protein sequence (named Cα(i-1) and Cβ(i-1), respectively). The sequential assignment is made by matching the alpha and beta carbon chemical shifts (see Figure 7).

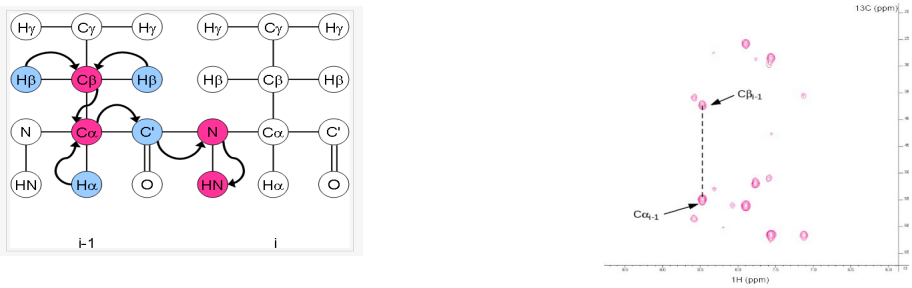


Fig. 7: CβCα(CO)NH On the left the schematic representation of the magnetization transfers between backbone nuclei, the arrow indicates the magnetization transfer. On the right an example of NH strip. Each strip contains two peaks: the Cα and the Cβ of the reside preceding the NH group (named Cα(i-1) and Cβ(i-1), respectively).

The Cα and Cβ chemical shift are strongly depending on the amino acid type. The most common values for each amino acid type are extensively characterized, and they are used to identity

particular sequence segments and math then to strip sequences. Therefore a particular ^{1}H- ^{15}N resonance in the HSQC spectrum can be assigned to a particular residue in the protein sequence

The HNCα and CβCα(CO)NH was performed using a sample of ^{13}C -^{15}N NSD1-P5C5 bound to unlabelled Nizp1-C2HR (ratio 1:1) and a sample of ^{13}C -^{15}N Nizp1-C2HR bound to unlabelled NSD1-P5C5 (ratio 1:1) at the final concentration of 0.6-0.7mM in buffer phosphate at pH 6.3, with 150mM NaCl and 2 mM of β-mercaptoethanol.

3.7. Protein Structure determination

The determination of a protein structure by NMR is based on distance restrains that place a particular atom in a defined position of the three-dimensional space relative to all other atoms in the molecule. In oder to observe these restraints, in a typically experiment, the magnetization is transferred thought the space in a phenomenon known as dipolar coupling. The transfer occurs between nuclei that are spatially close (< 5 Å) and it produces a measurable signals, called Nuclear Overhauser Effect (NOE). The NOE intensity is inversely proportional to the sixth power of the distance between these nuclei, so it can be used as measurement of this distance.

To determine the 3D structure of the protein by NMR, NOE distance restrains between the protones in the protein are observed using NOESY (NOE Spectroscopy) experiments. In the multidimensional NOESY experiments, the NOE effect gives raise to the correlation signals whose integration is a direct measurement of the inter-atomic distance. Homonuclear 2D ^{1}H - ^{1}H and heteronuclear 3D ^{15}N- NOESY spectra are used to acquire these restrains. Many intra-residue signals are usually observed, as well as signals with residues close in the sequence and in the space (see Fig. 8).

NOESY spectra are usually very crowed with signals and individual amino acids and hard to identify, as intra and inter-residue signals are mixed. To help with the assignment, TOCSY (Total Correlation Spectroscopy) experiments are acquired. In this experiment the magnetisation is transfer through the bound between nuclei connected by a limited number of covalent bonds. The coupling between two particular nuclei is described by field-independet coupling constant J (Hz) that is much smaller than the chemical shift. The spectrum displays the correlation peaks between all the atoms in a spin system. Each amino acid's said chain has a characteristic spin system, therefore all the nuclei corresponding to a single residue can be identified and assigned (see Fig. 8).

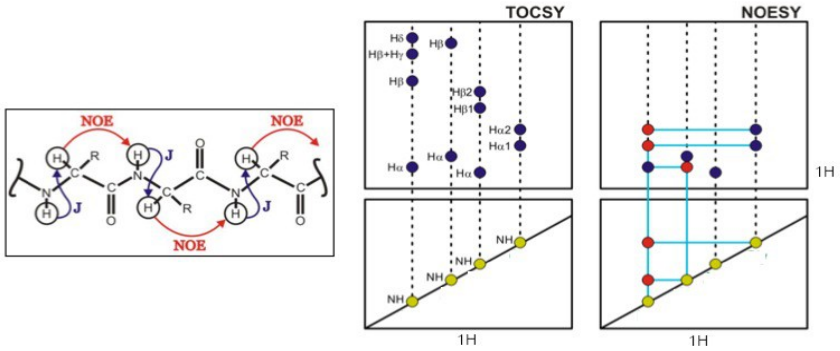


Fig. 8: Example of TOCSY and NOESY spectra. On the left the general chemical formula of a tripeptide is shown, while on the right there are part of TOCY and NOESY spectra. In the 2D ^{1}H - ^{1}H TOCSY the peaks on the diagonal are the NH group of the amino acids that are correlated with the nuclei in the own spin system (blue circles). In the 2D ^{1}H - ^{1}H NOESY there are both intra-residue correlations (blue circles), as well as inter-residue correlations (red circles), which allows us to find which residue is next to which in the peptide chain.

3.7.1 Resonances assignment

NMR samples contained typically 0.8 mM of unlabeled Nizp1-C2HR in 150 mM NaCl, 20 mM KH_2PO_4/K_2HP0_4 (pH 6.3) and 50 mM $ZnCl_2$. All NMR experiments were performed at 208C on Bruker Avance 600 Ultra Shield TM Plus 600 MHz spectrometersequipped with triple resonance cryoprobes and pulsed field gradients. Spectra were processed with Topspin and and analysed with CCPNMR (Vranken et al 2005). We acquired ^{1}H-^{1}H TOCSY experiments for spin systems identification and ^{1}H-^{1}H NOESY 2D spectra acquired in H_20 and in D_20, ^{15}N-HSQC NOESY to collect NOEs that were assigned manually.

3.7.2 Determination of the tautomeric form of histidines

NMR spectroscopy is used to the study of the histidine protonation state of the His residues of Nizp1-C2HR domain. ^{15}N NMR spectroscopy has been used extensively to study the histidine imidazole ring, as nitrogen assumes a typical chemical shift value according to the protonation state of the NH group (Figure 1, Pelton et al, 1993). By combining ^{15}N chemical shifts and J-couplings of the imidazole ring it is possible to construct the cross-peak patterns expected in a ^{1}H-^{15}N-HSQC spectrum for histidine in the three different protonation states (Figure 1, Pelton et al, 1993).

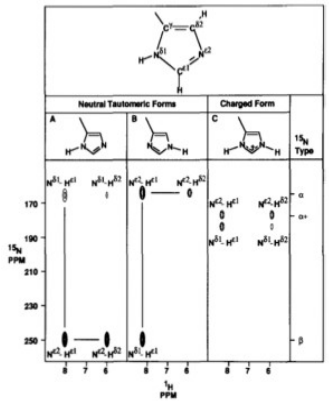


Figure 9. Correlation schemes according to the imidazole protonation states. α, α+ and β indicate the chemical shifts of the protonated sp3, protonated sp2 and non protonated sp2 nitrogen, respectively. Taken from Pelton et al., 1993.

For the Nδ1-H tautomer (Figure 9) two strong correlations (Nε2-Hε1 and Nε2-Hδ2) at nitrogen chemical shift near 250 ppm and a weaker Nδ1-Hε1 correlation at nitrogen chemical shift around 168 ppm are expected. In addition a very weak Nδ1-Hδ2 correlation peak could be visible.
For the Nε2-H tautomer (Figure 9) the pattern is inverted. A single strong Nδ1-Hε1 correlation at nitrogen chemical shift near 250 ppm and slightly weaker Nε2-Hε1 and Nε2-Hδ2 correlations at nitrogen chemical shift near 168 ppm are expected.
The correlation pattern for protonated, cationic histidines (Figure 9) is very different, therefore it is easy to identify. Because the chemical shifts of Nδ1 and Nε2 are nearly the same, four correlations peaks (Nε2-Hε1, Nε2-Hδ2, Nδ1-Hε1 and Nδ1-Hδ2) occur around a nitrogen chemical shift of 176 ppm. ^{1}H-^{15}N long range HSQC experiment was performed on a sample 0.5 mM of ^{15}N-Nizp1-C2HR in in 20 mM NaH_2PO_4/Na2HPO$_4$ pH 6.3, 150 mM NaCl, 2 mM DTT, 50 uM $ZnCl_2$, 10% (v/v) D_2O.

3.7.3 Structure calculation

Structure of Nizp1-C2HR was calculated using ARIA2.2 (Rieping et al 2007) in combination with CNS (Brunget et al. 1998) using the experimentally derived restraints (Table 1). All NOEs were assigned manually and calibrated by ARIA2.2. The standard procedure to determine the solution structure of a protein from NMR data is to minimize an objective function that incorporates experimental data (intensities of assigned NOE cross-peaks) and physical knowledge (physical and chemical properties of the protein nuclei and chemical bonds). A total of eight iterations (20 structures in the first six iterations) were performed: 200 structures were computed in the last two iterations. The ARIA2.2 default water refinement was performed on the 20 best structures of the final iteration. From initial calculations performed without a zinc ion, it was evident that the zinc coordination site was defined by the two thiol groups of Cys407, Cys410 and by the Nδ1 atom of His423. A zinc ion was included in the subsequent and final calculations, assuming a coordination with tree ligand.

3.8. Crystallization of NSD1-P5C5:Nizp1-C2HR complex

The determination of the crystal structure of a biological macromolecule requires the preparation of crystals suitable for x-ray analysis, usually single crystals which are characterized by a high degree of homogeneity, long-range order and suitable size, with the optimum being somewhere between 10^{-4} and 10^{-3} mm^3. Crystallization requires bringing the

macromolecule to supersaturation.

The sample should therefore be concentrated to the highest possible concentration without causing aggregation or precipitation of the macromolecule (usually 2-50 mg/ mL). Introducing the sample to precipitating agent can promote the nucleation of protein crystals in the solution, which can result in large three-dimensional crystals growing from the solution. There are two main techniques to obtain crystals: vapor diffusion and batch crystallization. In vapor diffusion, a drop containing a mixture of precipitant and protein solutions is sealed in a chamber with pure precipitant (Figure 10A).

The batch method relies on bringing the protein directly into the nucleation zone by mixing protein with the appropriate amount of precipitant (Figure 10B). In both the methods, the concentration of the precipitant in the reservoir is higher than in the drop. Due to the concentration gradient, the water from the drop will evaporate towards the reservoir, thus lowering the volume of the drop and increasing the protein concentration. At some moment, the concentration of the protein will reach supersaturation, and if all other parameters like pH, ionic strength, type of buffer, additives, etc. are appropriate, the protein may start to crystallize (Figure 10B, Point B).

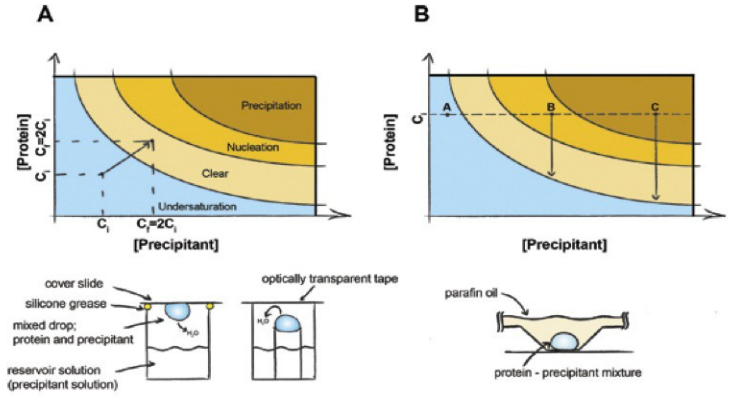


Fig 10:**The principle of protein crystallization.** In a vapor diffusion experiment (A) equal volumes of precipitant and protein are present in the drop. Water will diffuse out and both the precipitant and protein concentration will be doubled until equilibrium is achieved between the drop and the reservoir solution. In batch crystallization (B) the precipitant and protein concentration do not change during the experiment. Point A - Protein stays undersaturated no crystals can be formed, Point B - Protein nucleation occur, crystals began to form and the concentration of protein in solution drops to saturation. Point C - Protein precipitates, but crystals may still grow.

3.8.1 Crystallization screening

In order to determine the exact conditions for crystallization, we used the vapor diffusion method and tested a large number of conditions to find the most optimal combination of all parameters (pH, ionic strength, type of buffer, additives, etc...). To determine the appropriate sample concentration for the crystallization screening we carry out a pre-crystallization test (PCT, from Hampton

Research) which contains fourth conditions: reagent A1 (0.1 M TRIS hydrochloride pH 8.5, 2.0 M Ammonium sulfate), A2 (0.1 M TRIS hydrochloride pH 8.5, 0.2 M Magnesium chloride hexahydrate, 30% w/v Polyethylene glycol 4,000), reagent B1 (0.1 M TRIS hydrochloride pH 8.5, 1.0 M Ammonium sulfateand) and reagent B2 (0.1 M TRIS hydrochloride pH 8.5, 0.2 M Magnesium chloride hexahydrate, 15% w/v Polyethylene glycol 4,000). Then, we used two commercial kits, to screen the crystallization conditions: Index (96 conditions) and the Crystal screen (50 different conditions) both from Hampton Research.

3.9. HADDOCK web server

The HADDOCK web server was used to calculate the initial structure of the complex formed by NSD1-P5C5 and Nizp1-C2HR. HADDOCK (High Ambiguity Driven protein-protein DOCKing) is an information-driven docking approach for the modelling of biomolecular complexes. It is an approach that makes use of experimental data, such as the chemical-shift perturbations resulting from NMR titrations. The information on the interacting residues is introduced as Ambiguous Interaction Restraints (AIRs) to drive the docking. The interacting residues are distinguished between “active” and “passive” residues. In the case of NMR titration data, the active residues correspond to all residues showing a significant chemical shift perturbation upon complex formation as well as a high solvent accessibility in the free form protein. The passive residues correspond to the residues that show a less significant chemical shift perturbation and/or that are surface neighbours of the active residues and have a high solvent accessibility. The docking protocol requires the PDB files of the free proteins, and after calculation, the structures are clustered according to their intermolecular energy, that is sum of electrostatic, van der Waals, desolvation and AIR energy terms (HADDOCK score). The best cluster was selected on the basis of the lowest HADDOCK score and Z-score (Dominguez C et al, 2003).

4. Results

4.0 Introduction

The recruitment of NSD1 on the chromatin is mediated by Nizp1, through the interaction between NSD1-P5C5 and Nizp1-C2HR. The complex is probably responsible of the gene silencing. In order to investigate this mechanism we characterized by biophysical techniques (NMR titrations and ITC experiments) the binding between the two protein modules involved in the interaction. To perform this study we solved the structures of both the tandem domain NSD1-P5C5 and of Nizp1-C2HR. The NMR structure of NSD1-P5C5 was solved in a previous work, where I contributed to optimize the expression and purification of the construct (PhD Thesis, D. Spiliotopoulos). In the next paragraph, I will describe the structure of this tandem domain.

4.1 NMR structural characterization of NSD1-P5C5.

The tandem domain NSD1-P5C5 (residues Glu2115-Asp2206) is composed by two PHD finger domains that adopt a compact fold and form an unique structural units stabilized by four zinc ions coordinated by Cys-Cys-Cys-His motive (see Fig.1A and Fig.2A).

The first module (NSD1-PHD5, or NSD1-P5) comprising residues Glu2115-His2162, assumes a canonical PHD finger fold which contains a two-stranded antiparallel β-sheet (β1, residue Leu2130-Ser2132 and β2 residues Val2141-Tyr2142) followed by a short α-helix (Ala2144-Leu2147, α1) and the loop L3 (Leu2147-Cys2159). The α-helix (Cys2159-His2162, α2) is directly linked to the second PHD domain (NSD1-C5HCH or NSD1-C5) comprising residues Gln2163-Asp2206. This module is a degenerate PHD finger domain which contains two short antiparallel beta sheets, β3-β4 (Ser2173-Phe2174; Ser2181-Phe2182) and β5-β6 (Leu2191-Ser2194; Gly2198-Cys2202) that are arranged orthogonally to each other.

The single domain fold is stabilized by a network of hydrophobic interactions involving Tyr2142, Leu2147, Leu2149, Trp2157 in PHD5 and Phe2182, Leu2191, Ile2193 and Leu2200 in NSD1-C5HCH (see Fig.2B). Hydrophobic residues including Phe2225, Tyr2142, Trp2261, Phe2274, Pro2279, located at the protein interface, establish inter-domain interactions that stabilize the protein fold and create a hydrophobic groove surrounded on both sides by negatively charged residues (see Fig.1B and Fig.2C).

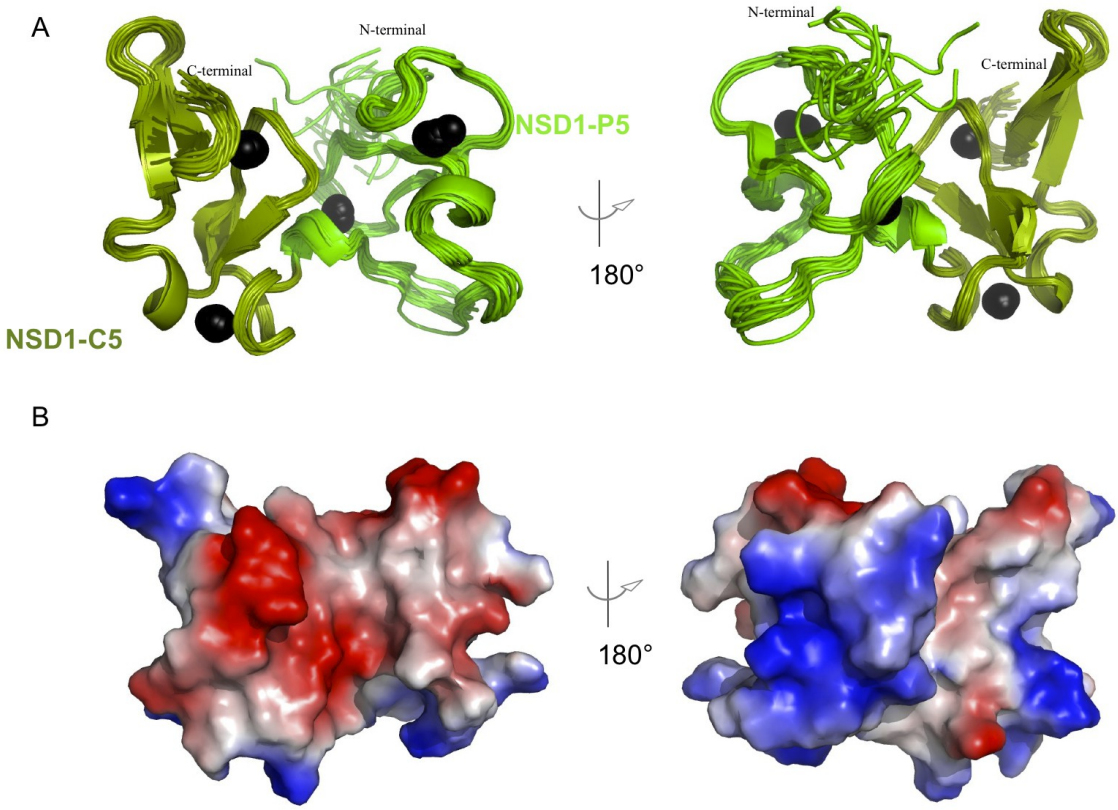


Fig.1. NSD1-P5C5 structure. (A) Cartoon representation of the NMR ensemble of the 15 lowest energy structures of NSD1-P5C5. The NSD1-P5 and the NSD1-C5 domain are shown in light green and dark green, respectively. (B) Electrostatic surface of NSD1-P5C5. The structure in the left side of the panel has the same orientation as in panel A. The structure in the right side of the panel is related to the former by a 180° rotation as indicated by the grey curved arrow. Blue and red indicate positive and negative electrostatic surface potentials, respectively.

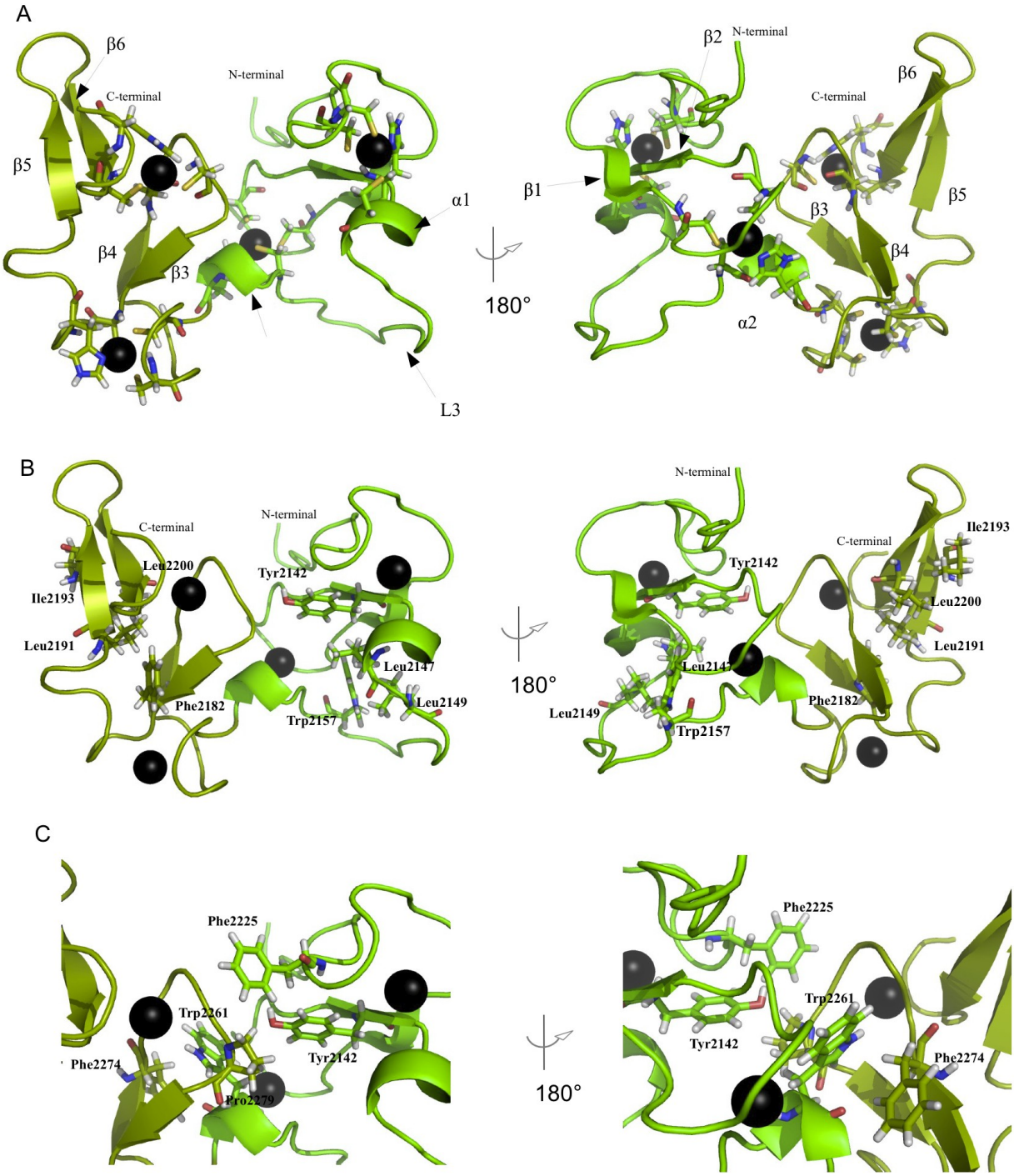


Fig.2. NSD1-P5C5 structure (A) The secondary structure elements are indicated by labels, while the zinc coordinating residues are shown in sticks. The zinc atoms are shown as black spheres. (B) The side chains of the residues forming NSD1-P5C5 hydrophobic core are shown as sticks (C) The hydrophobic residues that form the inter-domain interactions are shown in sticks.

Lucio-Eterovic *et al* 2010 proposed a functional model in which NSD1 mediates the repression of several gene involved in cellular growth and differentiation and the abrogation of this activity might

contribute to pathological conditions. The Sotos syndrome is an overgrowth disorder, probably linked to the loss of NSD1 transcriptional repression activity (Tatton-Brown e Rahman 2007). Interestingly, about two-thirds of pathological point mutations target the NSD1-P5C5 domain suggesting that this domain is functionally relevant in the pathology. In the next section, I will describe the effect of these pathological point mutations on the protein fold.

4.1.1 NSD1-P5C5 in Sotos Syndrome: structural analysis of NSD1-P5C5 pathological point mutations

Most of the Sotos pathological point mutations target the zinc-coordinating residues (Cys and His) with the exception of Tyr2142Asn, Arg2152Gln and Phe2182Ile that are outside the zinc coordinating sites. The cysteine mutants usually compromise the metal coordination a promoting the protein unfolding.

In order to characterize the impact on the protein fold of the non-cystein pathological point mutants, seven Sotos mutants have been generated (Tyr2142Asn, His2143Glu, His2143Tyr, Arg2152Gln, His2162Arg, Phe2182Ile and His2206Arg) and analysed by NMR spectroscopy comparing their 1H-1D spectra with the wild-type spectrum (see Fig3,4,5).

The 1H-1D spectra of Tyr2142Asn and Phe2182Ile show broad lines and reduced peak dispersion indicating that the mutations targeting the hydrophobic core of both domains strongly compromise the tandem domain fold (see Fig. 3B,C).

The His2143Glu and His2143Tyr substitutions are detrimental for the protein fold and both the spectra show that the protein is unfolded (see Fig. 3D,E) .

Unexpectedly, the His2162Arg and His2206Arg substitutions, that target the second and the forth zinc coordination site have a good chemical shift dispersion with respect to the 1H-1D spectra of the other mutants (see Fig. 3A, 4A, 5A). However, the broad line width, and an increase in intensity of the random coil signals (i.e. between 7.6 and 8.4 ppm) is indicative of a partial destabilization of the tertiary fold. In order to better investigate the fold of these two pathological mutants we collected 2D ^{1}H ^{15}N - HSQC spectra of the protein domains and we compared them with the spectrum of the wild type protein. The ^{1}H ^{15}N -HSQC spectrum His2162Arg is characterized by a large cluster of overlapped peaks in the middle of the spectrum with several peaks missing, indicating the presence of significant unstructured elements in the protein. On the contrary, the 2D spectrum of His2206Arg is a typical spectrum of a folded protein with well-dispersed peaks that can be distinguished and assigned. The residues whose chemical shift is affected by point mutants are the one that are close in sequence and in space to His2205, i.e. in the region comprising the strand

β5-β6 as shown in Fig. 5.

Finally, the 1H-1D and H ^{15}N -HSQC spectra of Arg2152Gln mutant are very similar to the wild type tandem domain (see Fig. 6). The 2D spectrum displays well dispersed signals typical of a folded protein: in the structure the side-chain of Arg2152 is totally solvent exposed and probably this residue has a functional rather than a structural role.

4.1.2 Summary

Conceivably NSD1-P5C5 Sotos pathological point mutations, lead to the loss of the protein domain function. In this thesis we studied the fold of seven Sotos mutations, five of which affect hydrophobic and histidine coordinating residues strongly compromising the protein fold. We observed that both His2162Arg and Arg2152Gln do not alter the protein fold.

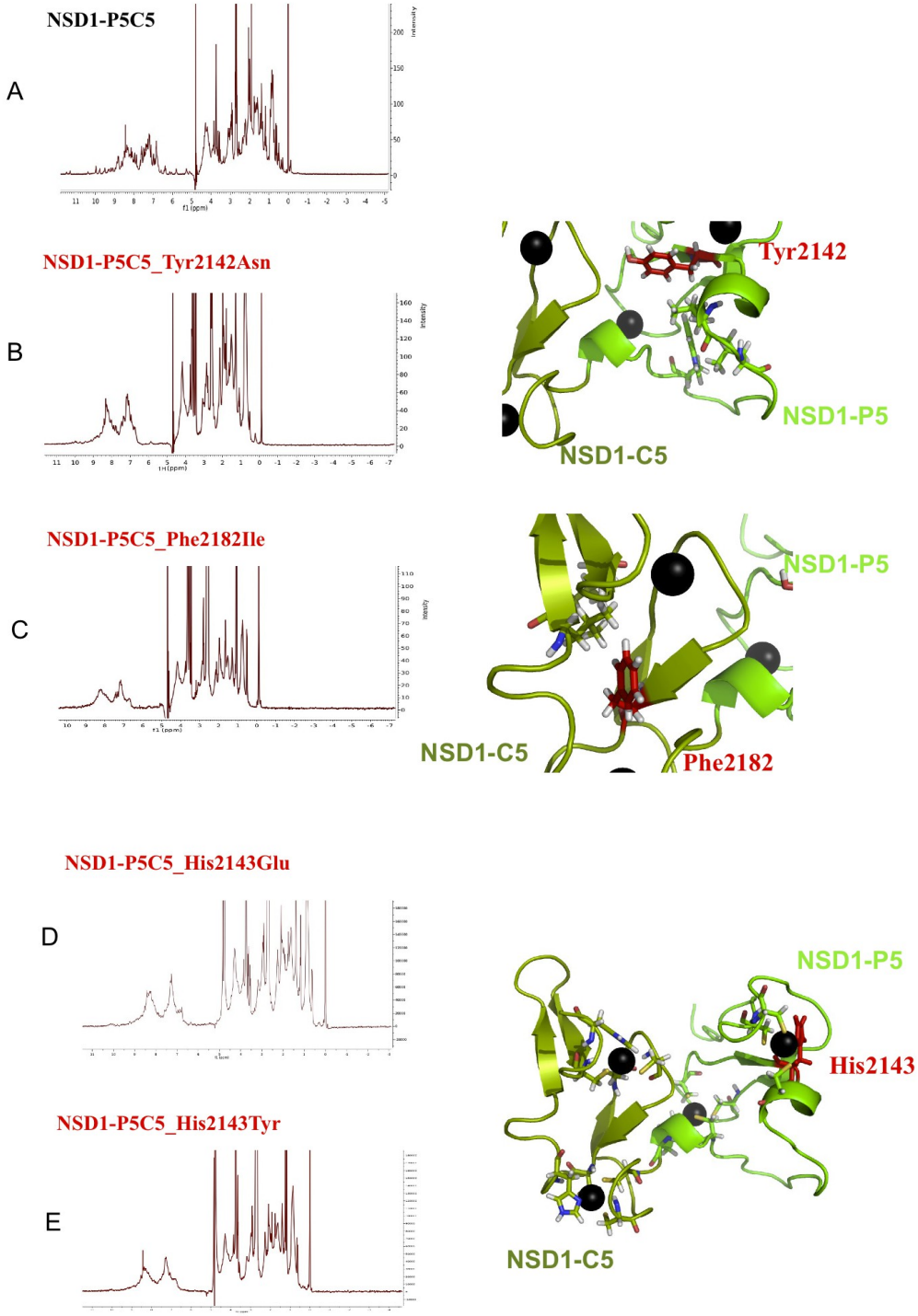


Fig. 3 Structural characterization of NSD1-P5C5 Sotos mutants. (A) 1H-1D spectrum of wild-type NSD1-P5C5. In the Fig. 3B,C,D,E on the left are shown the 1H-1D spectra of the NSD1-P5C5 Sotos mutants, on the right the residues affected by the mutation are highlighted in red on the structure.

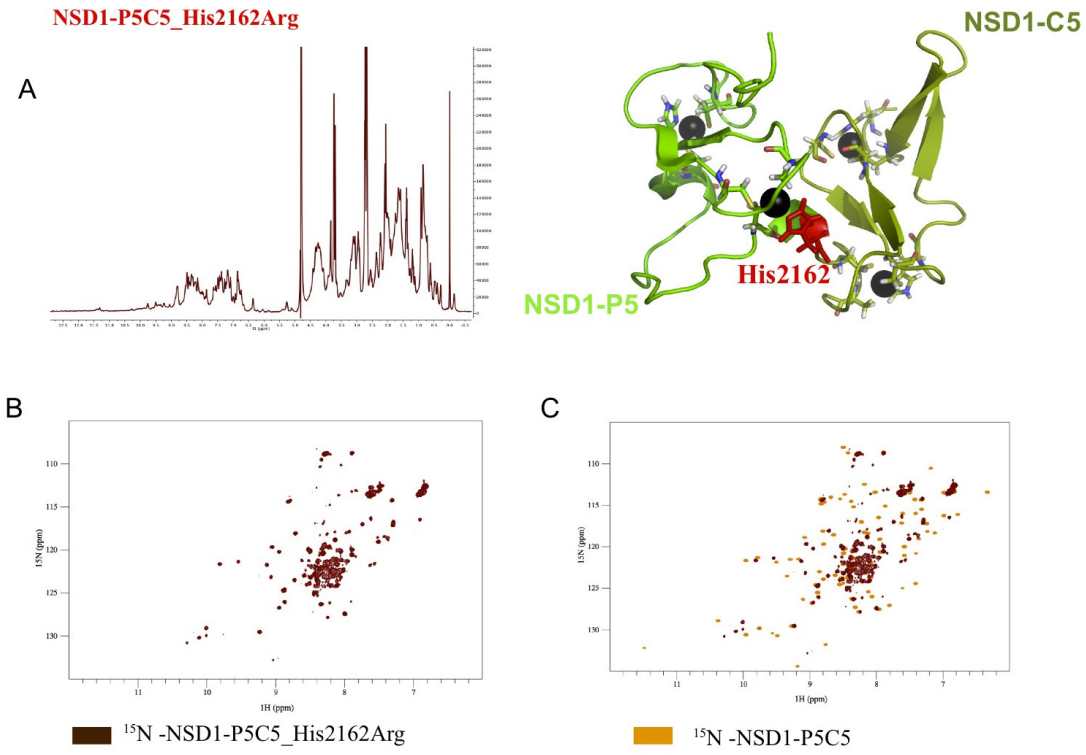


Fig. 4 The NSD1-P5C5_His2162Arg fold was analysed by NMR spectroscopy (A) 1H-1D spectrum of His2162Arg on the left, on the right is reported the NSD1-P5C5 structure with His2162 residue shown in red sticks. (B)^{1}H ^{15}N-HSQC spectrum of NSD1-P5C5_His2162Arg (C) Superposition of the ^{1}H ^{15}N-HSQC spectrum of NSD1-P5C5_His2162Arg (brown) and ^{1}H ^{15}N-HHSQC spectrum of NSD1-P5C5 (orange).

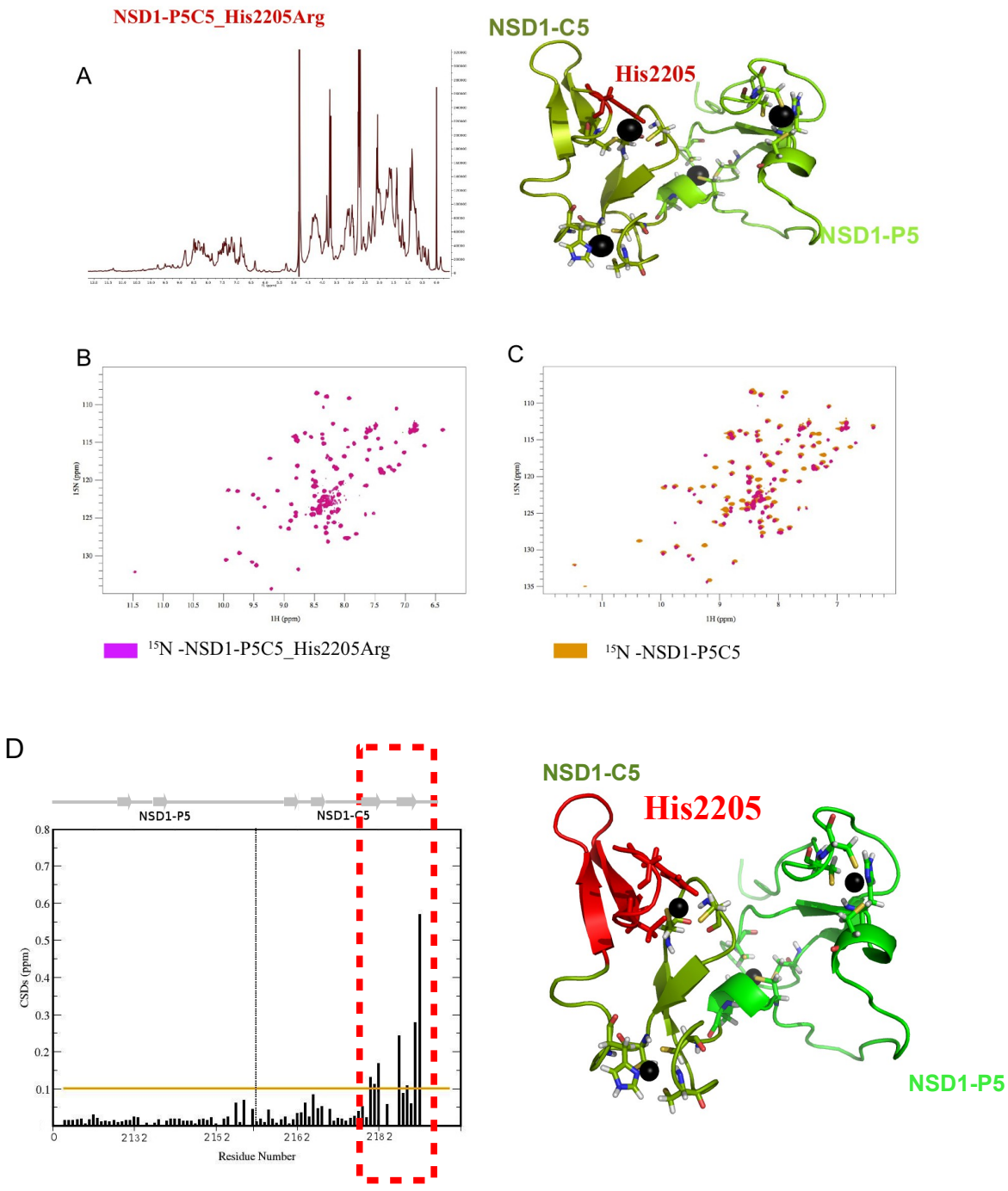


Fig. 5 The NSD1-P5C5_His2205Arg fold was analysed by NMR spectroscopy (A) 1H 1D spectrum of NSD1-P5C5_His2205Arg on the left, while on the right is reported the NSD1-P5C5 structure with the His2205 residue shown in red sticks. (B) ^{1}H ^{15}N-HSQC spectrum of NSD1-P5C5_His2205Arg (C) Superposition of the 1H 15N-HSQC spectrum of NSD1-P5C5_His2205Arg (pink) and ^{1}H ^{15}N-HSQC spectrum of NSD1-P5C5 (orange). (D) The histogram shows the difference of chemical-shift, between the wild-type and mutant domain, for each residue of the NSD1-P5C5_His2205Arg domain. Residues displaying values over 0.1 ppm (orange line) were considered affected by the pathological mutation.

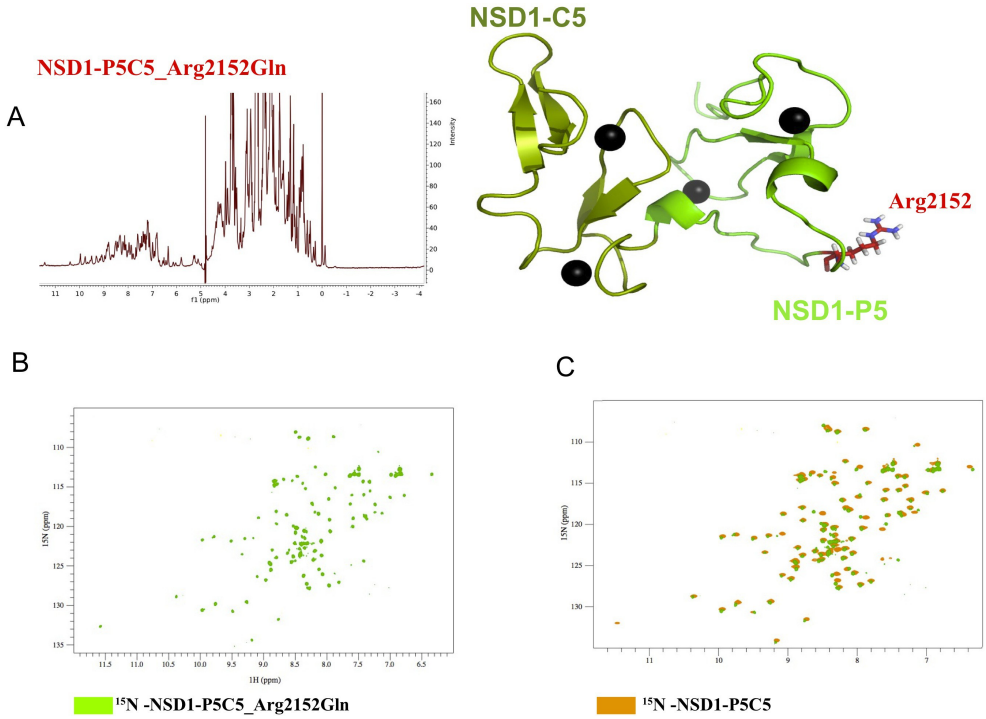


Fig. 6 Structural analysis of NSD1-P5C5_Arg2152Gln Sotos mutant (A) 1H 1D spectrum of NSD1-P5C5_ Arg2152Gln on the left, on the right is reported the NSD1-P5C5 structure with Arg2152 shown in red sticks. (B) ^{1}H ^{15}N-HSQC spectrum of Arg2152Gln (C) Superposition of the $^{1}H^{15}N$-HSQC spectrum of NSD1-P5C5_Arg2152Gln (light green) and 15N-HSQC spectrum of NSD1-P5C5 (orange).

4.2 NMR structural characterization of Nizp1-C2HR

4.2.1 Nizp1-C2HR binds Zn^{2+}

Nizp1-C2HR belongs to the family of zinc finger domains in which the fourth zinc-coordinating residue is missing. Generally, in the zinc finger proteins, the metal ion is tetra-coordinated by two Cys and two His residues (Cys2His2) or three Cys and one His residue (Cys2HisCys). Nizp1-C2HR contains two Cys residues (Cys407 and Cy410) and a single His residue (His423) that should constitute the zinc binding motif.. Notably the fourth coordinating position is occupied by an arginine (Arg427), an usual residue for zinc coordination.

Using a synthetic peptide corresponding to Nizp-C2HR domain (residues Glu397-Lys434), we analysed the domain fold by NMR spectroscopy. We collected ^{1}H ^{1}D spectrum on a protein samples that had been either depleted of zinc ions, adding EDTA and then dyalising agains NMR phosphate buffer (see Fig.7, red spectrum), or supplemented with $ZnCl_2$ in molar ratio of 1:1 (see Fig.7, green

spectrum). The comparison of the two spectra indicates that in the presence of zinc ions new peaks appear in the spectrum, for example between 6 and 9.5 ppm, and their dispersion increases, suggesting that the domain binds Zn^{2+}, adopting its native fold in presence of a equimolar amount of metal ions.

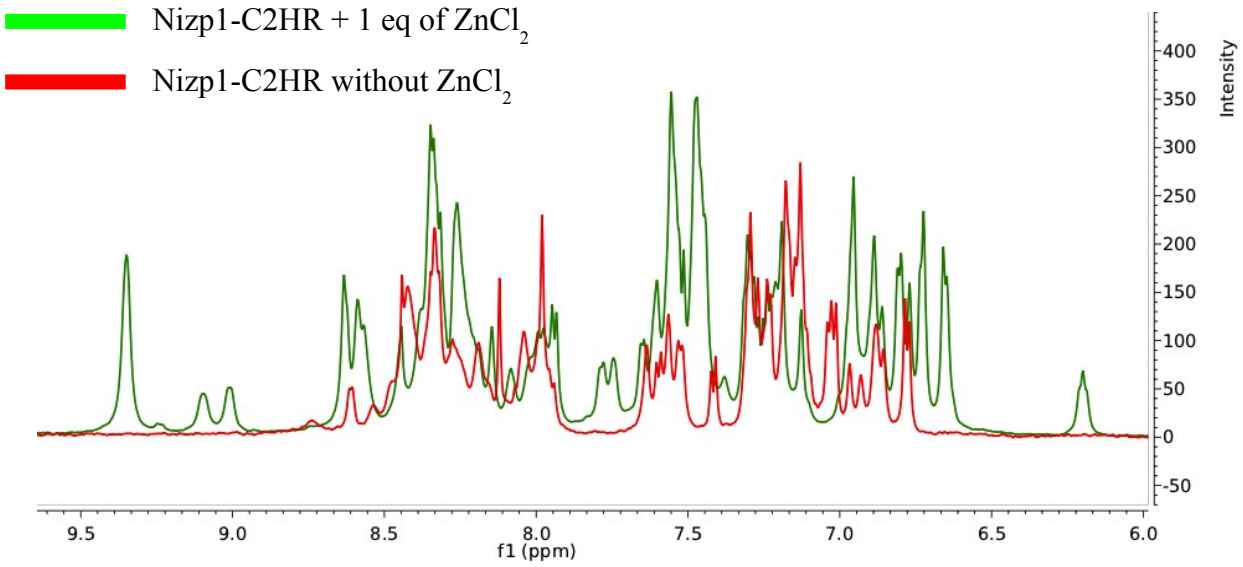


Fig. 7 Nizp1-C2HR is a zinc binding motif. The green spectrum is collected using a protein sample of 0.2mM in presence of equimolar amount of $ZnCl_2$ at pH 6.3, while the red spectrum is collected using a protein sample without $ZnCl_2$ at pH 6.3.

4.2.2 Optimization of Nizp1-C2HR recombinant protein preparation and labelling

In order to perform NMR titrations and 3D heteronuclear NMR experiments we generated a recombinant domain corresponding to Nizp1-C2HR module (37 residues, ~5kDa). Adequate expression conditions were found using a $His_{(6)}$-MBP tag construct and the purification of mgs amount of Nizp1-C2HR module was achived as described in Material and Method. The fold of the Nizp1-C2HR was checked by ^{1}H ^{1}D NMR experiment and it was compared with the spectrum of the synthetic peptide. The two peptides are characterized by a very similar peak dispersion suggesting that they adopt the same fold (see Fig. 8).

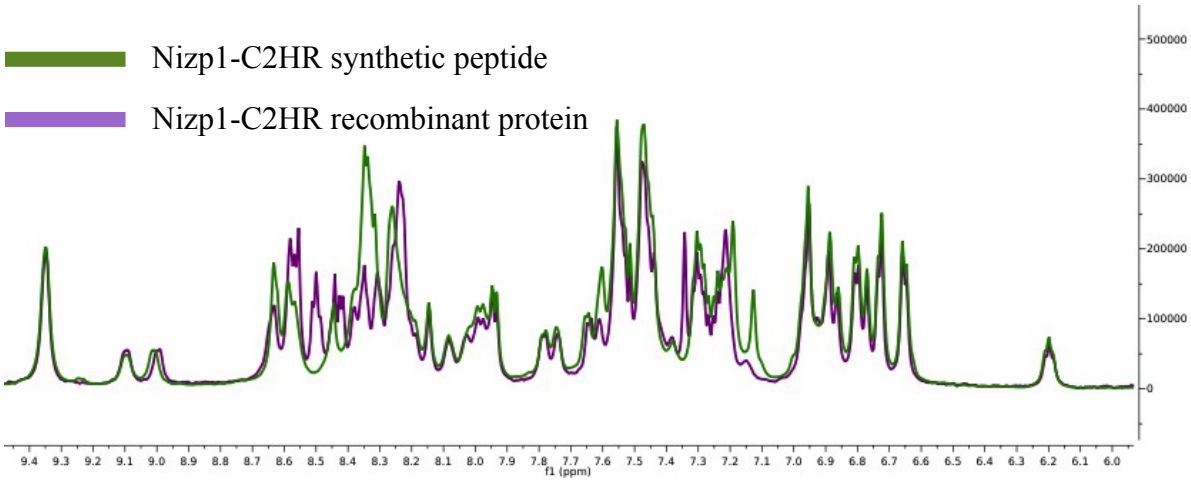


Fig. 8 Structural characterised of the Nizp1-C2HR recombinant protein fold by NMR spectroscopy. The ^{1}H ^{1}D NMR spectrum of Nizp1-C2HR synthetic peptide (green) is superimposed to the spectrum of the recombinant protein domain (violet).

A sample with labelled in ^{15}N was produced to acquire the 2D ^{1}H ^{15}N-HSQC of the protein domain, while a sample labelled with ^{15}N- ^{15}C was used to collect HNCα 3D NMR experiment to assign its backbone resonances. The 2D ^{1}H ^{15}N-HSQC of Nizp1-C2HR is reported in (see Fig. 9).

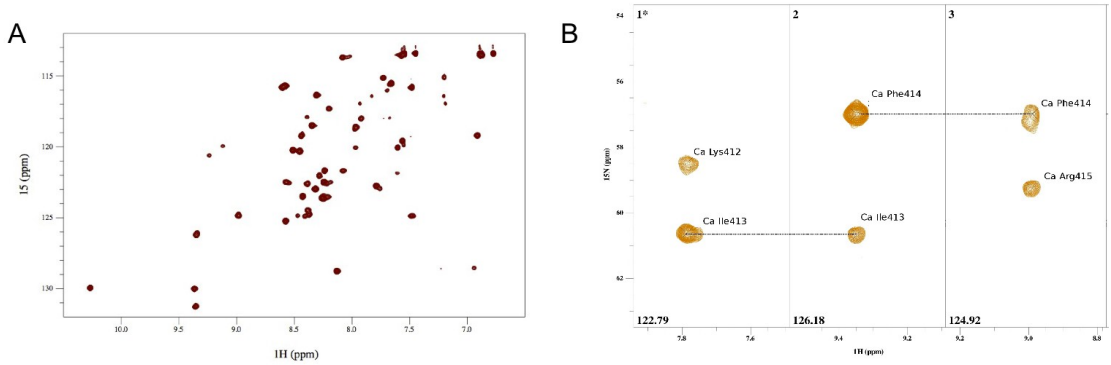


Fig. 9 Nizp1-C2HR backbone assignment (A) 2D ^{1}H ^{15}N-HSQC of Nizp1-C2HR (B) Selected strips extracted from of an HNCα 3D NMR experiment showing the assignment of three sequential residues of Nizp1-C2HR. Initially the protein sequence is know, but the assignment of HSQC is not. In the HNCα experiment each amide HN is correlated to its own Cα(i) and to the Cα of the preceding residue (i-1).

4.2.3 Structure of Nizp1-C2HR

We determined the structure of Nizp1-C2HR analyzing both homonuclear (^{1}H-^{1}H TOCSY ^{1}H-^{1}H NOESY) and heteronuclear multidimensional experiments (3D ^{1}H ^{15}N- SQC NOESY). From the anlysis of the 2D and 3D NOESY spectra we collected 236 restrains that were converted into interproton distances that were used in structure calculations

Initial calculations performed without a zinc ion, highlighted the spatial proximity of Cys407,

Cys410 and His423, strongly suggesting their involvement in Zn^{2+} binding. The backbone carbonyl group of Arg427 and the residue His433 did not appear to be involved in metal coordination as they were far away form these residues.

The protonation state of the histidine residues was assessed using HMQC long range experiment. For the His423 the spectrum shows that the correlations between the Nε2 and both Hε1 and Hδ2 (around 217 ppm) are stronger than the correlations between Nδ1 and Hε1, Hδ2 (at 174 ppm). This result indicate that the Nδ1 is protonated and the metal binding, via the imidazole rings, occurs through Nε2. For the His433 we could observe a set of correlations between Nε2 and Hε1, Hδ2 around 180ppm and between Nδ1 and Hε1, Hδ2 near 190 ppm. This set of peak correlations indicate that Nδ1 and Nε2 are protonated, again suggesting that the His433 is not involved in the coordination of the metal ion (see Material and Method).

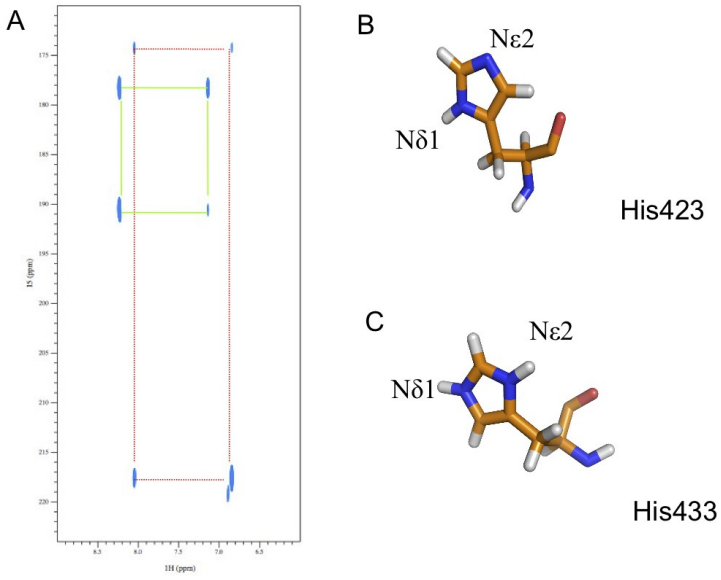


Fig. 10. The protonation state of the histidine residues: HMQC long range experiment. (A) ^{1}H ^{15}N long-range HSQC spectrum of ^{15}N- labelled Nizp1-C2HR. Correlations between the peaks are highlighted by red dashed lines for Hi423 (B) and green continuous lines for His433 (C).

A zinc ion was included in the subsequent and final calculations, assuming a coordination with tree ligands. In the final 20 structures, the Nizp1-C2HR module fold is composed by two-stranded antiparallel β-sheet (β1 and β2 residues Tyr405-Cys407 and Gly411-Phe414, respectively) linked by the loop (residues Arg415-Trp416) with the short α-helix (residue Arg417-Arg425), while the N- and C-terminal regions (Glu496-Ser502 and Ser426-Lys434, respectively) are largely unstructured (see Fig.11A,B).

The structure of Nizp1-C2HR is well defined, with a root-mean-square deviation (RMSD) of 0.629±0.126 Å and 0.521±0.097 Å on the heavy atom and on the backbones, respectively, for the region between the residues Ser404-Arg425 (see Fig.11A). This ββα fold is stabilized by a zinc ion coordinated by thiol groups of Cys407, Cys410 and Nε2 atom of His423 (see Fig. 13). The side chains of the residues Phe417, Phe420 and His423 form the hydrophobic core of the finger and stabilize its fold (see Fig12). In a classical C2H2 or C2HC zinc finger domain, the α-helix extends up to the final chelating residue, whereas in Nizp1-C2HR fingers the helix ends shortly after the position Arg425. Interestingly, the structure of Nizp1-C2HR exhibits a large positive charge face that could be involved in a protein-protein interaction (see Fig.11C).

Structural Statistic Nizp1-C2HR (residues Ser404-Arg425)	
Restrains Informations	
Total number of distance restraints	236
Inter-residual	138
Sequential	54
Short-medium	40
Long	4
Residual distance constraint violations	
Number > 0.5 Å	0
Coordinate rms deviation (Å)	
Ordered backbone atoms (N, Cα, C)	0.521±0.097
Ordered heavy atoms	0.629±0.126
Ramachandran quality parameters (%)	
Residues in most favoured regions	68.5
Residues in allowed regions	3.3
Residues in additional allowed regions	25.7
Residues in disallowed regions	2.5

Table 1. Structural Statistics of Nizp1-C2HR

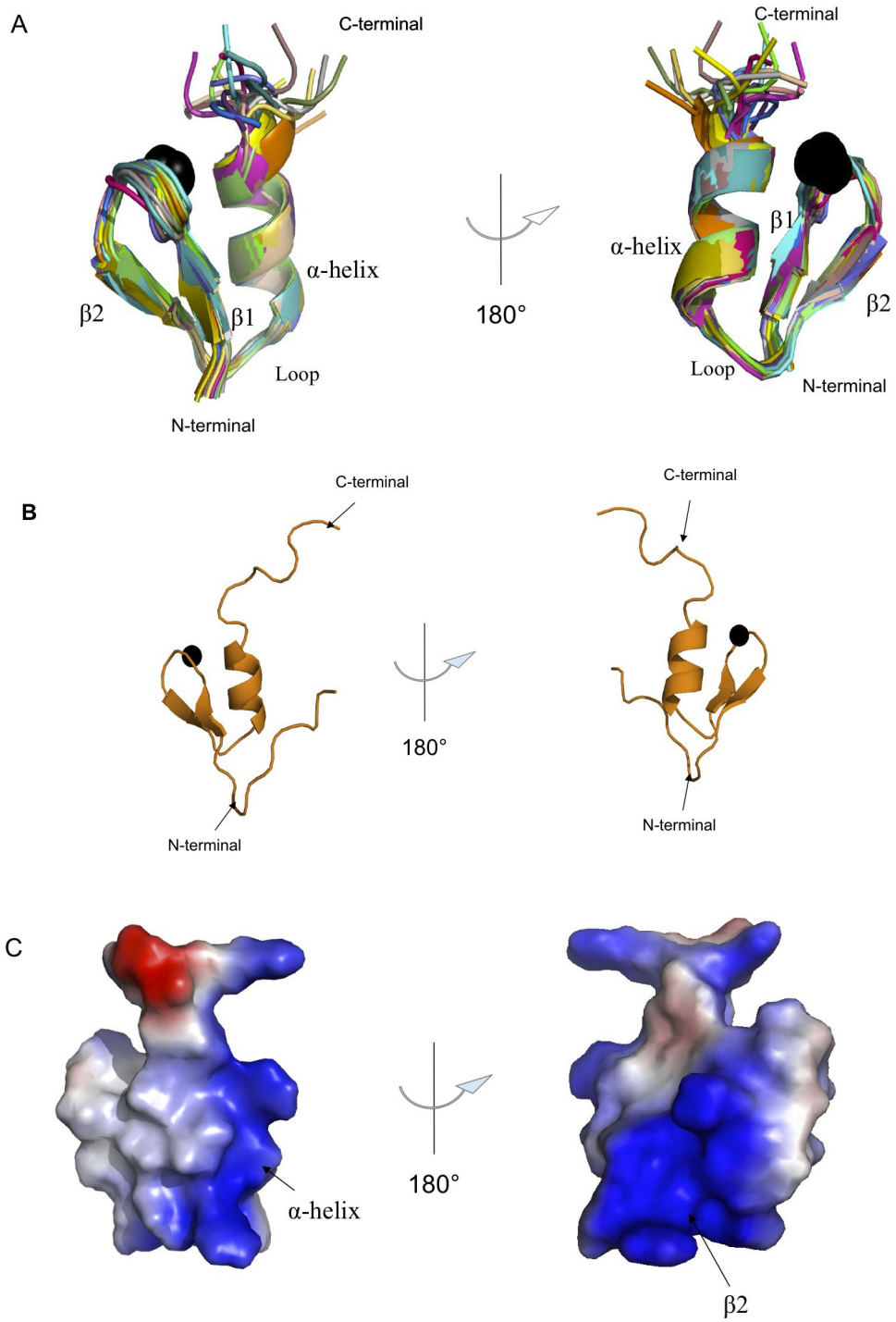


Fig 11. Nizp1-C2HR structure. (A) Cartoon representation of the NMR ensemble of the 20 lowest energy structures of Nizp1-C2HR. Superposition has been performed on the N, Ca and C' atoms of residues Ser401-Arg425 (B) Cartoon representation of the Nizp1-C2HR with the N- and C-terminal regions (Glu496-Ser502 and Ser426-Lys434, respectively) (C) Electrostatic surface of Nizp1-C2HR. The structure in the left side of the panel has the same orientation as in panel A. The structure in the right side of the panel is related to the former by a 180° rotation as indicated by the curved arrow. Blue and red indicate positive and negative electrostatic surface potentials, respectively.

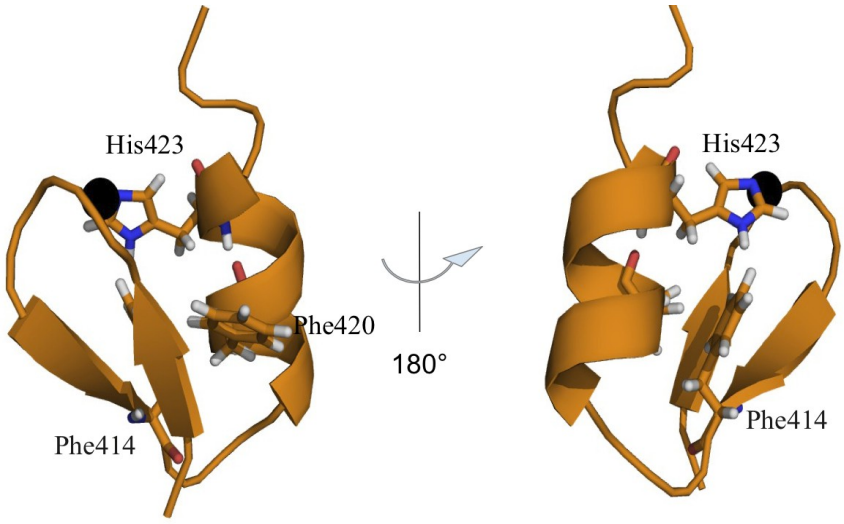


Fig12. Nizp1-C2HR hydrophobic core. The side chains of the residues forming Nizp1-C2HR hydrophobic core are shown as sticks.

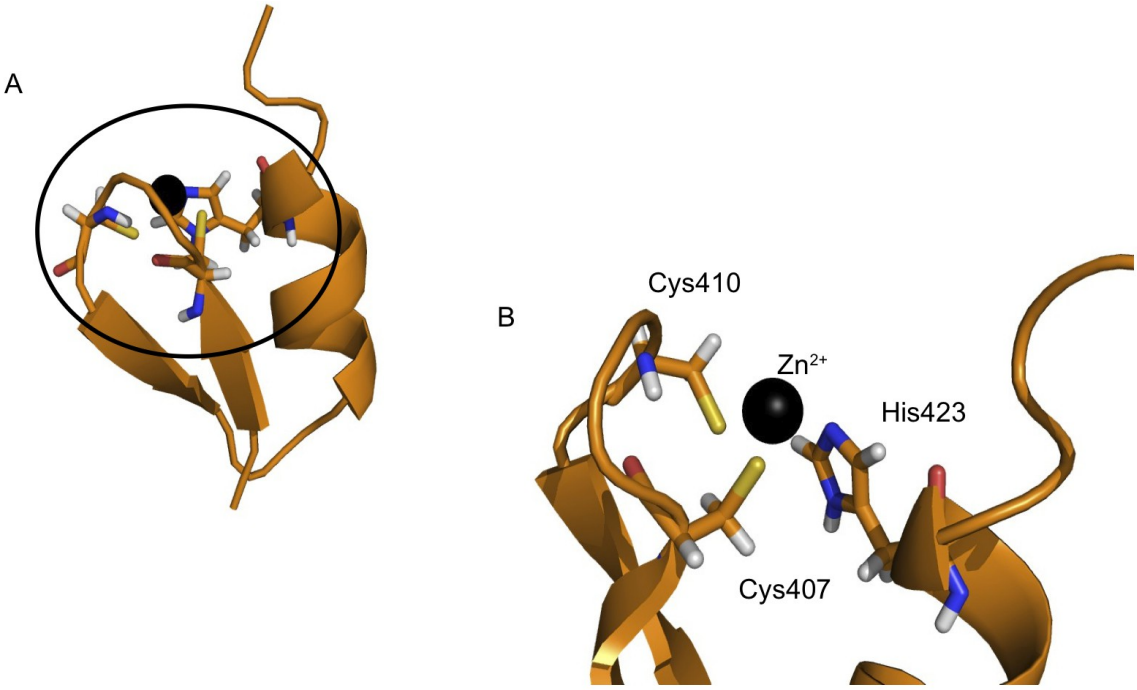


Fig 13. Nizp1-C2HR zinc binding site. (A)The zinc binding site of Nizp1-C2HR in the circle (B) Close-up view of the zinc binding site

4.2.4The structural role of Nizp1-C2HR-Asn409 and Nizp1-C2HR-Arg427

In Nizp1-C2HR domain the Zn^{2+} coordination occurs through two Cysteins (Cys407 and Cys410) and one Histidine (His423). The presence of only three coordinants is sufficient to bind the Zn^{2+} ion (Simpson et al. 2003) whereby the forth coordinants could be represented by a water molecule or a carboxylic group of the protein residue, as observed in other Zn^{2+} binding proteins like in AFV1p06 where the fourth ligand is the side-chain oxygen atom of the carboxylic group of Glu residue (Andreini et al. 2011, Guilliere F et al 2013).

However, in 4 out of the 20 structures defining the NMR ensemble, the carbonyl group in the side-chain of the residue Asn409 is close to the zinc ion binding site, suggesting a possible role in metal coordination. In order to investigate this hypothesis we analyzed the effect on the domain fold when the Asn409 is substituted by an alanine. The 1H 1D spectrum of Nizp1-C2HR-Asn409Ala mutant was collected in the presence of $ZnCl_2$ in molar ratio of 1:1. The spectrum shows a good peak dispersions and it is very similar (except for the peak that correspond to Asn409 around 9.1 ppm) to the spectrum of wild-type domain bound to the zinc ions. This result indicates that the carbonyl-group in the side-chain of Asn409 is not involved in the metal binding.

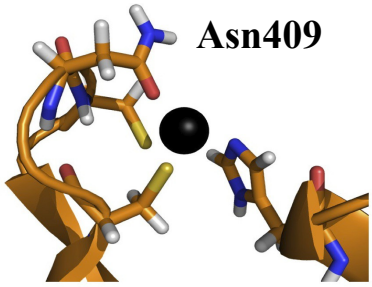


Fig 15 The zinc binding site of Nizp1-C2HR. The position of the Asn409 residue in 4 structures.

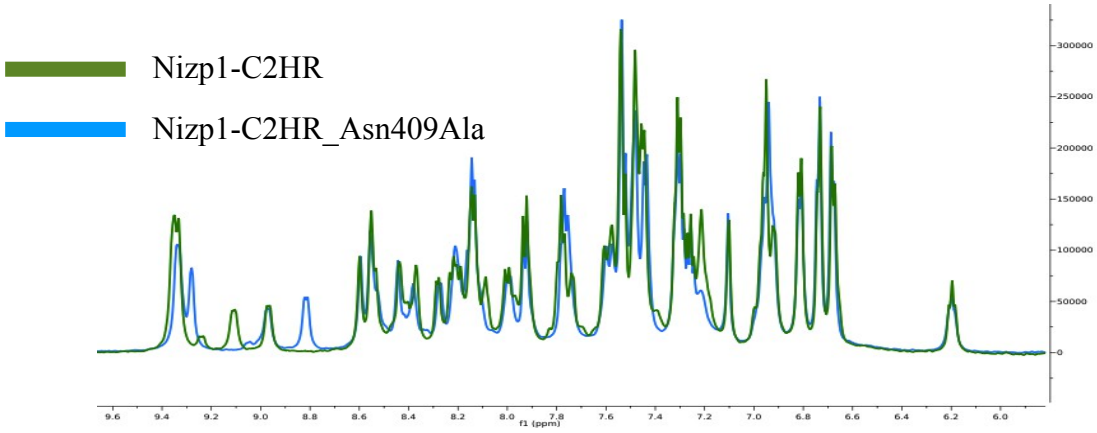


Fig 14. The Asn409 residue is not a zinc ligand. Superposition of 1H 1D spectrum of Nizp1-C2HR wild type (green) 1H 1D spectrum of Nizp1-C2HR-Asn409Ala mutant (cyan).

In the Nizp1-C2HR sequence the fourth zinc ligand is substituted by arginine residue (Arg427). By means of two hybrid assay Nielsen *et al* have studied the interaction between NSD1-P5C5 and Nizp1-C2HR and Nizp1-C2HR mutants (called Nizp1-C2HR-Arg427Ala and Nizp1-C2HR-Arg427His, respectively) using two hybrid assays. In their experimental conditions they observed that only the the Nizp1-C2HR-Arg427Ala mutant was able to bind with NSD1-P5C5.

These results prompted us to investigate the fold and the interaction of these mutants with NSD1-P5C5 by NMR experiments. To this aim we generated by site directed mutagenesis the Nizp1-C2HR-Arg427Ala mutant and we synthesized the peptide corresponding to Nizp1-C2HR_Arg427His mutant. The Nizp1-C2HR_Arg427Ala and Nizp1-C2HR_Arg427His fold has been studied by ^{1}H ^{1}D, ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY NMR spectra. We observed that the 1H-1D NMR spectrum of Nizp1-C2HR_Arg427Ala is is well dispersed indicating that the domain maintains its protein fold (see Fig.16). The result was confirmed by ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY spectra that indicated that for the domain establishes similar intermolecular NOEs as wild-type.

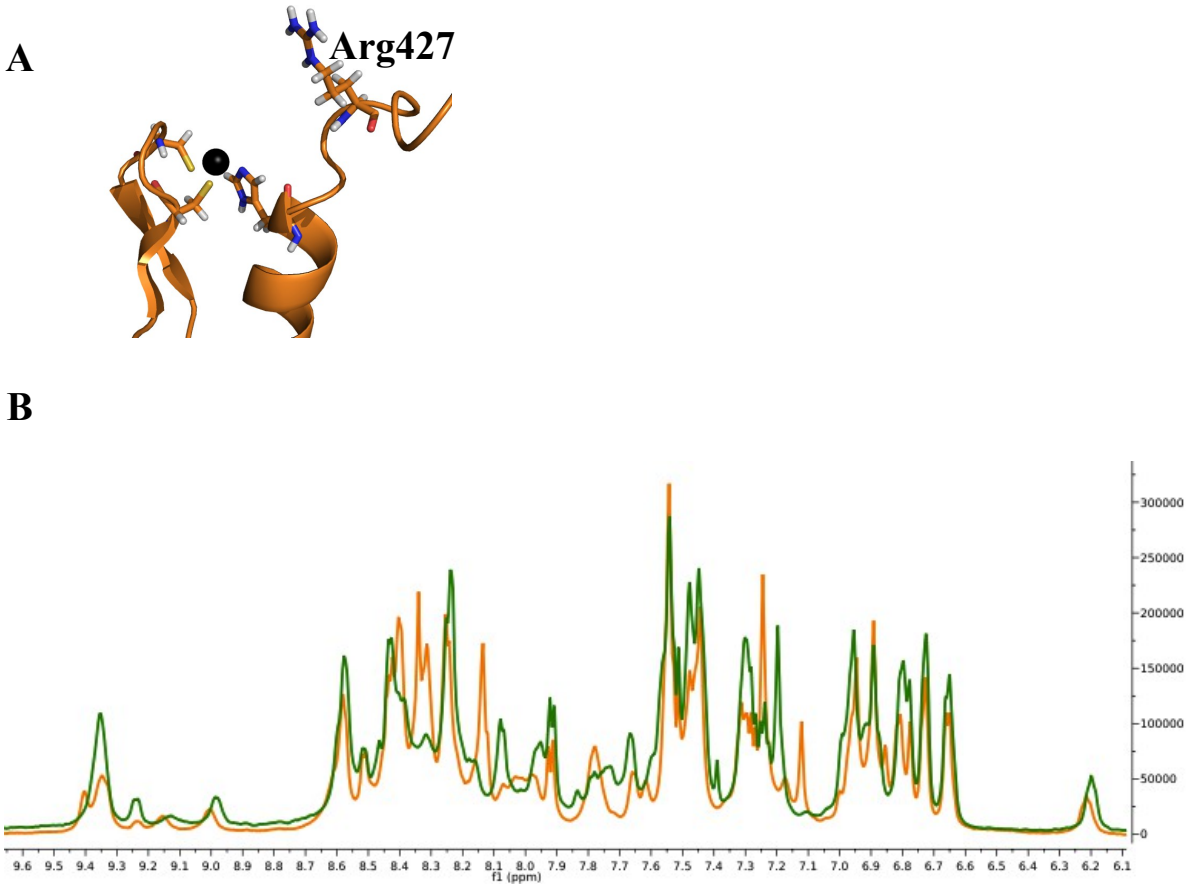


Fig 16. The Arg427Ala mutation does not affect the motif fold. (A) The position of the Arg427 residue in the NMR structure. (B) Superposition of 1H 1D spectrum of Nizp1-C2HR wild type (green) 1H 1D spectrum of Nizp1-C2HR_Arg427Ala mutant (orange).

In the Nizp1-C2HR_Arg427His mutant (see Fig 17), inter-residue NOEs indicate that His427 is near the zinc binding site suggesting an involvement of this residues in metal coordination. Based on the experimental restrains obtained from the analysis of the NOESY spectra we determined the structure of the Nizp1-C2HR_Arg427His: the solution structure is similar to the wild-type one, whereby the α-helix is slightly longer as compared to the wild type. The extension after position 427 is due to the presence of the additional histidine, that coordinates the zinc ion inducing an additional helical turn (see Fig 18).

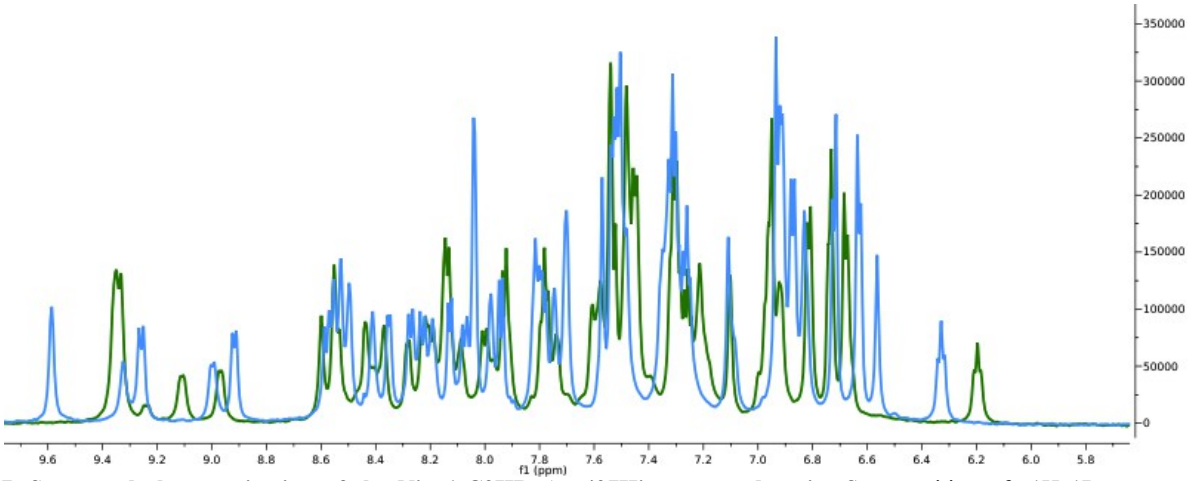


Fig 17. Structural characterization of the Nizp1-C2HR_Arg427His mutant domain. Superposition of 1H 1D spectrum of Nizp1-C2HR wild type (green) 1H 1D spectrum of Nizp1-C2HR-Arg427His mutant (light blue).

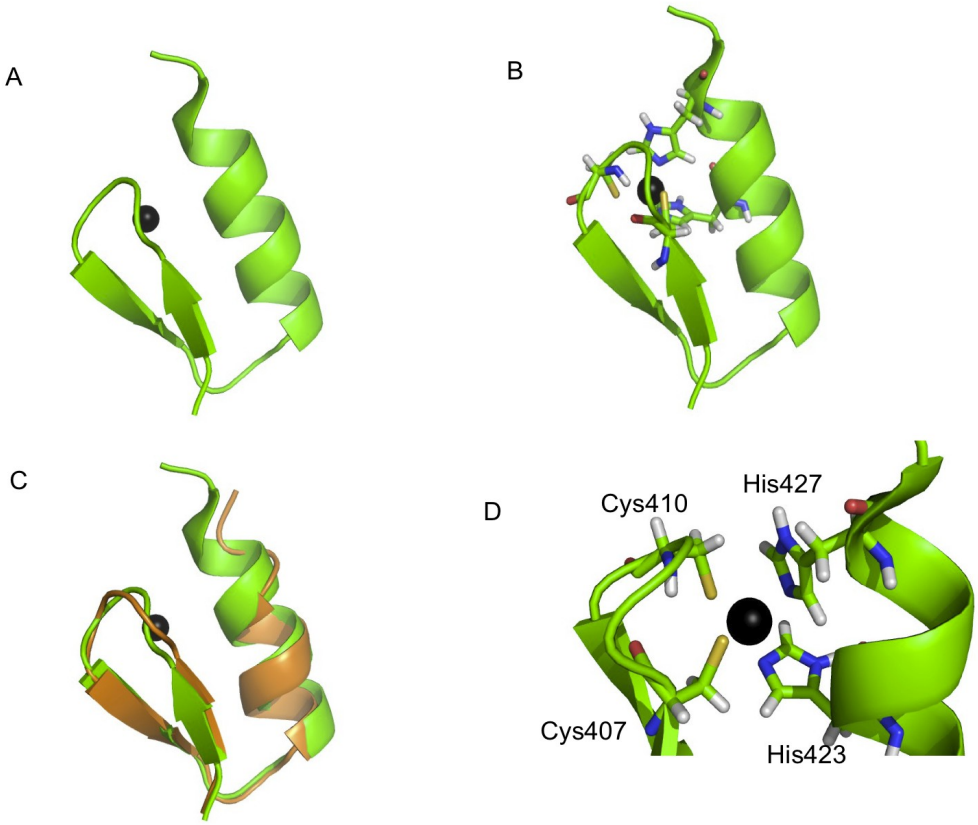


Fig.18 Structural characterization of Nizp1-C2HR_Arg427His mutant domain. (A) Cartoon representation of Nizp1-C2HR_Arg427His (in green) and its zinc binding site (B and D). (C) Superposition of Nizp1-C2HR (orange) and Nizp1-C2HR_Arg427His (green)

4.2.5 Summary

Nizp1-C2HR is a zinc finger domain that contains three zinc coordinating ligands in the metal binding motif. Interestingly, the Nizp1-C2HR synthetic peptide assumes its native fold in the presence of zinc ions suggesting that two Cys and one His are sufficient for Zn^{2+} coordination. In orde to perform multidimensional NMR experiments, we produced this small zinc finger domain (37 residues) as fusion protein in bacteria and the purified recombinant module folds as the synthetic peptide. We solved the NMR structure of Nizp1-C2HR, it adopts a typical zinc finger fold with a two-stranded antiparallel β-sheet and a short α-helix that ends after the coordinating histidine. The structural analysis of Nizp1-C2HR-Asn409Ala and Nizp1-C2HR-Arg427Ala and Nizp1-C2HR-Arg427His mutants, reveals that the mutants are well folded and that insertion of a forth Zn^{2+} coordinating residue extends the domain helix by one helical turn. The solved structures of NSD1-P5C5 and Nizp1-C2HR are the pre-requisite for the structural studies that will be described in the following chapters.

4.3 Characterization of the interaction between NSD1-P5C5 and Nipz1-C2HR

4.3.1 Introduction

NMR chemical-shift mapping method is a highly sensitive tool to prove binding and to map binding sites in protein-protein interactions (Rajagopal et al 1997). It requires ^{15}N-labeled protein in order to record 2D ^{1}H-^{15}N HSQC spectra before and after the addition of increasing amounts of the ligands. The HSQC provides the correlation between the nitrogen and amide proton, and each amide yields a peak in the HSQC spectrum. Peaks corresponding to a residue involved in the recognition process might move, duplicate or broaden depending on the exchange regime of complex formation (see Material and Methods). In the system object of our study we have two protein domains (NSD1-P5C5 and Nizp1-C2HR) and using NMR chemical-shift mapping method we can identify the binding surface on both of them. In order to identify the protein backbone NH groups affected by the binding, we performed NMR titrations of ^{15}N labelled NSD1-P5C5 with unlabelled Nizp1-C2HR and reversed titrations where ^{15}N labelled Nizp1-C2HR was titrated with unlabelled NSD1-P5C5.

4.3.2 Identification of the binding surface on NSD1-P5C5

The NMR titration of ^{15}N-NSD1-P5C5 with unlabelled Nizp1-C2HR synthetic peptide was performed up to final 1:1 ratio. In Fig. 19 in blue is reported the 2D ^{1}H ^{15}NHSQC spectrum of NSD1-P5C5 in free state and in red the spectrum of the complex NSD1-P5C5:Nizp1-C2HR (see Fig 19A,B). Superposition of the two spectra shows that several peaks move in the spectra upon Nizp1-C2HR addition suggesting that the two proteins interact directly. The exchange regime of the reaction is slow in the chemical-shift time scale: residues affected by the binding show reduction of their peak intensities at sub-stoichiometric ratio, while new peaks, belonging to the protein in the bound form, appear at different chemical shifts, with an increasing intensity upon ligand addition (see Fig 19C).

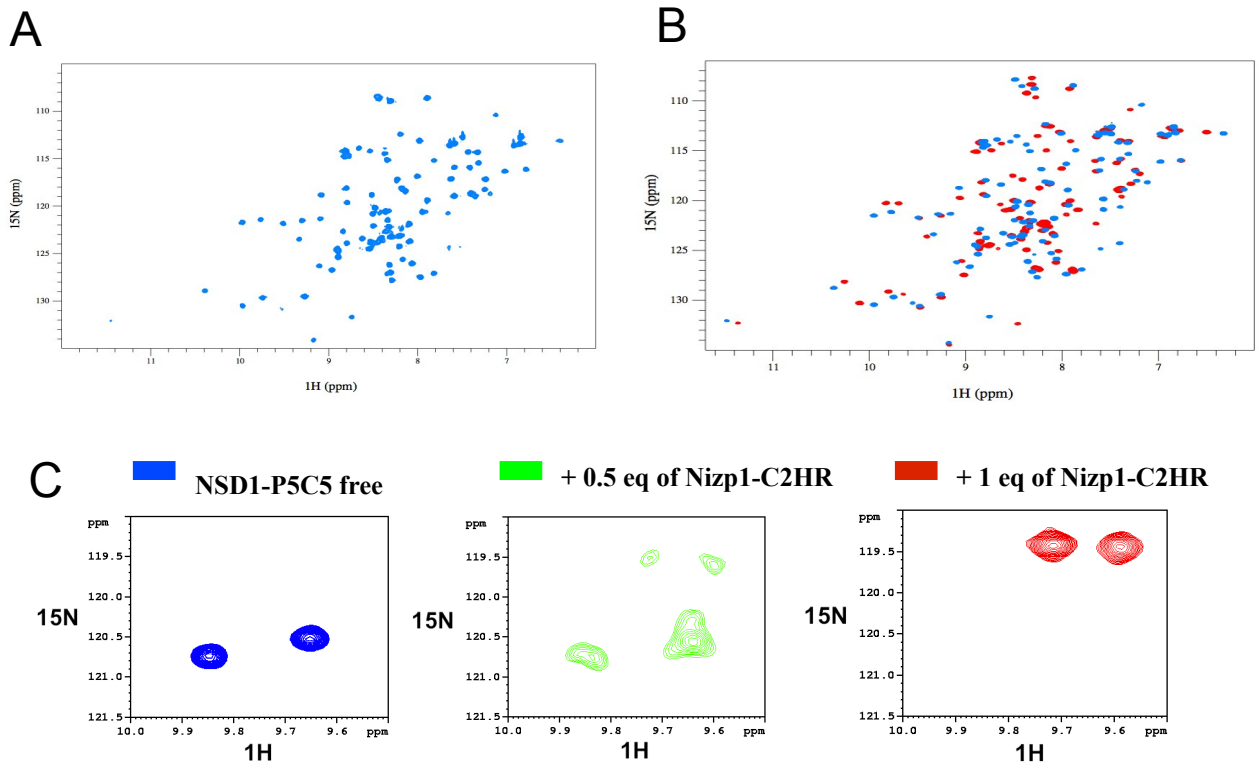


Fig. 19: NSD1-P5C5 binds Nizp1-C2HR. (A) ^{1}H ^{15}N-HSQC spectrum of NSD1-P5C5 in free state (blue). (B) ^{1}H ^{15}N-HSQC spectra resulting from the titration. Shown in blue the spectrum of the free NSD1-P5C5. The red spectrum corresponds to the same sample after the addition of 1 equivalent of Nizp1-C2HR. (C) Close-up view of two peaks during the NMR titration. The exchange regime of the reaction is slow in the chemical-shift time scale: with the progressive addition of ligand, new resonances appear at the chemical shift of the bound form, whereas the resonances of the free state progressively disappear.

The residues shifting upon peptide addition were identified performing 3D NMR experiments, HNCα and CβCα(CO)NH spectra (see Material and Methods), that permit to assign the backbone NH group of NSD1-P5C5 in complex with Nizp1-C2HR.

To compare the free and the bound forms, the chemical shift displacements (CSDs) of the protein backbone NH groups upon ligand addition are reported in the histogram (Fig. 20A). We can observe that the backbone NH groups of the first PHD finger are more perturbed than the NH groups of the second module. In the CSD plot we considered NSD1-P5C5 aminoacids significantly affected by ligand binding those residues showing a CSD higher than the average (green line) plus one standard deviation (red line) (see Fig. 20A) and these residues are Asp2119, Phe2122, Ser2123, Cys2124, His2143, Leu2147, Trp2157, Glu2158, Met2177.

Also the chemical-shift of the Cα and Cβ in the residue side-chains were perturbed by the binding event and provided information about the interaction surface. The differences of the Cα and Cβ chemical shift are plotted in the histograms in Fig. 20C and D and we considered significant the residues whose CSDs is equal or higher than 0.5 ppm. The amino acids involved in the interaction are Phe2122, Ser2123, Cys2124, Ser2132, Cys2133, His2143, Tyr2142, Cys2146, Leu2147, Ala2152, Trp2155, Glu2156, Glu2176, Cys2178, Pro2179, Ser2180, Ser2181, Glu2204 and His2205. All the residues affected by the binding were mapped on the tandem domain structure (see Fig. 20D).

The majority of these residues clustered around a common region localized at the interface between the PHD finger tandem domain, while other residues far from this region are indirectly affected by the interaction. In the tandem protein domain structure the interface region between the two modules is composed by hydrophobic residues that form direct inter-domain interactions, creating an hydrophobic groove surrounded on both sides by negatively charged residues that can establish polar contacts and hydrophobic interactions with the Nizp1-C2HR (see Fig. 20).

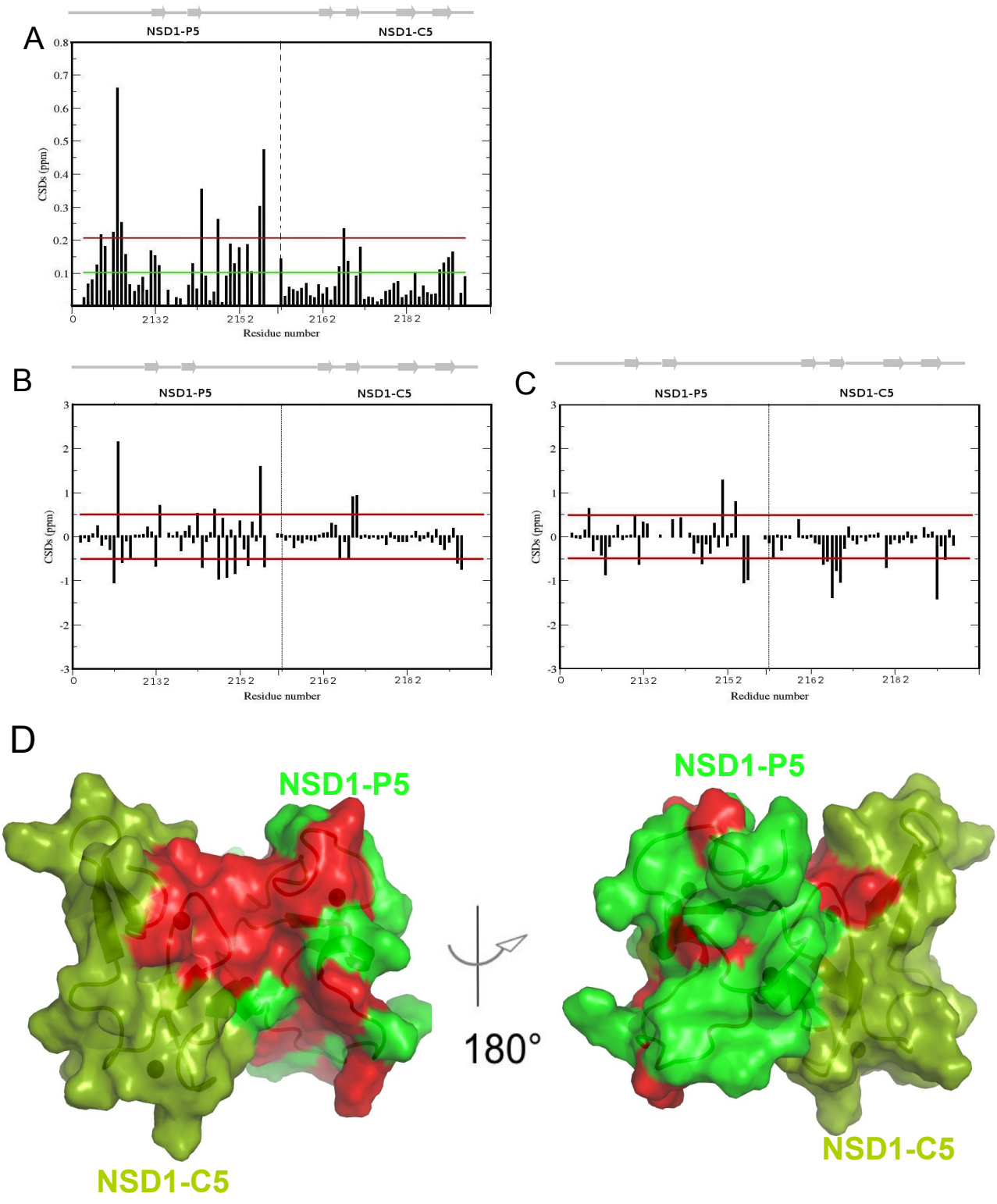


Fig.20 Binding site mapping on the surface of the NSD1-P5C5 domain. (A) The histogram shows the backbone NH group Chemical Shift Displacement (CSD, in ppm) value per each assigned residue of NSD1-P5C5 in complex with Nizp1-C2HR. Residues displaying values higher than the average (green line) plus one standard deviation (red line) were considered for the Nizp1-C2HR binding site map. (B) and (C) The histograms show the differences of the Cα and Cβ chemical shift, respectively. Residues displaying values over 0.5 ppm (red line) were considered significantly affected by the binding. (D)Representation of the Nizp1-C2HR binding site on the surface of the NSD1-P5C5 domain. Residues displaying a significant CSD are coloured in red. The NSD1-P5 is coloured in light green, while the NSD1-C5 is coloured in dark green.

4.3.2 Identification of the binding surface on Nizp1-C2HR

The recombinant Nizp1-C2HR domain was labelled with ^{15}N and used in NMR titration with unlabelled NSD1-P5C5. An overlay of the 2D ^{1}H-^{15}N HSQC spectra of free Nizp1-C2HR (brown) and in the presence of NSD1-P5C5 (blue, final ratio 1:1) is shown in Fig. 21. The interaction is in the slow exchange regime in the chemical shift time scale and several peaks move in the spectrum. The backbone of Nizp1-C2HR in complex with NSD1-P5C5 was assigned by 3D NMR experiments (HNCA experiment, see Material and Method). The chemical shift displacements of the protein backbone NH groups free and bound to NSD1-P5C5 are plotted in the histogram in Fig. 22. Residues showing CSD higher than the average (green line) plus one standard deviation (red line) included Arg417, Val418, Phe420, Arg422 and His423 (Fig. 22A). As described before for NSD1-P5C5, we also considered the chemical-shift perturbations of the Cα to identify the binding surface. The residues whose CSDs were higher than of 0.2 ppm are localized on the second β2-strand (Lys412, Phe414), on the loop (Trp416) and on the short α-helix (Arg417, Val418, Asn419, Phe420, Ile421, Arg422). The majority of these residues clustered on the short α-helix suggesting that this structural element is involved in the interaction.

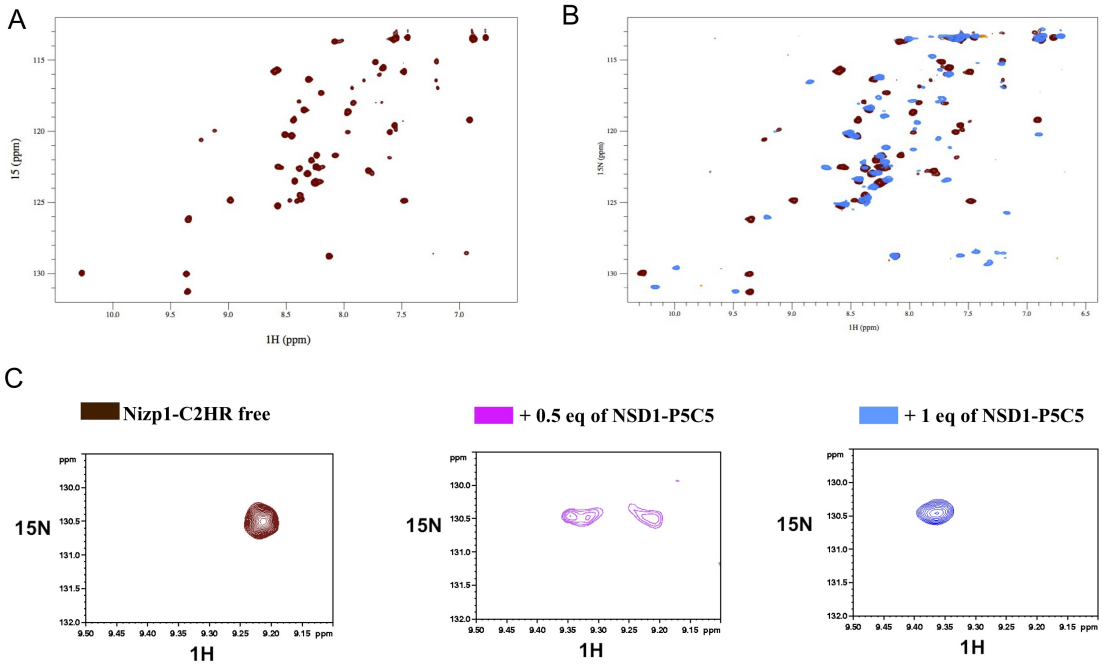


Fig. 21. Nizp1-C2HR binding to the NSD1-P5C5 monitored by NMR. (A) ^{1}H ^{15}N HSQC spectrum of Nizp1-C2HR in free state (blue). (B) ^{1}H ^{15}N HSQC spectra resulting from the titration of increasing amounts of unlabelled NSD1-P5C5 domain on a ^{1}H ^{15}N -Nizp1-C2HR sample (C) Close-up view of two peaks during the NMR titration. The exchange regime of the reaction is slow in the chemical-shift time scale.

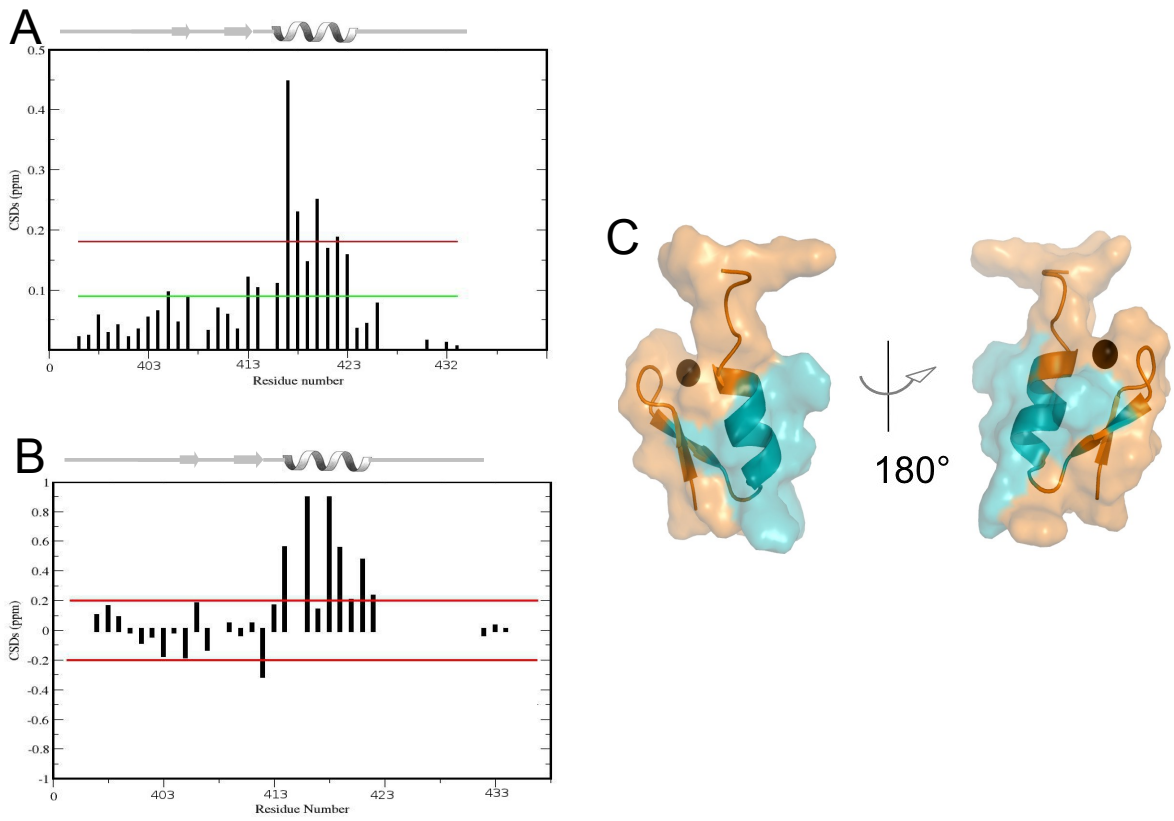


Fig.22: Binding site mapping on the surface of the Nizp1-C2HR domain. (A) The histogram shows the backbone NH group Chemical Shift Displacement (CSD, in ppm) value for each assigned residue of Nizp1-C2HR in complex with NSD1-P5C5 Residues displaying values higher than the average (green line) plus one standard deviation (red line) were considered for the binding site map. (B) The histograms show the differences of the Cα. Residues displaying values over 0.2 ppm (red line) were considered significant affected by the binding. (C) Representation of the NSD1-P5C5 binding site on the surface of the Nizp1-C2HR domain. Residues displaying a significant CSD are coloured teal.

4.3.3 Summary

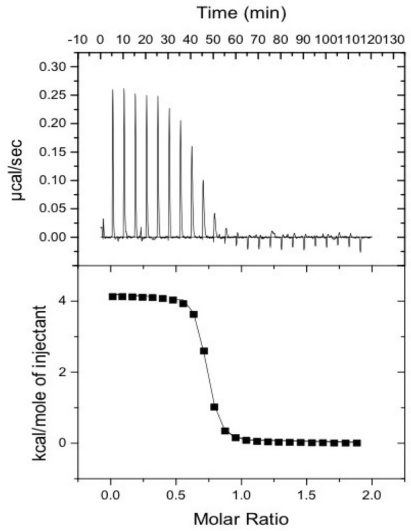
By means of NMR titrations we have validated and characterized the interaction between NSD1-P5C5 and Nizp1-C2HR. Using the CSDs we identified the binding surface both on NSD1-P5C5 and Nizp1-C2HR. For NSD1-P5C5, we observed that both the PHD finger domains are involved in the interaction. In particular, the interface region between the two domains, composed by hydrophobic and negatively charged residues, forms the binding pocket for the zinc finger domain. For Nizp1-C2HR, the CSDs indicates that the short α-helix is involved in the interaction. In the next chapter the thermodynamic parameters that govern this protein-protein interaction will be presented.

4.3.4 Thermodynamic characterization of the interaction between NSD1-P5C5 and Nizp1-C2HR

Isothermal titration calorimetry experiment were performed to determine the binding affinity and the binding enthalpy and entropy of the interaction between NSD1-P5C5 and Nizp1-C2HR. The heat of the interaction of Nizp1-C2HR (200 μM) with NSD1-P5C5 (15 μM) was determined in 20mM phosphate buffer, 150 mM of NaCl and 2mM β-mercaptoethanol at pH 7.2 and 23°C. Fig. 23 shows the typical titration for this interaction.

Under these conditions, complexation of NSD1-P5C5 and Nizp1-C2HR was endothermic (positive peaks in the ITC output) with saturation around 1:1 stoichiometry. Integrated heat data yielded thermal binding curve that could be analyzed by standard nonlinear regression methods to give estimates of the binding stoichiometry (n), equilibrium affinity constant (K_a) for the association ($K_a = 1/K_d$, where K_d is the dissociation constant), and the enthalpy of binding (ΔH), see Fig. 23. To estimate accurate enthalpy value the contributions of the buffer enthalpy and protein dilution enthalpy in sample cell were subtracted to the enthalpy of the reaction. The free energy of the reaction (ΔG) was calculated from the relationship $\Delta G = -RTInK_a$, and the change in entropy (ΔS) was calculated from $T\Delta S = \Delta H - \Delta G$ (see Fig. 23).

The measured K_d is in the range of μM that is typical for this biological system and, the stoichiometry of the reaction is 1 as indicated by the value of n close to 1. The binding reaction is characterized by enthalpically unfavourable (ΔH>0) but entropically favourable (ΔS>0) process and the large positive entropy is compensated by a small positive enthalpy change. This data suggest that the reorganization of the ordered solvation network and the exclusion of water molecules from the complex surface (these contributes are entropically favourable) are the driving force of the reaction.



K_a (M^{-1})	(2,63 ±0,46) •10^5
K_d (M)	(3,80±0,66) •10^6
n	0.802±0.004
ΔH (kcal/mol)	+4.14
TΔS (kcal/mol)	+14.19
ΔG (kcal/mol)	-10.05

Fig. 23. ITC titration of Nizp1-C2HR motif on the NSD1-P5C5 domain. The measured differential heat in function of time is shown in the top panel. The lower panel shows the integrated data and the calculated residuals in function of the ligand molar ratio, as well as the fitted binding isothermal. The thermodynamic parameters are displayed in the table in the right.

4.3.5 Characterization of the interaction between NSD1-P5C5 pathological point mutants and Nizp1-C2HR

The NSD1-P5C5 Sotos pathological point mutations conceivably lead to a loss of the protein domain function. As previously described (chapter 3.1.1) we studied the fold of seven Sotos mutations. Five of these mutations that affect hydrophobic and histidine coordinating residues strongly compromise the protein fold, whereas His2162Arg and Arg2152Gln mutations preserve the protein fold.

In this section, in order to investigate the role of NSD1-P5C5 under pathological conditions we characterized the interaction between the Arg2152Gln and His2162Arg mutants and Nizp1-C2HR.
We followed the amide resonances shifts of uniformly ^{15}N-labeled NSD1-P5C5_Arg2152Gln and NSD1-P5C5_His2206Arg mutants upon the addition of unlabelled Nizp1-C2HR. Both reactions were in slow-exchange regime in the chemical-shift time scale. In the first experiment we observe binding between NSD1-P5C5_Arg2152Gln and Nizp1-C2HR, (see Fig. 24A). The superposition of the 2D spectra (see Fig.24B) of the the complex ^{15}N-NSD1-P5C5_Arg2152Gln:Nizp1-C2HR and ^{15}N-NSD1-P5C5:Nizp1-C2HR shows that the entity of the shifts of the mutant protein upon Nizp1-C2HR are comparable to the wild-type protein, suggesting a similar interaction pattern and a similar interaction strength. These results indicate that the Arg2152Gln substitution doesn't compromise the interaction with Nizp1-C2HR and that the binding is structurally similar. In the NSD1-P5C5 structure the Arg2152 is localized far from the putative binding surface mapped on tandem domain structure (see chapter 4.3.2, and Fig 24E) suggesting that this residue is not involved in protein-protein interaction.

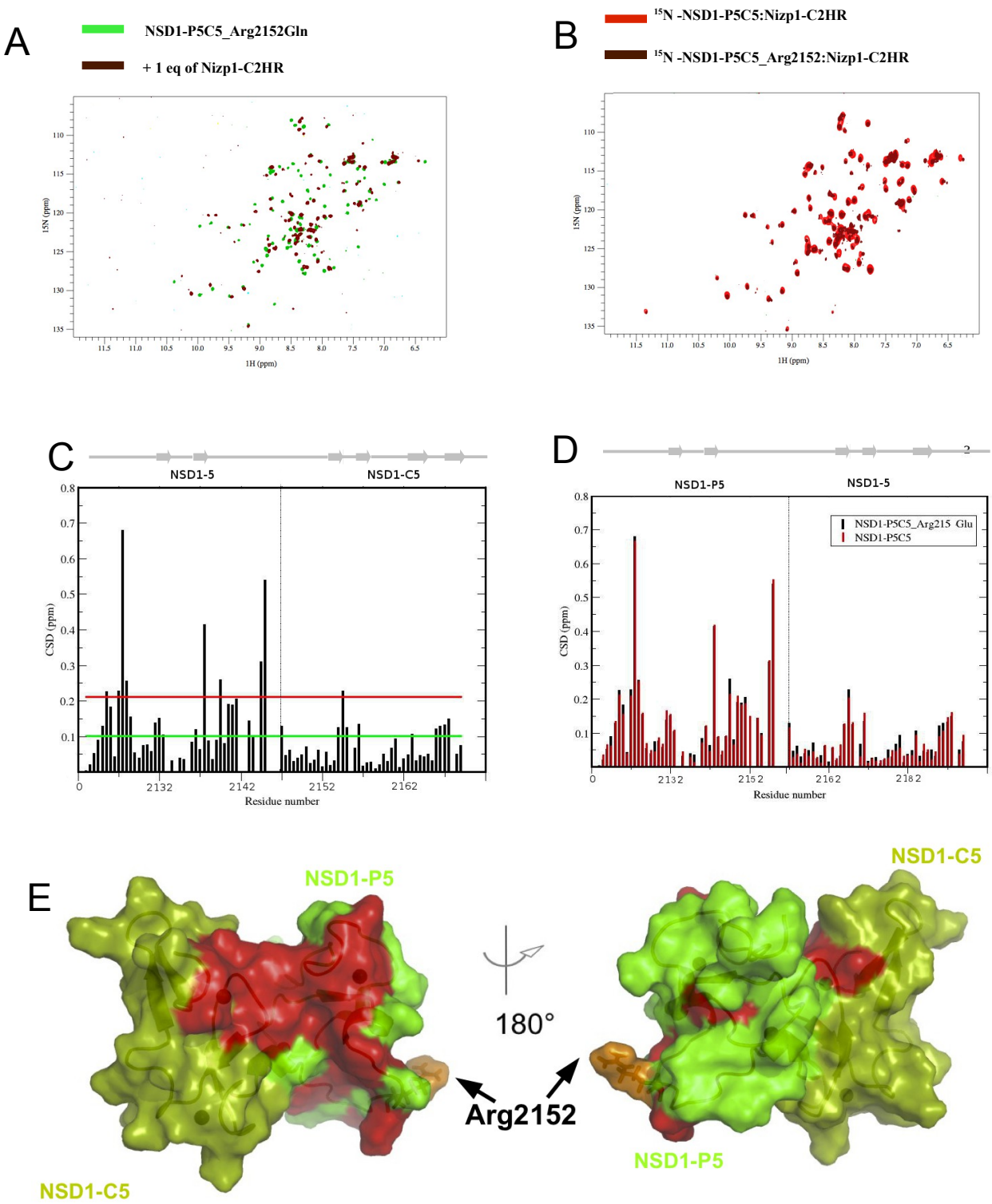


Fig.24 : NSD1-P5C5_Arg2152Gln binds Nizp1-C2HR. (A) HSQC spectra of the ^{15}N-NSD1-P5C5_Arg2152Gln in free state (green) and bound to Nizp1-C2HR (brown). (B)Superposition of the HSQC spectra of the ^{15}N-NSD1-P5C5:Nizp1-C2HR complex (red) and ^{15}N-NSD1-P5C5_ Arg2152Gln:Nizp1-C2HR (brown). (C)Histogram representation of the chemical shift displacement (CSD) observed per each residue. (D) Superposition of the CSD plots resulting from the NMR titration with wild-type NSD1-P5C5 (red) and NSD1-P5C5_Arg2152Gln (black). (E) Position of the Arg2152 residue.

The second pathological substation, NSD1-P5C5-His2206Arg, is near the Nizp1-C2HR binding pocket to (see Fig. 26).

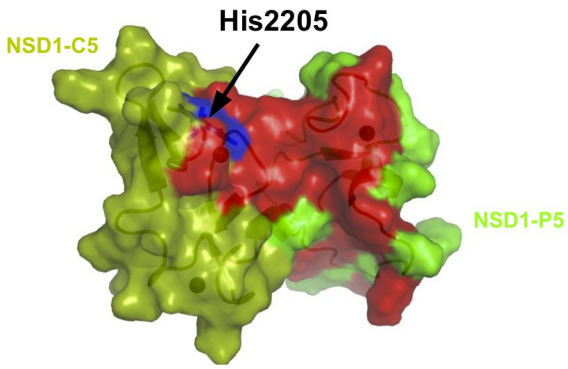


Fig. 26 Localization of the His2205 residue. The Nizp1-C2HR binding surface is shown in red, the surface of the His2205 residue is coloured in blue.

The NMR titration (see Fig. 27A) indicates that the two protein domains interact, however the entity of the shifts upon Nizp1-C2HR addition is smaller as compared to the wild-type protein, suggesting a weaker interaction for this mutant (see Fig. 27B).

Finally comparison between CSDs of the wild-type protein with those of NSD1-P5C5_His2206Arg and with those of NSD1-P5C5_Arg2152Gln upon binding to Nizp-C2HR shows that the His2206Arg mutant has an higher impact in complex formation as compared to Arg2152Gln (Fig. 28). For a quantitive valuation of the impact of these mutations on binding to Nizp1-C2HR in future we measure by ITC titration the association constant between the pathological mutants and Nizp1-C2HR.

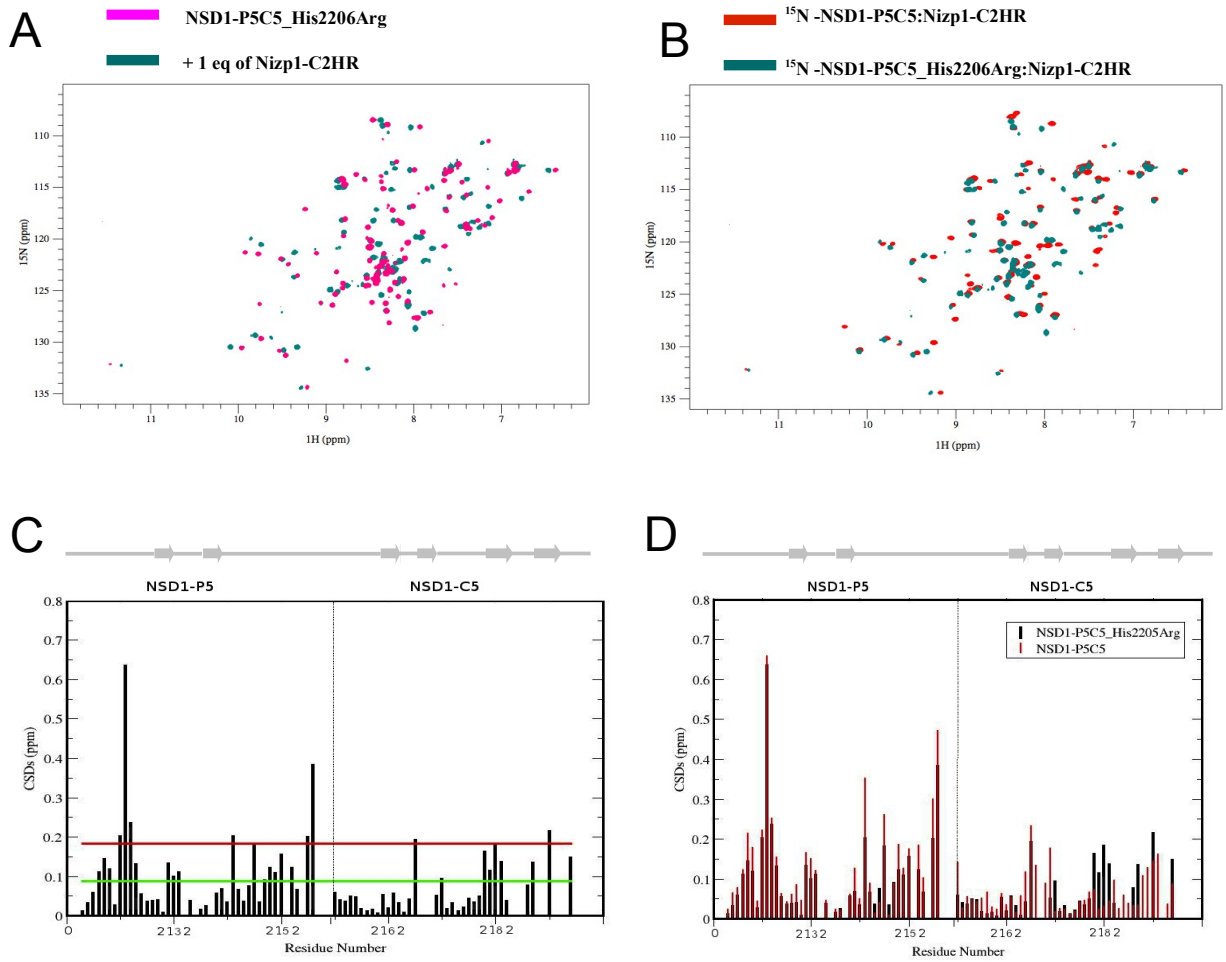


Fig.27: NSD1-P5C5_His2206Arg binds Nizp1-C2HR. (A) HSQC spectra of the ^{15}N-NSD1-P5C5_His2206Arg in free state (pink) and bound to Nizp1-C2HR (teal). (B)Superposition of the HSQC spectra of the ^{15}N-NSD1-P5C5:Nizp1-C2HR complex (red) and ^{15}N-NSD1-P5C5_His2206ArgNizp1-C2HR (teal). (C)Histogram representation of the chemical shift displacement (CSD) observed per each residue. (D) Superposition of the CSD plots resulting from the NMR titration with wild-type NSD1-P5C5 (red) and NSD1-P5C5_His2206Arg (black).

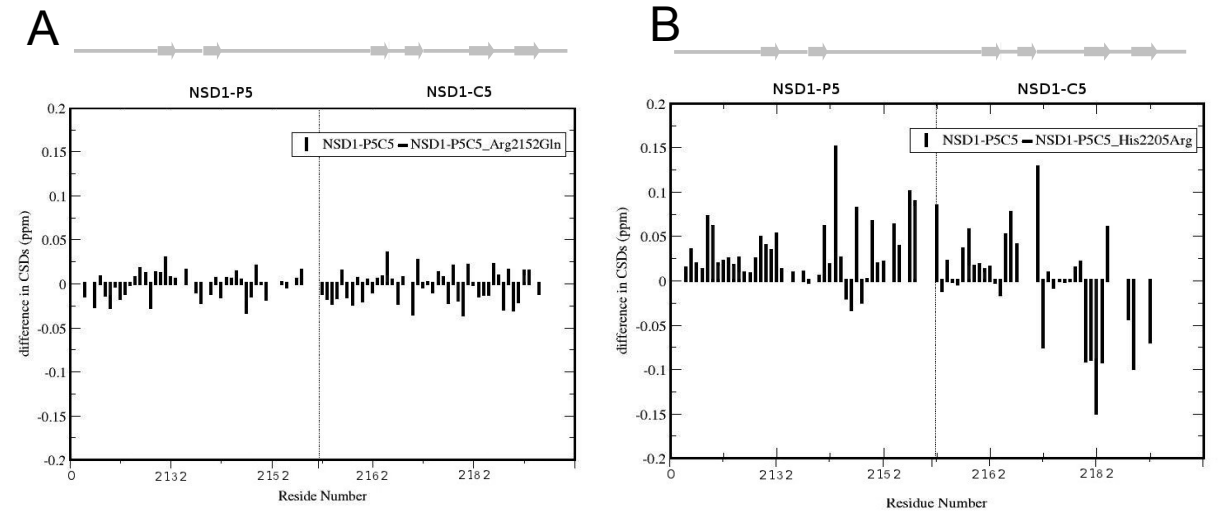


Fig.28. The binding mode of NSD1-P5C5_Arg2152Gln and Nizp1-C2HR.NSD1-P5C5_His2206Arg. (A) The histogram shows the differences in CSDs between the wild-type NSD1-P5C5 domain and NSD1-P5C5_Arg2152Gln mutant. (B)The histogram shows the differences in CSDs between the wild-type NSD1-P5C5 domain and NSD1-P5C5_His2206Arg mutant.

4.3.6 Summary

In this chapter we have characterized the effects of the pathological point mutants on the protein interaction activity of NSD1-P5C5. The CSDs indicate that the Arg2152Gln substitution does not compromise the protein binding with the Nizp1-C2HR while the His2206Arg influences the interaction with the zinc finger module. Most likely diminishing the interaction strength.

In the following section we will investigate the role in protein-protein interaction of Nizp1-C2HR_ Asn409, the residue near the metal binding site, and Nizp1-C2HR_Arg427, the residue that was supposed to substitute the fourth zinc coordinating ligand and that is part of the C2HR signature.

4.3.7 Functional role of Nizp1-C2HR-Asn409 and Nizp1-C2HR-Arg427

In the sections 3.4.5 and 3.4.3 we have characterized the structure of Nizp1-C2HR-Asn409, Nizp1-C2HR-Arg427Ala and Nizp1-C2HR-Arg427His mutants. In this paragraph, the interactions of these three mutants with NSD1-P5C5 and mutants were investigated by NMR titrations.

We did three experiments, in which ^{15}N-NSD1-P5C5 was titrated either with unlabelled Nizp1-C2HR-Asn409, or with unlabelled Nizp1-C2HR-Arg427Ala, or with Nizp1-C2HR-Arg427His (see Fig. 29, 30, 31).

The three NMR titrations indicate that NSD1-P5C5 interacts equally well with the three mutants. In Fig 29B, 30B, 31B is reported an overlay of the 2D ^{1}H-^{15}N HSQC spectra of the complex formed by ^{15}N-NSD1-P5C5:Nizp1-C2HR (in red) and in green, orange and blue the spectra of the complex composed by ^{15}N-NSD1-P5C5:Nizp1-Asn409, ^{15}N-NSD1-P5C5:Nizp1-C2HR-Arg427Ala, ^{15}N-NSD1-P5C5:Nizp1-C2HR-Arg427His and ^{15}N-NSD1-P5C5:Nizp1-Asn409, respectively.We observed that the peaks in the three ^{1}H-^{15}N HSQC spectra have similar chemical shift dispersion suggesting that the mutations do not compromise the protein-protein interaction.

The ^{1}H-^{15}N HSQC spectra of the three complexes were assigned and the chemical-shift perturbations were plotted in three histograms, see Fig. 29C, 30C, 31C. The residues of NSD1-P5C5 affected by the binding event with Asn409Ala and Nizp1-C2HR-Arg427Ala/His mutants are the same as the one involved in interaction with the Nizp1-C2HR wild type domain see Fig. 29D, 30D, 31D. The differences of CSDs between the Nizp1-C2HR wild-type domain and its mutants are smaller than 0.04 ppm (Fig. 29E, 30E, 31E). Collectively, these results suggest that the three mutant domains and the wild type zinc finger module interact with the same binding surface and that the Arg427 and Asn409 side-chain not directly involved in the recruitment. The Arg427His substitution, induces a conformational change that extends the fold of the α-helix but does not alter its protein interaction activity. This data indicates that only the α-helix region between the residues Arg417-His423 is involved in the binding.

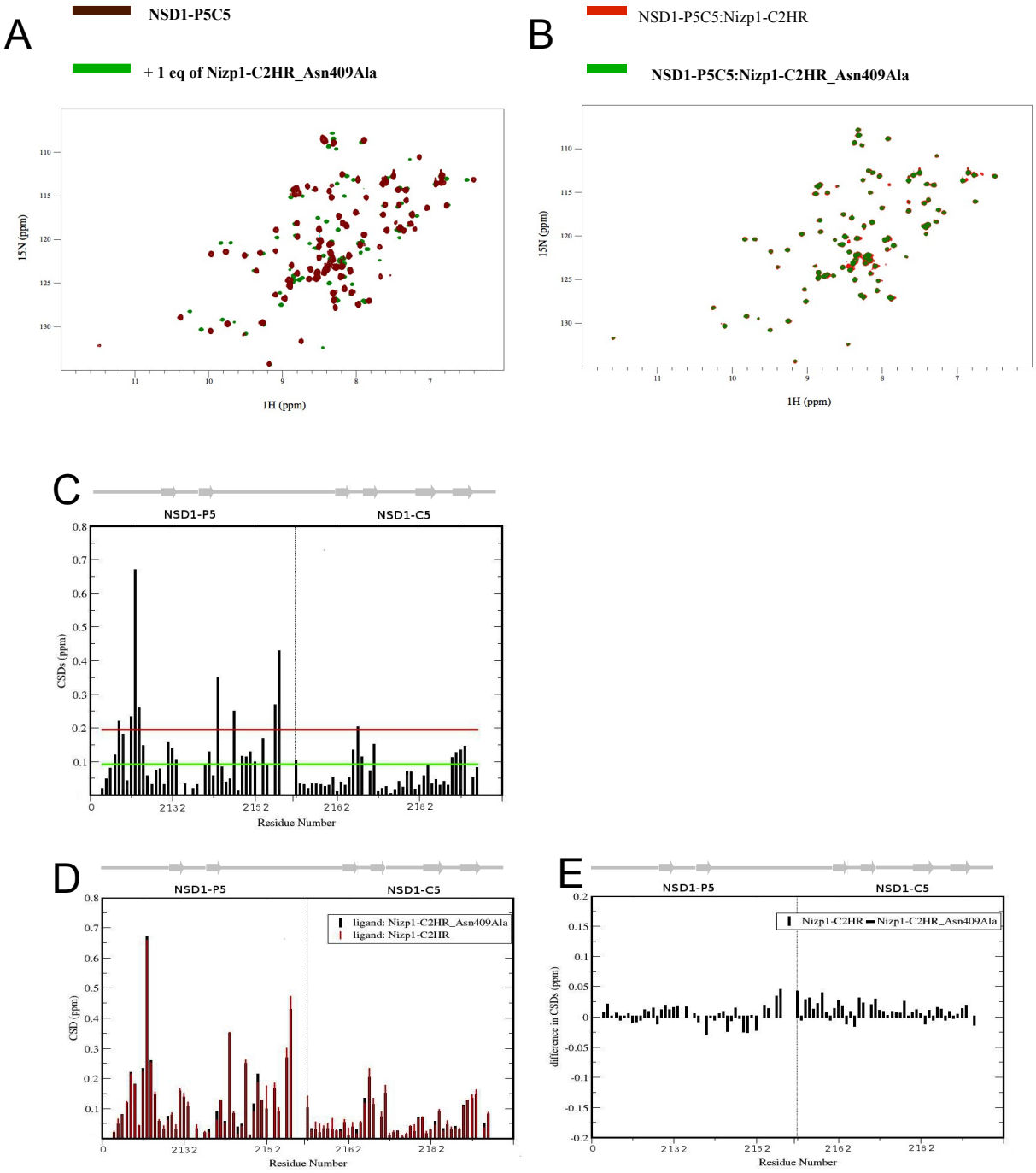


Fig 29: NSD1-P5C5 binds Nizp1-C2HR_Asn409Ala. Data resulting from NMR titration of the unlabelled Nizp1-C2HR_Asn409Ala on a ^{15}N-NSD1-P5C5 domain sample (A) HSQC spectra resulting form the NMR titration of ^{15}N-NSD1-P5C5 with Nizp1-C2HR_Asn409Ala. In brown the HSQC spectrum of the free ^{15}N-NSD1-P5C5 domain and in green the HSQC spectrum of the ^{15}N-NSD1-P5C5 domain bound to Nizp1-C2HR_Asn409Ala. (B) Superposition of the HSQC spectra of the ^{15}N-NSD1-P5C5:Nizp1-C2HR complex (red) and ^{15}N-NSD1-P5C5-C2HR_Asn409Ala complex (green). (C)Histogram representation of the chemical shift displacement (CSD) observed per each residue of the NSD1-P5C5 domain titrated with Nizp1-C2HR_Asn409Ala .(D) Superimposition of the CSD plots resulting from the NMR titration of the NSD1-P5C5 domain with wild-type Nizp1-C2HR (red) and Nizp1-C2HR_Asn409Ala (black). (E) The histogram shows the differences in CSDs upon binding to Nizp1-C2HR and Nizp1-C2HR_ Asn409Ala.

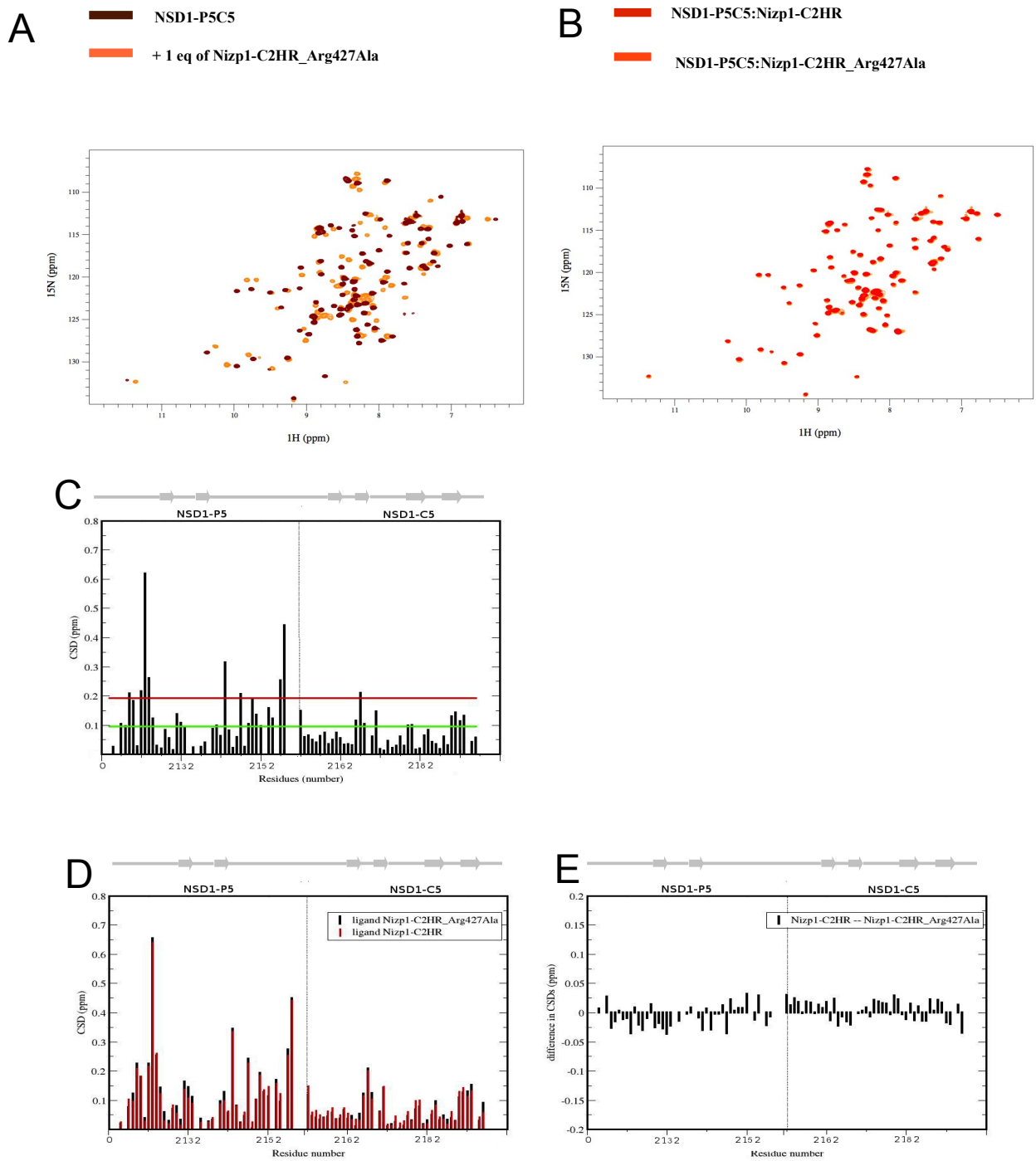


Fig.30: NSD1-P5C5 binds Nizp1-C2HR_Arg427Ala. Data resulting from NMR titration of the unlabelled Nizp1-C2HR_Arg427Ala on a ^{15}N-NSD1-P5C5 domain sample (A) HSQC spectra resulting form the NMR titration of ^{15}N-NSD1-P5C5 with Nizp1-C2HR_Arg427Ala. In brown the HSQC spectrum of the free ^{15}N-NSD1-P5C5 domain and in green the HSQC spectrum of the ^{15}N-NSD1-P5C5 domain bound to Nizp1-C2HR_Arg427Ala. (B) Superposition of the HSQC spectra of the ^{15}N-NSD1-P5C5:Nizp1-C2HR complex (red) and ^{15}N-NSD1-P5C5-C2HR_Arg427Ala complex (green). (C)Histogram representation of the chemical shift displacement (CSD) observed per each residue of the NSD1-P5C5 domain titrated with Nizp1-C2HR_Arg427Ala. (D) Superimposition of the CSD plots resulting from the NMR titration of the NSD1-P5C5 domain with wild-type Nizp1-C2HR (red) and Nizp1-C2HR_ Arg427Ala (black). (E) The histogram shows the differences in CSDs upon binding to Nizp1-C2HR and Nizp1-C2HR_Arg427Ala.

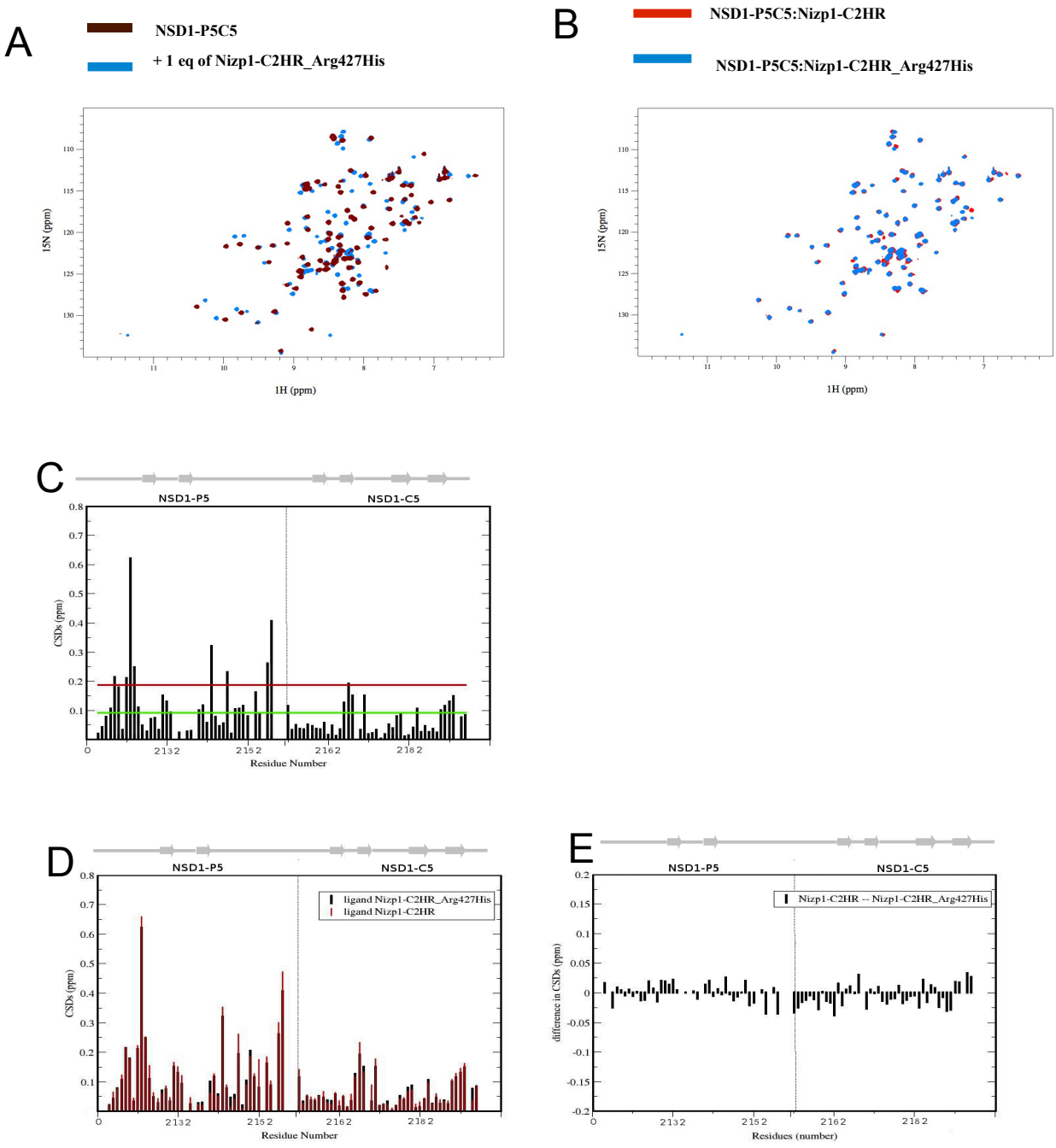


Fig.31: NSD1-P5C5 binds Nizp1-C2HR_Arg427His. Data resulting from NMR titration of the unlabelled Nizp1-C2HR_Arg427His on a ^{15}N-NSD1-P5C5 domain sample (A) HSQC spectra resulting form the NMR titration of ^{15}N-NSD1-P5C5 with Nizp1-C2HR_Arg427His. In brown the HSQC spectrum of the free ^{15}N-NSD1-P5C5 domain and in green the HSQC spectrum of the ^{15}N-NSD1-P5C5 domain bound to Nizp1-C2HR_Arg427His. (B) Superposition of the HSQC spectra of the ^{15}N-NSD1-P5C5:Nizp1-C2HR complex (red) and ^{15}N-NSD1-P5C5-C2HR_ Arg427His complex (green). (C) Histogram representation of the chemical shift displacement (CSD) observed per each residue of the NSD1-P5C5 domain titrated with Nizp1-C2HR_Arg427His . (D) Superimposition of the CSD plots resulting from the NMR titration of the NSD1-P5C5 domain with wild-type Nizp1-C2HR (red) and Nizp1-C2HR_ Arg427His (black). (E) The histogram shows the differences in CSDs upon binding to Nizp1-C2HR and Nizp1-C2HR_Arg427His.

4.4. Crystallization screening

In order to better understand the structural basis of the interaction between NSD1-P5C5 and Nizp1-C2HR we aimed to solve the structure of the protein complex by protein crystallography. The crystallization trials were performed on a 1:1 complex.

In the first step we searched the optimal concentration of the protein sample by a pre-crystallization test (PCT kit, from Hampton Research). We tested three initial sample concentrations, 5 mg/ml, 10 mg/ml and 15 mg/ml, using the four conditions of the kit (reagents: A1,B1,A2,B2. See Material and Method). We observed that at 5 mg/ml the protein complex did not precipitate in any condition of the kit (the crystallization drops were clear), while at 10 mg/ml and 15 mg/ml we observed that it precipitated heavily. Following the protocol we diluted the protein sample to 7.5 mg/ml and repeated the test. In this case, we observed a light granular precipitate in the A1/B1 reagents and no precipitation for the protein complex in A2/B2 reagents. This result indicated that the optimal concentration to perform the following screenings was around 7.5 mg/ml.

Next, we performed two commercial screenings (Index and the Crystal-screen), which allowed us to test 146 different conditions. After one week we selected a condition in which we observed evidence of nucleation (0.15 M magnesium chloride hexahydrate, 0.1 M BIS-TRIS ph6.5, 25% w/v PEG3.350). Nucleation is the initial formation of the molecular clusters from which the crystals grow (see Fig 32 A). Using as starting point this condition, we performed a more rigorous screening around it, varying the pH, the salt and precipitant concentrations. The better result of this optimization of the conditions was found in 0.1 M BIS-TRIS pH 6.5, 0.3 M magnesium chloride hexahydrate and 27% w/v PEG3.350. Single crystal were obtained, but further screenings are needed to increase the size and thickness of them.

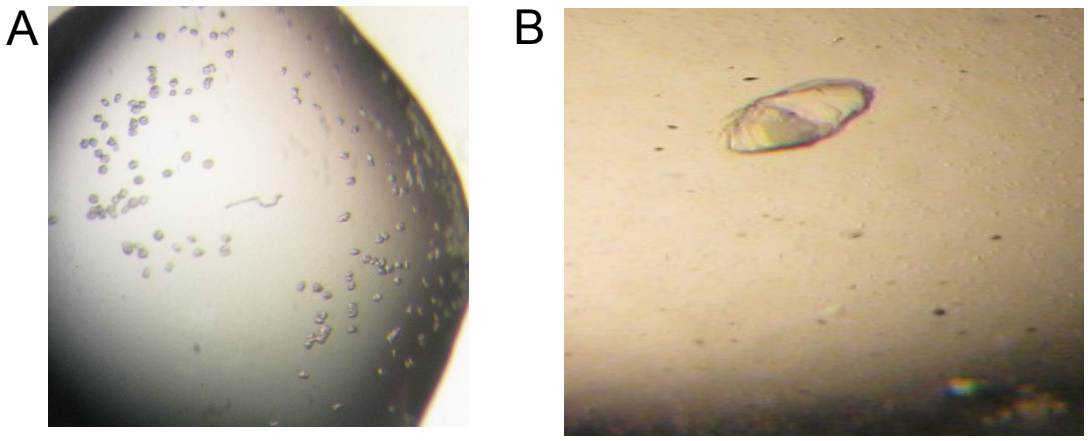


Fig 32 Crystallization screening. (A) The small spheres correspond to the nuclei of the nucleation process (B) Single crystal of the complex NSD1-P5C5:Nizp1-C2HR.

4.5. Molecular modelling of the NSD1-P5C5 and Nizp1-C2HR complex

A model of the NSD1-P5C5 and Nizp1-C2HR complex has been computed by HADDOCK server using the experimental data from the NMR titrations. The HADDOCK server is a data driven docking program commonly used to generate three-dimensional model of interacting proteins using experimental data. The structure of NSD1-P5C5 has been docked onto the structure of Nizp1-C2HR, using as restrains the residues shifting in NMR titrations. In particular, for both domains, we use as active residues the residues whose backbone NH groups are affected by the binding and as passive residues the residues whose carbon atom CSDs is equal or higher than 0.5 ppm. After the calculation HADDOCK clustered 162 structures in 10 clusters. Fig 33 shows the HADDOCK score (intermolecular energy, see Material and Method) as a function of the iRMSD (backbone RMSD at the interface) from the target, that is the the lowest energy complex structure. The cluster 3 is the best cluster with a HADDOCK score of -145.0±9.1, iRMSD = 1.5±0.8 Å and a Z-score of -1.5. The statistic of the cluster is reported in the Table 2.

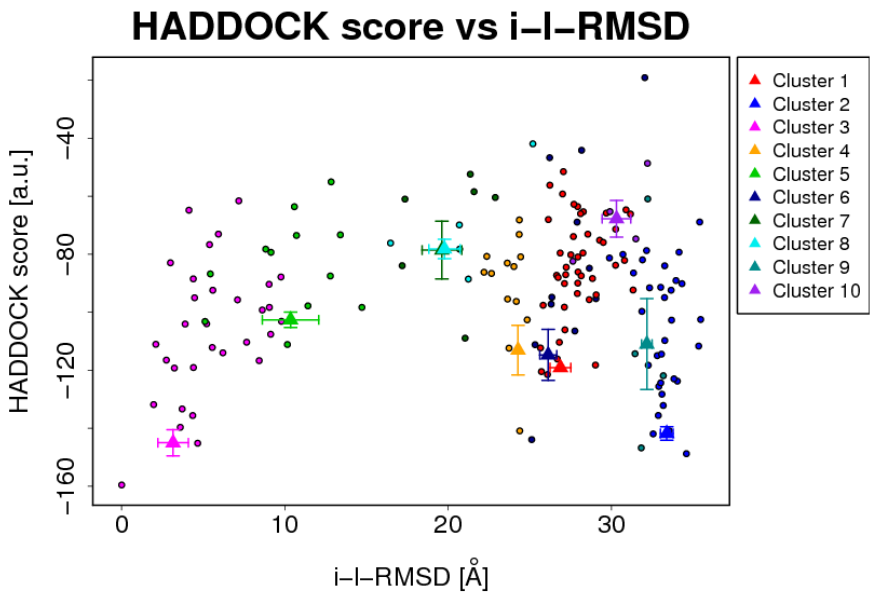


Fig 33. Graphic representation of the HADDOCK score in function of iRMSD

HADDOCK score	**-145.0± 9.1**
Cluster size	**31**
RMSD from the overall lowest-energy structure (Å)	**1.5 ± 0.8**
Van der Waals energy (kcal/mol)	**-48.3 ± 7.8**
Electrostatic energy (kcal/mol)	**-545.2 ± 70.2**
Desolvation energy(kcal/mol)	**7.8 ± 7.6**
Restraints violation energy	**45.0± 26.32**
Buried Surface Area	**1704.1 ± 182.4**
Z-Score	**-1.5**

Table 2. Statistic NSD1-P5C5:Nizp1:C2HR HADDOCK complex.

The model suggested that the formation of the complex involves the NSD1-P5C5 residues at the interface between the two PHD domains and the short α-helix of Nizp1-C2HR as indicated by CSDs (see Fig. 34). The interaction is stabilized by hydrophobic contacts and polar contacts between the two protein domains. Specifically, the Ser2143, Asp2165 and Glu2204 of NSD1-P5C5 interact with the Asn419, Arg422 and Arg417 of Nizp1-C2HR, respectively. The aromatic residue Trp416 of Nizp1-C2HR establishes hydrophobic interaction with the side-chain of Phe2122 and Met2177 of NSD1-P5C5. In future, we aim to improve this model adding to the docking calculation the molecules NOEs between the two proteins and then validate the model by site-specific mutagenesis.

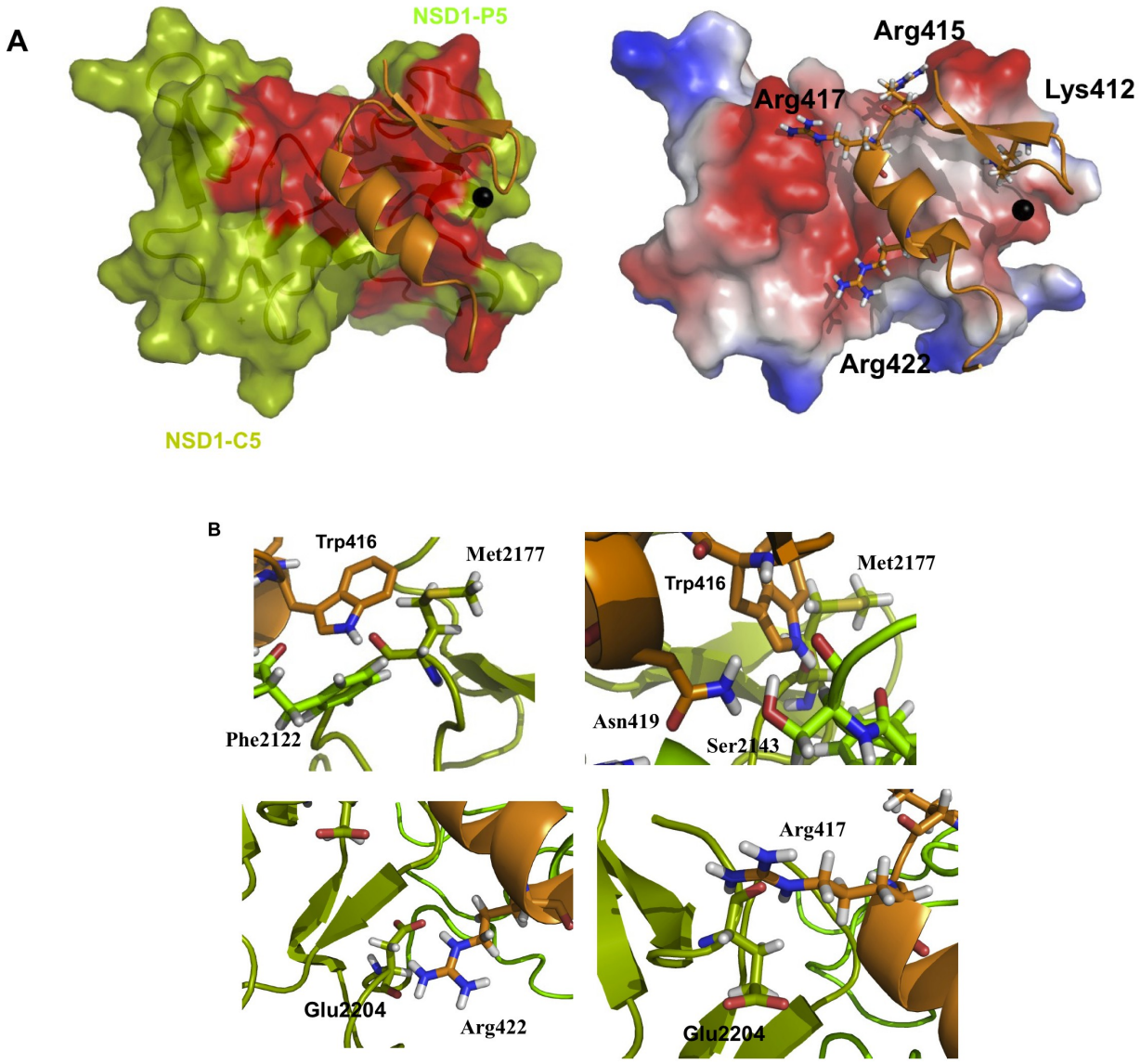


Fig 34 Haddock model of NSD1-PD5C5:Nizp1-Nizp1-C2HR complex. (A) On the left, cartoon representation of the complex model computed using the experimental data from the NMR titrations. The binding surface is coloured in red. On the left Electrostatic surface of NSD1-P5C5. The charge residueS of the Nizp1-C2HR are shown as stick (B) Close up view of the interactions that stabilize the protein complex.

5. Discussion

5.1 Introduction

The histone lysine methyltransferase NSD1 (Nuclear SET domain containing protein-1) is a transcriptional regulator that is important in the control of the expression levels of genes involved in development and differentiation. The divalent specificity to methylated H4K20 and H3K36 determines its involvement in gene silencing and activation (Rayasam et al 2003). NSD1 is a multi-modules protein that contains a SET domain, PHD finger and PWWP domains that are typically involved in numerous epigenetic events. This domain architecture is a common feature in human transcription factors. For example, the MLL1/KMT2A, a SET-domain containing protein, contains three consecutive PHD finger domains in the N-terminal part, mediating gene repression, and one PHD finger-Bromo tandem domain, in C-terminal region, having transcriptional activator function (Muzaffar A. et al 2013).

NSD1 plays a key role in human diseases: chromosomal translocation and point mutations in *NSD1* gene are linked with acute myeloid leukemia (AML) and overgrowth Sotos Syndrome, respectively. In particular, *NSD1* gene fusion to *NUP98* leads to 5% of human acute myeloid leukemia cases. The binding of the chimeric protein to its genetic targets is essential for leukaemogenesis and it is mediated by NSD1-P5C5 (Wang et al 2007). The physiological importance of this tandem domain is further highlighted by the fact that numerous mutations associated with Sotos syndrome lie in this two PHD domains. It was proposed that NSD1-P5C5 works both as histone reader and as mediator that interacts with the C2HR zinc finger domain of the transcriptional repressor Nizp1 (Nielsen et al. 2004).
Nizp1 contains five zinc finger modules, that in tandem array display DNA and protein interaction activities, and a KRAB domain that has been shown to silence the transcription through the recognition of the corepressor protein TIF1β/KAP-1 (Abrink M et al 2001). It was speculated that the binding of NSD1 to Nizp1, probably mediates NSD1 recruitment on the chromatin leading to gene repression (Nielsen et al. 2004).
Biochemical or genetic studies should be assisted by further biophysical studies to better understand this protein-protein interaction mechanism between two or more functional units. The molecular details on the interaction of NSD1-P5C5 with Nizp1-C2HR, presented

in this work, enabled us to explain the at molecular levels structural determinants at the basis of the interaction between a PHD finger and a zinc finger motif. This is the first time to our knowledge that an interaction between two zinc binding domains has been described at structural level. In the following I will discuss the structural and thermodynamic features of the interaction, that have been studied in this thesis.

5.2 NSD1-P5C5 is a tandem PHD finger involved in Sotos Syndrome.

The NSD1-P5C5 NMR structure shows that the NSD1-PHD5 presents the typical fold of the PHD fingers, while the C5HCH domain displays an atypical topology for a PHD finger with two additional antiparallel β strands in the C-terminal region. The two domains have a large interface region that holds them together and determines the formation of a highly negatively charged concave surface between the two domains (see Fig.1).

Overall the tandem domain creates a single structural unit, similarly to what observed in the homologue NSD3-P5C5 structure (RMSD = 1.49 Å) adopting a "face to side orientation" at variance to what observed in other tandem PHD fingers such as the tandem domains of MOZ (monocytic leukemia zinc finger protein, a histone acetyltransferase) or of Dpf3b (an adaptor component of BAF chromatin remodeling complex) that adopt a "head to tail orientation" (Dreveny I et al 2013, Zeng L et al 2010). The different domains orientation determines that MOZ and Dpf3b recognize the unmodified N-terminal residues and acetylated K14 of H3 with PHD2 and PHD1 respectively, wheres NSD3-P5C5 forms a complex with H3 N-terminal peptide only through the first PHD (NSD3-PHD5 or NSD3-P5) of the tandem module (see Fig. 35).

NSD3-P5C5 was identified as an epigenetic module able to recognize the trimethyled K9 on the H3 histone tail with low binding affinity (290 ±8 μM) but no binding was detected with Nizp1-C2HR by GST-pull down (He C et al 2012). Interestingly, NSD1-P5C5 and NSD3-P5C5 share about 65% sequence similarity and high structural homology, however our NMR titration and ITC experiments did not confirm the interaction with H3 histone tail (data not show).

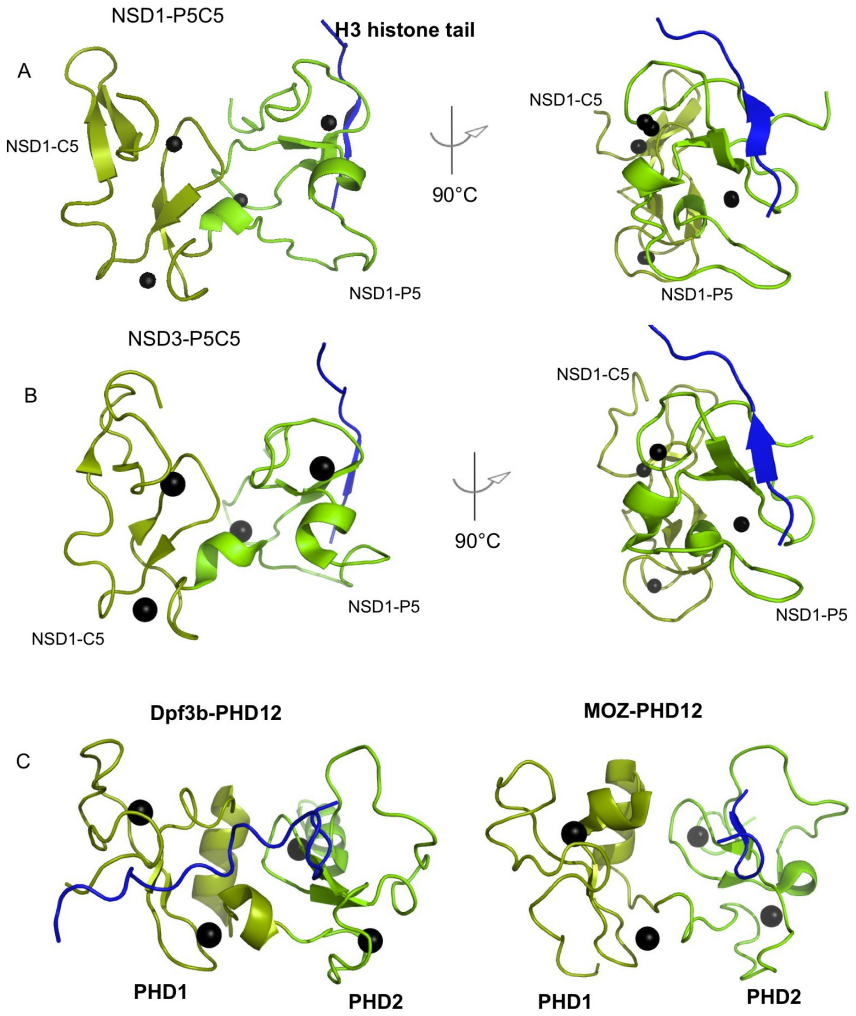


Fig 35. The PHD finger orientation. (A) Cartoon representation of the model of NSD1-P5C5 in complex with H3 histone tail (in blue). The model has been generated through the superposition with the structure of NSD3-P5C5 in complex with the H3 histone tail (B). (B) Cartoon representation of NSD3-P5C5 in complex with H3 histone tail. (C) Cartoon representation of Dpf3b-PHD12 in complex with H3 histone tail. (D) Cartoon representation of MOZ-PHD12 in complex with H3 histone tail

5.2.1 NSD1-P5C5 in the Sotos Syndrome

In the context of Sotos Syndrome NSD1 missense mutations occur in the catalytic SET domain, PWWP, and PHD finger modules, suggesting that each domain has a role in the pathology. The crystal structure of the SET domain of NSD1 and the histone lysine methyltransferase activity assays show that the missense mutations within the SET domain are responsible for the loss or reduction of the histone lysine methyltransferase activity of NSD1, which may induce aberrant lysine methylation pattern during development (Qiao et al 2011).

The NSD1-P5C5 structure allowed us to rationalize the impact of the Sotos mutations on the

tandem domains showing that Tyr2142Asn, Arg2152Gln, Phe2182Ile, His2143Glu, His2143Tyr, His2164Arg and His2206Arg, compromise the protein fold. The histidine mutations, affect the zinc ion binding sites, destabilizing the protein structure (His2143Glu, His2143Tyr, His2164Arg) or partial alterating the secondary structural element (His2206Arg). Interestingly, the Tyr2142 and Phe2182 mutations, that involve important residues of the hydrophobic core have a detrimental effect on the protein structure leading to a complete unfolding of the tandem domain. Of note only one mutation, Arg2152Gln, that is located in a solvent exposed patch on the surface of NSD1-P5C5, does not affect the protein fold. The misfolding mutations probably determine the loss of functions associated with the protein domain and thus alter the biological role of NSD1, such as cellular localization or its recruitment on the chromatin. Conversely, the mutations that only partially affect the protein fold or are solvent exposed could compromise the interaction with other regulatory partners. To verify this hypothesis we have characterized the effects of the NSD1-P5C5_His2206Arg and NSD1-P5C5_Arg2152Gln pathological mutants on the interaction with the zinc finger module C2HR of the corepressor Nizp1 with (see Results, section 4.3.4). The first disease-associated mutation (NSD1-P5C5_His2206Arg) is localized near the Nizp1-C2HR interaction region and probably destabilizes this binding event (see Result 3.3.4). The second pathological substitution (Arg2152Gln) is far from the zinc finger binding surface, thus it might not have impact on Nizp1-C2HR interaction, as indicated by the NMR titration. We hypothesise that the protein surface near the Arg2152 is involved with the recruitment of other specific NSD1 cofactors that we aim to identify in future by proteomic studies.

In the last years, several disease-associated mutations affecting the PHD fingers have been structurally and functionally characterized. For example, the transcriptional factor AIRE containing two PHD fingers (PHD1 and PHD2), normally localizes in punctate nuclear bodies, where it interacts with unmodified H3 histone tail through its AIRE-PHD1, and it partners with a number of chromatin related proteins controlling gene expression of a repertoire of peripheral-tissue antigens (PTAs) to mediate deletional tolerance, thereby preventing self-reactivity (Gaetani M et al 2012). Of note, the two PHD fingers are sites of pathological mutations that cause the autoimmune disorder APECED (autoimmune polyendocrinopathy ectodermal dystrophy) (Lindsey AB et al 2008). In particular, the

Cys443Gly mutation in AIRE-PHD2, impairs metal binding, completely destroying the fold of the domain. Herewith it compromises the nuclear protein localization reducing the activation of AIRE-dependent genes (Gaetani M et al 2012). The mutation Val301Met in AIRE-PHD1 it is solvent exposed and maintainers the structure fold. It is localized in a distinct region with respect to histone binding site, thus it does not affect the interaction with H3 histone tail (Chignola F et al 2009). Conversely, it has a stronger impact on AIRE interactome reducing the interactions with protein partners involved in the transcriptional activation (Gaetani M et al 2012).

These observations support the notion that different surface in a PHD finger are involved in several protein interactions. We therefore expect that the Arg2152Gln mutation will affect NSD1 interactome, thus compromising the recruitment of important co-factors to chromatin and possibly affecting the transcriptional events regulated by NSD1. In future proteomics studies comparing NSD1 interactome with its wild-type and mutated form will give useful insights in NSD1 function.

5.3 Nizp1-C2HR is a zinc finger module with unusual Zn^{2+} coordination

No direct interactors of NSD1 are known except Nizp1 that interacts with NSD1-P5C5 via its C2HR domain (Nielsen AL et al 2004). Nizp1-C2HR is an atypical Cys2His2-type zinc finger in which the fourth chelating residue is substituted by arginine residue (Arg427).

Zn^{2+} ion is crucial for the protein fold as the zinc depletion unfolds the domain. Nizp1-C2HR has a typical zinc finger fold consisting of an α-helix (residue Arg417-Arg425) and a β-hairpin (β1 and β2 residues Tyr405-Cys407 and Gly411-Phe414, respectively) hold together by one zinc ion. The small hydrophobic core is composed by Phe417, Phe420 and His423. The Trp416 is not conserved in the others Nizp1 zinc fingers, and it is partially solvent exposed suggesting its possible involvement in a protein binding event. The electrostatic potential of Nizp1-C2HR on the solvent exposed side of the α-helix is positive suggesting a role as binding region.

The metal binding site is composed by three ligands, two Cys (Cys407 and Cys410) and one His (His423). This kind of coordination is not unusual in zinc finger modules: in several domains the fourth zinc ligand is not necessarily an His or Cys residue and it can be

substituted by an aspartate or glutamic acid. In other cases a water molecule can function as fourth Zn^{2+} coordinating ligand. For example, in the AFV1p06-zinc finger structure the fourth ligand of the zinc ion is the side-chain oxygen atom of the carboxylic group of glutamic acid residue (Guilliere F et al 2013), while in NEMO zinc finger a water molecule coordinates the Zn^{2+} ion upon mutation of a cysteine into phenylalanine, indicating that mutations of Zn^{2+}chelating residues not always lead to domain unfolding (Cordier F et al 2008).

In conclusion, three zinc chelating residues can be sufficient to stably coordinate metal ion. From our structure and NOE data, it was evident that the C-terminal region (Ser426-Lys434) is disordered. The backbone carbonyl oxygen of the Arg427 is far from the metal binding site. Moreover mutation of this residue into alanine produces a well folded protein indicating the side-chain of the residue in position 427 is not involved in Zn^{2+} binding. We exclude also that Glu429 and His433 coordinate zinc as they are structurally distant from the Zn^{2+} binding site. We observed that Asn409 is near the zinc binding site, however, its replacement by Ala residue did not affect the protein fold, suggesting that it is not involved in metal coordination. We speculate that a water molecule could be close to the zinc ion thus acting as a fourth ligand, a common situation in catalytic zinc-binding site of enzymes (Sankaranayanan R et al 1999). We next mutagenized Arg427 into a His residue: the presence of the zinc-ligand in this position determines the formation of a classical zinc binding site (Cys2His2). The structure of the mutant calculated by NMR spectroscopy shows that it is similar to the wild type one, and the histidine residue induces an additional helical turn that extends the alfa-helix after the position 427.

Taken together, our data are consistent with those previously published on other the zinc finger modules. The biophysical characterization of several Cys2HisX mutants (X=Asp, Asn, Glu, Arg or Ala) showed that the fourth chelating residue is not essential for zinc binding and that three zinc-ligands are sufficient to maintain the zinc finger fold and its functionality. An example of Zn^{2+} coordination by only three ligands is given by the transcriptional repressor BKLF. It contains a tandem array of zinc finger domains (Cys2His2) whereby the last zinc finger module is involved in the DNA binding (Simpson RJ et al 2003). In this module, the substitution of the fourth coordinant (Histidine) with an

aspartate, glutamate, asparagine, glutamine, alanine, and arginine did not affect the metal binding and the ability to recognize the DNA sequence, thus maintaining its ability to promote the transcriptional activation (Simpson RJ et al 2003).

We have characterized the interaction between Nizp1-C2HR_Asn409Ala and Nizp1-C2HR_Arg427Ala/His mutants with NSD1-P5C5 (see Results section 4.3.5). The NMR titrations show that the alanine substitutions did not affect the binding with NSD1-P5C5. Interestingly, also the Arg427His substitution maintains its interaction with NSD1-P5C5 as the wild type domain. This latter result is not in agreement with the two hybrid assays published by Nielsen *et al* in which the Arg427His substitution seems to hinder the interaction with NSD1-P5C5. This point will be further discussed in the next section.

5.4 NSD1-P5C5 binds Nizp1-C2HR: identification of the binding surfaces by NMR titrations

5.4.1 NSD1-P5C5 binding site mapping

We performed NMR titrations to establish the interaction surfaces between NSD1-P5C5 and Nizp1-C2HR. In NSD1-P5C5, the NMR chemical-shift mapping shows that the binding site is localized at the interface between the two PHD fingers involving residues Phe2122, Ser2123, Cys2124, Ser2132, Cys2133, His2143, Tyr2142, Cys2146, Leu2147, Ala2152, Trp2155, Glu2156, Glu2176 on NSD1-PHD5, and residues Cys2178, Pro2179, Ser2180, Ser2181, Glu2204 and His2205 on NSD1-C5HCH. These data indicate that both the modules, combined in a single functional unit, are involved in the Nizp1-C2HR recruitment (see Results 4.3.2). This concave binding surface is formed by hydrophobic resides surrounded by polar residues that could interact with the zinc finger motif through hydrophobic and polar contacts, respectively. We observed that the surface involved in this protein interaction is distinct from the predicted histone binding site (see Fig. 36). Further structural studies are needed to clarify the role, if any, of the NSD1-P5C5:Nizp1-C2HR complex in the recognition of histone tails.

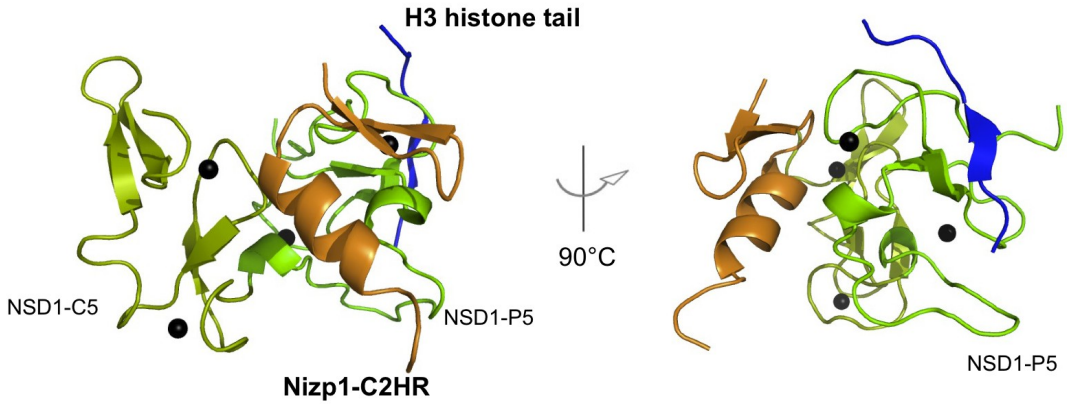


Fig. 36. Model of NSD1-P5C5:Nizp1-C2HR in complex with the H3 histone tail. Cartoon representation of the tertiary complex in which the NSD1-P5 and NSD1-C5 are colored in light green and dark green, respectively. The Nizp1-C2HR domain and histone tail are colored in orange and blue, respectively.

In the literature other cases have been described in which the combination of PHD fingers or PHD finger and other epigenetic modules (such as bromodomains), function together to read epigenetic marks and to interact with non-histone proteins. An example is given by MLL1 PHD3-bromodomain. In this case the PHD3 finger and the bromodomain adopt a compact fold whereby the bromodomain regulates the function of the PHD3 that can bind simultaneously, through distinct surfaces, both the histone H3K4me2/3 markers and the suppressor factor Cyp33 . In the apo-state the bromodomain increases the binding affinity of the PHD3 finger for H3K4me3 by 20 fold. In the presence of Cyp33 the interactions between the PHD3 and bromodomain are disrupted and the PHD finger can interact with the RNA recognition motif (RRM) of Cyp33. The PHD3:Cyp33 complex is also involved in the interaction with H3K4me2/3 and the dissociation constant for the histone peptides is doubled when compared to the PHD-bromodomain alone. The ability of PHD3 to bind both histone peptides and non-histone protein reinforces the concept that PHD fingers can function not only as reader but also as mediators of interactions whit other regulatory proteins.

5.4.1 Nizp1-C2HR binding site mapping

The Nizp1-C2HR region involved in the interaction with NSD1-P5C5 comprises part of the β2-strand (Lys412, Phe414), the loop (Trp416) and the short alfa-helix (Arg417, Val418, Asn419, Phe420, Arg422 and His423). The strongest CSDs affects the alfa-helix (Arg417-His423) indicating that this structural element is fundamental for the interaction (see Results

4.3.3).

Usually, in zinc finger modules the alfa-helix has both DNA and protein binding activity. For example the human transcriptional regulator FOG-1 ('Friend of GATA-1') contains nine classical zinc finger domains, whereby five can mediate protein–protein interactions binding the N-terminal GATA-1 transcription factor (Fox et al. 1999; Simpson et al. 2004). This interaction mediates the recruitment of FOG-1 to gene promoters giving rise to changes in gene expression that are essential for normal erythro and megakaryopoiesis in mammals. Structural and mutagenesis studies suggest that the protein binding motif is localized on the α-helix of FOG zinc finger domains (Simpson et al. 2004; Garriga-Canut e Orkin 2004), of note, the very same region is involved in DNA-binding other FOG-1 zinc fingers, highlighting the multitasking role of these domains (Gamsjaeger et al. 2007).

5.4.2 HADDOCK model of the complex NSD1-P5C5:Nizp1-C2HR

Based on NMR chemical shift mapping data we have obtained a data driven docking model (using HADDOCK) of the protein complex. In the model the short alfa-helix of Nizp1-C2HR is localized at the NSD1-P5C5 interface and hydrophobic and polar contacts stabilize the protein complex (see Fig 34). In particular, Ser2143, Asp2165 and Glu2204 of NSD1-P5C5 interact with Asn419, Arg422 and Arg417 of Nizp1-C2HR, respectively. The interactions involve not only the residues on the Nizp1-C2HR α-helix but also Trp416 that inserts its side-chain between the Phe2122 and Met2177 of NSD1-P5C5 at the PHD fingers interface (see Result section 4.5). Further experimental data (side specific mutagenesis) will be necessary to confirm these interactions. Moreover we aim to collect inter-protein NOE restrains that will be introduced in new data driven docking calculations, that will improve the quality of the model. Finally, a detailed analysis on the interactions between NSD1-P5C5 and Nizp1-C2HR will have to await the solution of the protein complex by X-ray crystallography.

5.5 The binding reaction between NSD1-P5C5 and Nizp1-C2HR is entropy-driven.

The thermodynamic characterization of the interaction, by ITC experiment, shows that the binding reaction has a $K_d = (3{,}80 \pm 0{,}66) \cdot 10^6$ M and is entropy-driven ($\Delta H = +4.14$ kcal/mol

and TΔS = +14.19 kcal/mol).

The changes in entropy are related to solvent effects such as loss of ligand-solvent interactions and solvent reorganization near the protein surface. Any binding event displaces water molecules from the interaction surface. While the vast majority of water molecules in the binding site are mobile and easy displaceable, some are tightly bound to the protein structure. The release of these bound water molecules upon ligand binding affects the thermodynamic signature of the binding event resulting in an entropically favourable process (Breiten B et al 2013). In the binding reaction the entropic and enthalpy contributions are related. This phenomenon, referred to as enthalpy/entropy compensations describes the linear relationship between the change in enthalpy and in entropy. This means that favourable changes in the binding entropy are balanced by opposite changes in binding enthalpy and vice versa, resulting in very small changes in the overall free energy (Bronowska AK, 2012).

In our case, the large positive entropy is compensated by small positive enthalpy changes. Interestingly, in the HADDOCK model, the NSD1-P5C5:Nizp1-C2HR complex is stabilized by several polar contacts between charged residues. The side-chain of these residues are likely to form hydration shells with the water molecules in the free form. Thus, it appears that the thermodynamic parameters observed in the formation of the complex reflect both the release of water molecules from the binding site and the desolvation of the ligand. Similar mechanisms are observed in other biomolecular interactions, such as the interaction between the peripheral subunit-binding domain (PBSD) of the dihydrolipoyl acetyltranferase and the dihydrolipoyl dehydrogenase (Jung H et al 2002), in which the excluding water from the binding surface serves as driving force in molecular association.

5.6 Hypothesis on NSD1 gene regulation mechanism

Berdasco *at al 2010* showed that the silencing of NSD1 is involved both in up-regulation and in down-regulation of several development and differentiation genes. This duality is probably linked to NSD1 ability to methylate both H4K20 and H3K36 (Rayasam GV et al 2003). The H4K20 methylation appears to be associated with transcription silencing and the H3K36 methylation has been found primarily in active genes throughout the gene body, even though it is also present in inactivated genes, depending on the context and genome

loci. In this scenario, we can speculate that all the NSD1 epigenetic modules interacting with histone post-transcriptional modifications and non-histone proteins can direct the NSD1 histone methylation activities, depending on cellular context.

NSD1 is member of NSD protein family that contains other two proteins called NSD2 and NSD3, that possess a C-terminal tandem domain. However, these domains fail to interact with Nizp1, suggesting that this protein-protein interaction is specific for NSD1 (Nielsen at al 2004). Several data are available on the NSD2 methylation activity and the role of its PHD fingers (Kuo JA et al 2011). This NSD protein shows a single methylation specificity for H3K36. Experimental data indicate that the PHD1,2,3 are involved in recruitment of NSD2 on the chromatin, and in particular only PHD2 is fundamental for the methylation of H3K36, while the P5C5 tandem domain has no impact on histone methylation. The NSD1-PHD1,2,3 share a high homology protein sequence (40%) with NSD2-PHD1,2,3. It is conceivable that also the first three NSD1-PHD fingers contribute to NSD1 chromatin localization and that the interaction of NSD1-P5C5 with the repressor Nizp1 via its C2HR motif regulates methylation activity of NSD1 to specific promoters.

5.7 Conclusions and future works

NSD1 is a versatile protein that can act as coactivator and corepressor. This diversified function is probably due to the presence of different chromatin related domains that can have opposite outcomes, depending on the cellular context.

Here, we have investigated the structure of the tandem PHD finger NSD1-P5C5 and characterized the structural effects of Sotos pathological mutations by NMR spectroscopy. We have also characterized its interaction with Nizp1-C2HR, an unconventional Zinc-finger motif, using a combination of NMR, ITC and computational methods. NMR titrations showed that Sotos mutations affect binding to Nizp1, in future we will perform ITC measurements to thermodynamically characterize the effects of these mutations on the binding. Finally the X-ray structure of the complex will give useful and definite insights into the details of the interaction. Future work will be dedicated to the investigation of the interplay between binding of Nizp-C2HR to NSD1-P5C5 and the interaction with histone tails.

Bibliography

Adwan, L., and Zawia, N.H. (2013). Epigenetics: a novel therapeutic approach for the treatment of Alzheimer's disease. Pharmacol. Ther. *139*, 41–50.

Andreini, C., Bertini, I., and Cavallaro, G. (2011). Minimal functional sites allow a classification of zinc sites in proteins. PLoS ONE *6*, e26325.

Archakov, A.I., Govorun, V.M., Dubanov, A.V., Ivanov, Y.D., Veselovsky, A.V., Lewi, P., and Janssen, P. (2003). Protein-protein interactions as a target for drugs in proteomics. Proteomics *3*, 380–391.

Arrowsmith, C.H., Bountra, C., Fish, P.V., Lee, K., and Schapira, M. (2012). Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov *11*, 384–400.

Baker, L.A., Allis, C.D., and Wang, G.G. (2008a). PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. Mutat. Res. *647*, 3–12.

Baker, L.A., Allis, C.D., and Wang, G.G. (2008b). PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. Mutat. Res. *647*, 3–12.

Baxevanis, A.D., and Landsman, D. (1996). Histone Sequence Database: a compilation of highly-conserved nucleoprotein sequences. Nucleic Acids Res. *24*, 245–247.

Bedford, M.T., and Clarke, S.G. (2009). Protein arginine methylation in mammals: who, what, and why. Mol. Cell *33*, 1–13.

Bedford, M.T., and Richard, S. (2005). Arginine methylation an emerging regulator of protein function. Mol. Cell *18*, 263–272.

Berdasco, M., Ropero, S., Setien, F., Fraga, M.F., Lapunzina, P., Losson, R., Alaminos, M., Cheung, N.-K., Rahman, N., and Esteller, M. (2009). Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. Proc. Natl. Acad. Sci. U.S.A. *106*, 21830–21835.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature *447*, 407–412.

Botuyan, M.V. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell *127*, 1361–1373.

Bouhouche, N., Syvanen, M., and Kado, C.I. (2000). The origin of prokaryotic C2H2 zinc finger regulators. Trends Microbiol. *8*, 77–81.

Chignola, F., Gaetani, M., Rebane, A., Org, T., Mollica, L., Zucchelli, C., Spitaleri, A., Mannella, V., Peterson, P., and Musco, G. (2009). The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. Nucleic Acids Res. *37*, 2951–2961.

Davie, J.R., and Spencer, V.A. (1999). Control of histone modifications. J. Cell. Biochem. *Suppl*

32-33, 141–148.

Dreveny, I., Deeves, S.E., Fulton, J., Yue, B., Messmer, M., Bhattacharya, A., Collins, H.M., and Heery, D.M. (2014). The double PHD finger domain of MOZ/MYST3 induces α-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. Nucleic Acids Res. *42*, 822–835.

Fox, A.H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999). Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. EMBO J. *18*, 2812–2822.

Gamsjaeger, R., Liew, C.K., Loughlin, F.E., Crossley, M., and Mackay, J.P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. Trends Biochem. Sci. *32*, 63–70.

Garriga-Canut, M., and Orkin, S.H. (2004). Transforming acidic coiled-coil protein 3 (TACC3) controls friend of GATA-1 (FOG-1) subcellular localization and regulates the association between GATA-1 and FOG-1 during hematopoiesis. J. Biol. Chem. *279*, 23597–23605.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Yuung, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–78.

Guillière, F., Danioux, C., Jaubert, C., Desnoues, N., Delepierre, M., Prangishvili, D., Sezonov, G., and Guijarro, J.I. (2013). Solution structure of an archaeal DNA binding protein with an eukaryotic zinc finger fold. PLoS ONE *8*, e52908.

Huang, N., vom Baur, E., Garnier, J.M., Lerouge, T., Vonesch, J.L., Lutz, Y., Chambon, P., and Losson, R. (1998). Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. EMBO J. *17*, 3398–3412.

Hung, T., Binda, O., Champagne, K.S., Kuo, A.J., Johnson, K., Chang, H.Y., Simon, M.D., Kutateladze, T.G., and Gozani, O. (2009). ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. Mol. Cell *33*, 248–256.

Hyllus, D., Stein, C., Schnabel, K., Schiltz, E., Imhof, A., Dou, Y., Hsieh, J., and Bauer, U.-M. (2007). PRMT6-mediated methylation of R2 in histone H3 antagonizes H3 K4 trimethylation. Genes Dev. *21*, 3369–3380.

Iberg, A.N., Espejo, A., Cheng, D., Kim, D., Michaud-Levesque, J., Richard, S., and Bedford, M.T. (2008). Arginine methylation of the histone H3 tail impedes effector binding. J. Biol. Chem. *283*, 3006–3010.

Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., Schultz, D.C., Psulkowski, E., Fredericks, W.J., White, D.E., Maul, G.G., et al. (2007). PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. Mol. Cell *28*, 823–837.

Jakovcevski, M., and Akbarian, S. (2012). Epigenetic mechanisms in neurological disease. Nat. Med. *18*, 1194–1204.

Kouzarides, T. (2007a). Chromatin modifications and their function. Cell *128*, 693–705.

Kouzarides, T. (2007b). SnapShot: Histone-modifying enzymes. Cell *131*, 822.

Kurotaki, N., Harada, N., Yoshiura, K., Sugano, S., Niikawa, N., and Matsumoto, N. (2001). Molecular characterization of NSD1, a human homologue of the mouse Nsd1 gene. Gene *279*, 197–204.

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature *448*, 718–722.

Li, Y., Trojer, P., Xu, C.-F., Cheung, P., Kuo, A., Drury, W.J., Qiao, Q., Neubert, T.A., Xu, R.-M., Gozani, O., et al. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. J. Biol. Chem. *284*, 34283–34295.

Lucio-Eterovic, A.K., Singh, M.M., Gardner, J.E., Veerappan, C.S., Rice, J.C., and Carpenter, P.B. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. Proc. Natl. Acad. Sci. U.S.A. *107*, 16952–16957.

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

Malecek, K., and Ruthenburg, A. (2012). Validation of histone-binding partners by peptide pull-downs and isothermal titration calorimetry. Meth. Enzymol. *512*, 187–220.

Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G., and Mackay, J.P. (2011). Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. J. Biol. Chem. *286*, 11779–11791.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat. Rev. Mol. Cell Biol. *6*, 838–849.

Mellor, J. (2006). It takes a PHD to read the histone code. Cell *126*, 22–24.

Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. *4*, 1609–1614.

Miller, T.C.R., Mieszczanek, J., Sánchez-Barrena, M.J., Rutherford, T.J., Fiedler, M., and Bienz, M. (2013). Evolutionary adaptation of the fly Pygo PHD finger toward recognizing histone H3 tail methylated at arginine 2. Structure *21*, 2208–2220.

Morishita, M., and di Luccio, E. (2011a). Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1. Biochem. Biophys. Res. Commun. *412*, 214–219.

Morishita, M., and di Luccio, E. (2011b). Cancers and the NSD family of histone lysine methyltransferases. Biochim. Biophys. Acta *1816*, 158–163.

Musco, G., and Peterson, P. (2008). PHD finger of autoimmune regulator: an epigenetic link between the histone modifications and tissue-specific antigen expression in thymus. Epigenetics *3*, 310–314.

Nielsen, A.L., Jørgensen, P., Lerouge, T., Cerviño, M., Chambon, P., and Losson, R. (2004). Nizp1, a novel multitype zinc finger protein that interacts with the NSD1 histone lysine methyltransferase through a unique C2HR motif. Mol. Cell. Biol. *24*, 5184–5196.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.

Pasillas, M.P., Shah, M., and Kamps, M.P. (2011). NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome. Hum. Mutat. *32*, 292–298.

Patel, D.J., and Wang, Z. (2013a). Readout of epigenetic modifications. Annu. Rev. Biochem. *82*, 81–118.

Peterson, C.L., and Laniel, M.-A. (2004). Histones and histone modifications. Curr. Biol. *14*, R546–551.

Pírez, N., Christmann, B.L., and Griffith, L.C. (2013). Daily rhythms in locomotor circuits in Drosophila involve PDF. J. Neurophysiol. *110*, 700–708.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat. Rev. Mol. Cell Biol. *10*, 192–206.

Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., and Xu, R.-M. (2011). The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. J. Biol. Chem. *286*, 8361–8368.

Rahman, N. (2005a). Mechanisms predisposing to childhood overgrowth and cancer. Curr. Opin. Genet. Dev. *15*, 227–233.

Rahman, N. (2005b). Mechanisms predisposing to childhood overgrowth and cancer. Curr. Opin. Genet. Dev. *15*, 227–233.

Ramón-Maiques, S., Kuo, A.J., Carney, D., Matthews, A.G.W., Oettinger, M.A., Gozani, O., and Yang, W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. Proc. Natl. Acad. Sci. U.S.A. *104*, 18993–18998.

Rathert, P., Dhayalan, A., Ma, H., and Jeltsch, A. (2008). Specificity of protein lysine methyltransferases and methods for detection of lysine methylation of non-histone proteins. Mol Biosyst *4*, 1186–1190.

Rayasam, G.V., Wendling, O., Angrand, P.-O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. EMBO J. *22*, 3153–3163.

Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., et al. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat. Struct. Mol. Biol. *19*, 1155–1160.

Sanchez, R., and Zhou, M.-M. (2011). The PHD finger: a versatile epigenome reader. Trends

Biochem. Sci. *36*, 364–372.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.-H.A., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G.E., Fingerman, I.M., Briggs, S.D., et al. (2007). Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. J. Biol. Chem. *282*, 2450–2455.

Simpson, R.J.Y., Cram, E.D., Czolij, R., Matthews, J.M., Crossley, M., and Mackay, J.P. (2003). CCHX zinc finger derivatives retain the ability to bind Zn(II) and mediate protein-DNA interactions. J. Biol. Chem. *278*, 28011–28018.

Simpson, R.J.Y., Yi Lee, S.H., Bartle, N., Sum, E.Y., Visvader, J.E., Matthews, J.M., Mackay, J.P., and Crossley, M. (2004). A classic zinc finger from friend of GATA mediates an interaction with the coiled-coil of transforming acidic coiled-coil 3. J. Biol. Chem. *279*, 39789–39797.

Tatton-Brown, K., and Rahman, N. (2007). Sotos syndrome. Eur. J. Hum. Genet. *15*, 264–271.

Tatton-Brown, K., Douglas, J., Coleman, K., Baujat, G., Cole, T.R.P., Das, S., Horn, D., Hughes, H.E., Temple, I.K., Faravelli, F., et al. (2005). Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with NSD1 aberrations. Am. J. Hum. Genet. *77*, 193–204.

Taverna, S.D., Ilin, S., Rogers, R.S., Tanny, J.C., Lavender, H., Li, H., Baker, L., Boyle, J., Blair, L.P., Chait, B.T., et al. (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. Mol. Cell *24*, 785–796.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Trojer, P. (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. Cell *129*, 915–928.

Tsai, W.-W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.-Y., Shi, X., Schwarzer, D., Plunkett, W., et al. (2010). TRIM24 links a non-canonical histone signature to breast cancer. Nature *468*, 927–932.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., et al. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967–980.

Visser, R., Landman, E.B.M., Goeman, J., Wit, J.M., and Karperien, M. (2012). Sotos syndrome is associated with deregulation of the MAPK/ERK-signaling pathway. PLoS ONE *7*, e49229.

Wang, G.G., Cai, L., Pasillas, M.P., and Kamps, M.P. (2007). NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nat. Cell Biol. *9*, 804–812.

Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.-L., Patel, D.J., and Allis, C.D. (2009). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature *459*, 847–851.

Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.-M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature *466*, 258–262.

Ringraziamenti

Al termine di questo lavoro voglio ingraziare il mio capo Giovanna Musco per il supporto e la formazione che mi ha offerto durante questi tre anni di dottorato. Non è stato un progetto semplice e gli errori sono stati frequenti, ma alla fine abbiamo trovato un modo per capirci e capire le cose su cui stiamo lavorando. Sono molto felice di essere stato un membro del tuo gruppo durante questi anni, è stata una bella esperienza di vita. Voglio ringraziare Giacomo Quilici per il supporto nella parte sperimentale legata alla spettroscopia NMR, senza le sue competenze non sarei riuscito a portare a termine il mio progetto. Ringrazio Maria (la spagnola) per avermi aperto un nuovo mondo legato alla cristallografia. Ringrazio Davide Gaudesi per avermi aiutato nell'espressione e nella purificazione di una proteina un po' “capricciosa”. Ringrazio Andrea Spitaleri per le risate e i momenti di svago. Ringrazio Michela Ghitti per i consigli e il supporto informatico specialmente nelle ultime settimane. Ringrazio Valeria Mannella per il supporto grafico e Chiara Zucchelli per il supporto nella vita da laboratorio, i suoi sono stati consigli molto preziosi. Ringrazio Cristina (la ragazza di Predore) per i suoi script che mi aiutano a lavorare meglio. Ringrazio Ferdando (lo spagnolo) per aiutarmi tutte le volte che ho un problema informatico. Ringrazio Francesca Nardella per i caffè e il tempo passato insieme. Ringrazio Laurette (la francese) per il cioccolato. Ringrazio Paola Tornaghi per la sua saggezza e per aver ascolto i mie confusi pensieri. Ringrazio Claudia Minici per i suoi consigli lavorativi. Ringrazio Pierluigi Previtali per il lavoro che fa, senza di lui non avrei neanche un computer.