SAPIENZA
UNIVERSITÀ DI ROMA

DOTTORATO DI RICERCA IN BIOCHIMICA

CICLO XXV (A.A. 2009-2012)

Structural and functional characterization of Nucleophosmin domain associated with Acute Myeoloid Leukemia

Docente guida

Prof. Maurizio Brunori

Coordinatore

Prof. Paolo Sarti

Dottorando

Carlo Lo Sterzo

Dicembre 2012

SAPIENZA
UNIVERSITÀ DI ROMA

DOTTORATO DI RICERCA IN BIOCHIMICA

CICLO XXV (A.A. 2009-2012)

Structural and functional characterization of Nucleophosmin domain associated with Acute Myeoloid Leukemia

Docente guida

Prof. Maurizio Brunori

Coordinatore

Prof. Paolo Sarti

Dottorando

Carlo Lo Sterzo

Immagine in Copertina:

Pablo Picasso
"Violino"
1912
Museo Picasso, Parigi

Aknowledgments

I wish to thank Prof. Maurizio Brunori for welcoming me in his team and for always encouraging me and believing in my potential. His stimulating suggestions were always of great value for my human and scientific development.

I wish to thank Prof. Ivano Bertini and Prof. Lucia Banci and their group, expecially Angelo Gallo and Leonardo Gonnelli, for welcoming me in Florence and for their precious contribution both in performing NMR experiments and in manuscript writing and revision.

My thanks go also to Prof. Brunangelo Falini for introducing us to the fascinating and challenging world of Nucleophosmin and for his precious suggestions.

I like to aknowledge Prof. Alessandro Arcovito for his precious contribution in Biacore experiments and for his brilliant intuitions in protein-DNA studies.

I am extremely thankful to Luca Federici and Adele Di Matteo for their continuous support during my PhD period. Their fundamental help in the laboratory has been a great growing experience for my interest and passion for science, and for my entire life.

I want to express my most sincere appreciation to Flavio Scaloni who has always been extremely generous and taught me the fundamentals of this work, Linda Celeste Montemiglio, for her smiling company, Sara Chiarella for her constant presence, Patrizio Di Micco and Fulvio Saccoccia that always showed me their sincere friendship. I want also to thank the non mentioned friends from the lab for the time spent together.

1 Introduction

1.1 Nucleophosmin

Nucleophosmin (NPM1, also known as B23, Numatrin or NO38; Grisendi et al. 2006), was first identified as a nucleolar phosphoprotein expressed at high levels in the granular region of the nucleolus (Spector et al. 1984). NPM1 was soon recognized to play a role in the regulation of cell growth, proliferation and transformation (Feuerstein et al. 1988), based on the observation that its expression rapidly increases in response to mitogenic stimuli, and that increased amounts of the protein are detected in highly proliferating and malignant cells (Chan et al. 1989). Since then, the understanding of NPM1 biology has considerably increased, revealing a complex scenario: NPM1 proved to be a multifunctional protein that is involved in many cellular activities, and which has been related to both proliferative and growth-suppressive roles in the cell (Di Fiore 2008). Furthermore, in the mouse, *NPM1* is an essential gene whose inactivation in the germ line leads to a series of developmental defects and embryonic lethality at mid-gestation (Grisendi et al. 2005).

Much of the interest in *NPM1* has been fostered by the fact that this gene is implicated in human tumorigenesis. On the one hand, NPM1 is frequently over-expressed in solid tumors of diverse histological origin, whereas on the other hand, the *NPM1* locus is involved in chromosomal translocations or deleted in various kinds of haematological malignancies and solid tumors. Strikingly, *NPM1* has also been found mutated and aberrantly localized in the cytoplasm of leukaemic blasts in a high proportion (around 35%) of patients

with acute myeloid leukaemia (AML) (Falini et al. 2005). This makes of *NPM1* the most frequently mutated gene in AML.

1.2 Nucleophosmin is a nuclear-cytoplasmic shuttling protein

NPM1 is a ubiquitously expressed protein belonging to the nucleoplasmin family of nuclear chaperones (Frehlick et al. 2007). It is encoded by the *NPM1* gene that, in humans, maps to chromosome 5q35 (Chan et al. 1989). Three nucleophosmin isoforms are generated through alternative splicing. NPM1 (or B23.1), the dominant isoform (accession n. NM_002520) (Wang et al. 1993), is a 294-amino acid phospho-protein of about 37 kDa that is expressed in all tissues. NPM1.2 (or B23.2) uses an alternative terminal exon compared with variant 1, resulting in a 259-amino acid protein (accession n. NM_001037738) with a distinct C-terminus and it accounts for minimal nucleophosmin content in tissues. A third variant (accession number NM_199185) lacks an alternate in-frame exon compared with variant 1, resulting in a shorter protein whose functions and expression pattern are still under investigation. Interestingly, NPM1 (B23.1) and NPM1.2 (B23.2) have different subcellular distribution patterns (Wang et al 1993): while NPM1 protein is localized mainly in the nucleolus (Spector et al. 1984, Cordell et al. 1999), NPM1.2 is only found in the nucleoplasm (Dalenc et al. 2002). NPM1 contributes to build-up the nucleolar compartment of which it is one of the most abundant protein components among the approximately 700 identified by proteomics (Ahmed et al. 2009).

Despite its nucleolar localization, experiments on NPM1 migration provided conclusive evidence that the protein shuttles constantly back and forth

between nucleus and cytoplasm (Borer et al. 1989). However, the flux of nucleophosmin at the nucleolus–nucleoplasm interface is unbalanced towards the nucleolus. This is the consequence of the very efficient nucleolar-localization signal (NoLS) that NPM1 contains at its C-terminus.

Nuclear-cytoplasmic shuttling of NPM1 is critical for most of its functions, including regulation of ribosome biogenesis and control of centrosome duplication (Grisendi et al. 2006). NPM1 shuttles across cytoplasm and nucleoplasm as well as between nucleoplasm and nucleolus. Distinct NPM1 functional domains regulate its different activities as well as its shuttling properties.

1.3 Nucleophosmin functional domains

The NPM1 polypeptide chain has a modular structure containing distinct sequence motifs. Various functional domains that reside in mainly independent, but slightly overlapping segments, have been identified within the protein (Fig. 1.1).

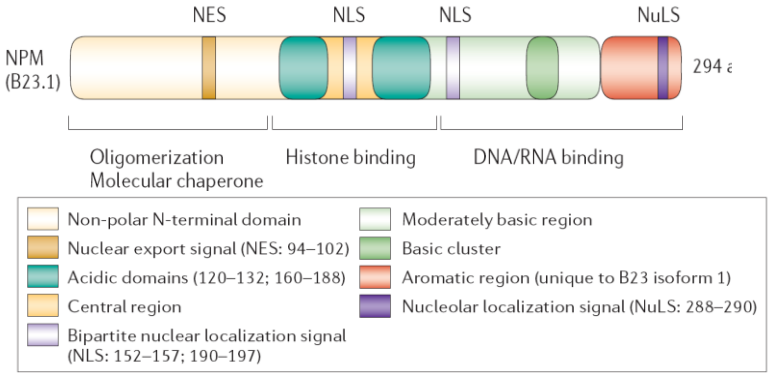

Fig. 1.1 Structural and functional organization of NPM1, the different portions of the protein and their respective function are depicted in a colour coded representation. (adapted from Grisendi et al., 2006)

Starting from the N-terminus, the protein contains a nucleoplasmin-like domain that is involved in protein oligomerization and chaperone activities, followed by two acidic stretches that are important for binding to histones. The central portion between the two acidic stretches is required for ribonuclease activity, together with the C-terminal domain, which contains basic stretches involved in nucleic-acid binding (Hingonari et al. 2000). The basic clusters are followed by an aromatic stretch, which contains two tryptophan residues (288 and 290) that are required for nucleolar localization of the protein (Nishimura et al.2002). In addition, NPM1 includes a nuclear-localization signal (NLS), and a nuclear-export signal (NES) (Wang et al.2005). As stated above, the N-terminal domain of NPM1 share 50% homology with the *Xenopus laevis* protein nucleoplasmin (Chan et al. 1989; Borer et al. 1989). Nucleoplasmins are nuclear chaperone proteins that function as histone assembly and chromatin decondensation factors. Their main function is to bind to core histones and transfer DNA to them in a reaction that does not require ATP. All the members of this protein family share a similar core region in the N-terminal half (nucleoplasmin domain), and one or more acidic domains rich in aspartic and glutamic acid residues. Compared with other members of the family, NPM1 possesses additional domains and functions in the C-terminal region, such as ribonuclease enzymatic activity and a nucleic-acid binding domain.

1.4 Nucleophosmin: Physiological role

By shuttling between cellular compartments, NPM1 takes part in various cellular processes. These include the transport of pre-ribosomal particles and ribosome biogenesis (Olson et al. 1986), the response to stress stimuli such as UV irradiation and hypoxia (Li et al. 2004; Wu et al. 2002), the maintenance of genomic stability through the control of cellular ploidy (Naoe et al. 2006) and the participation in DNA-repair processes (Wu et al. 2002), and the regulation of DNA transcription through modulation of chromatin condensation and decondensation events (Wang et al. 1994; Dumbar et al. 1989).

Finally, NPM1 is involved in regulating the activity and stability of crucial tumour suppressors such as p53 and p14ARF (Bertwistle et al. 2004; Kurki et al. 2004; Colombo et al. 2006; Bolli et al. 2009) and oncogene products such as c-MYC (Boon et al. 2001; Yung 2004; Zeller et al. 2001).

One main feature of NPM1 is its ability to function as a molecular chaperone for both proteins and nucleic acids (Szebeni and Olson 1999; Okuwaki et al. 2001). *In vitro* experiments on various protein substrates have shown that NPM1 is active in preventing the aggregation of proteins in the congested cellular environment (Szebeni and Olson 1999), and that it functions as a histone chaperone that is capable of histone assembly, nucleosome assembly and increasing acetylation-dependent transcription (Okuwaki et al. 2001).

Through its molecular domains, NPM1 is able to bind to many partners in distinct cellular compartments, including nucleolar factors (for example, nucleolin, fibrillarin and snoRNPs), transcription factors (for example, interferon regulatory factor 1 (IRF1), YY1 and nuclear factor kB (NFkB)),

histones (for example, H3, H4 and H2B), proteins involved in cell proliferation (for example, DNA polymerase), mitosis (for example NUMA) and proteins involved in response to oncogenic stress (for example, p14ARF and p53) (all reviewed in Grisendi et al. 2006). In addition NPM1 has been reported to bind both the oncogenic protein c-MYC (Boon et al. 2001) and its E3-ubiquitin ligase Fbwγ (Bonetti et al. 2008) in the nucleus while, in the cytoplasm, interactions with caspases 6 and 8 (Leong et al. 2010) and with proteins belonging to the Fanconi anemia complex have been reported (Li et al. 2007).

Eventually, NPM1 is able to associate with the second messenger phosphatidylinositol-(3,4,5)-trisphosphate (PIP3) in the nucleus in response to anti-apoptotic factors (Ahn et al. 2005).

The diversity of the cellular activities in which NPM1 is involved in makes it both a potential oncogene and a potential tumour suppressor, depending on the expression levels and the cellular context.

1.5 Structural features of Nucleophosmin

Under native conditions, NPM1 exists as an oligomer (Herrera et al. 1996): the structure of the X*enopus laevis* nucleoplasmin family member NO38 (Spector et al. 1984), a nucleolar histone chaperone that is highly homologous to human NPM1, appears to be able to form pentamers and decamers (Naboodiri et al. 2004).

More recently the crystal structure of human NPM1 N-terminal domain was determined and confirmed that a crown-shaped pentamer is formed by this core domain, with two pentamers loosely associating to form a head-to-head

dimer of pentamers that account for the fairly large molecular weight of appromatively 300 KDa (Lee et al. 2007) (Fig. 1.2).

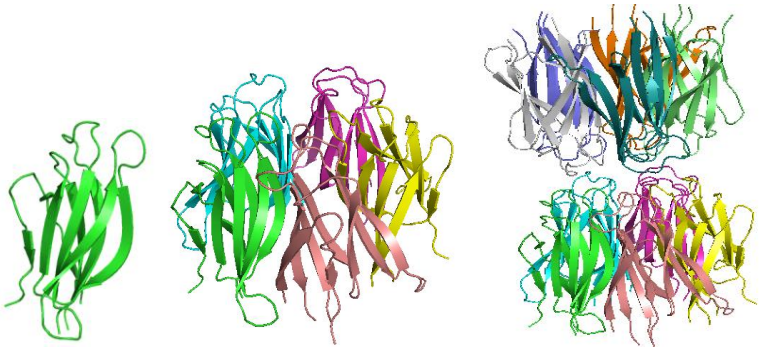

Fig. 1.2 Organization of Nucleophosmin N-terminal domain.

The ability of NPM1 N-terminal domain to form oligomers is involved in its translocation to the nucleolus. Indeed, artificial NPM1 mutants that contain the nucleolar-binding domain but are unable to form oligomers cannot enter the nucleolus, suggesting that the NoLS load of the particle has to reach a threshold to enable nucleolar localization (Bolli et al. 2007).

The structure of NPM1 C-terminal domain was solved by NMR spectroscopy (Grummitt et al. 2008) and consists of a well-defined 3-helix right-handed bundle (Fig. 1.3).

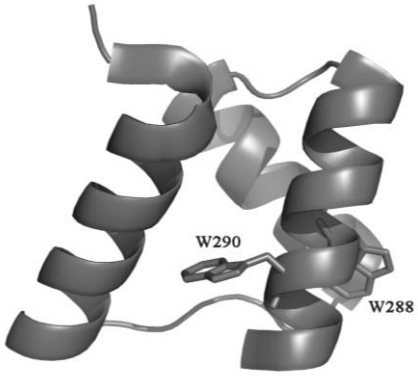

Fig. 1.3 NMR 3D structure of NMP1-Cter (pdb code, 2VXD). The two tryptophans at positions 288 and 290 are highlighted in sticks.

The central region of NPM1, which is endowed with many protein's function including histone binding and ribonuclease activity, is predicted to be natively unstructured by the vast majority of predictive algorithms (Hingonari et al. 2000).

Finally and importantly, several post-translational modifications have been reported, including phosphorylation, ubiquitination and SUMOylation, and may contribute to the final shape of the protein, regulate the cellular traffic of NPM1 and its functions (Colombo et al., 2011).

1.6 NPM1 is altered in human cancer

NPM1 has been directly implicated in human tumorigenesis. The NPM1 protein is overexpressed in a variety of tumours, and it has been proposed as a marker for gastric (Tanaka et al. 1992), colon (Nozawa et al. 1996), ovarian (Shields et al. 1997) and prostate carcinomas (Subong et al. 1999). In some cases, the expression levels of NPM1 have been correlated with the stage of tumour progression. For instance, overexpression of *NPM1* mRNA is independently associated with the recurrence of bladder carcinoma and progression to a more advanced stage of disease (Tsui et al. 2004). On the other hand, *NPM1* is one of the most frequent targets of genetic alterations in haematopoietic tumours. Two well-characterized genomic events that involve the *NPM1* gene are chromosomal translocations in both lymphoid and myeloid disorders (Raimondi et al. 1989, Redner et al. 1996), and mutations in Acute Myeloid Leukemia (AML) (Falini et al. 2005).

NPM1 alterations lead to the formation of either oncogenic fusion proteins or mutant NPM1 products, and to the concomitant loss of one functional allele of the gene (Berger et al. 2006). The role of the NPM1 moiety in the various chimeric products (NPM1–ALK (anaplastic lymphoma kinase), NPM1–RAR (retinoic acid receptor) and NPM1–MLF1 (myelodysplasia/myeloid leukaemia factor1)) is not fully characterized, and it is conceivable that NPM1 might serve as a homo- or heterodimerization interface (Bischof et al. 1997).

In 2005, Falini and coworkers identified heterozygous NPM1 mutations as the most common genetic lesion in adult acute myeloid leukaemia (AML) (about 30% of cases) (Falini et al. 2005). They

subsequently provided evidence that AML with mutated NPM1 exhibits distinctive biological and clinical features (Falini et al. 2007) that supported its inclusion as a provisional entity in the new WHO classification (Falini et al. 2009).

Approximately 40 NPM1 mutations have been identified to date (Falini et al. 2009), and despite their heterogeneity they all lead to similar abnormalities in the mutated protein. Mutations all consist of short base sequence duplications or insertions at the terminal exon of the gene (Fig. 1.4). These lead to a shift in the reading frame, and a consequent alteration in the protein sequence which is longer by four residues and different in the last seven of the reference one. Notably both W288 and W290 are replaced in the most common mutants, while W288 is retained in less frequent mutations (Fig. 1.4).

Fig. 1.4 Representative NPM1 mutations and their consequences on protein sequence**.**

The salient feature of all mutated genes is that they generate a variant protein that is aberrantly localized in leukaemic blast cytoplasm (Nakagawa et al. 2005) (Fig. 1.5). This varied distribution of NPM1 mutants in AML cells is due to the concerted action of tryptophans loss, with the alteration of the so-called nucleolar localization signal, and the generation of a new nuclear export signal (NES) motif at the NPM1 protein C-terminus (Falini et al. 2006). Under normal conditions, the two tryptophans are critical for keeping the C-terminus globular domain of wild-type NPM1 folded (Grummitt et al. 2008, Scaloni et al., 2009-2010), and this in turn appears essential for NPM1 binding to the nucleolus. In leukaemic cells, mutation-induced tryptophan changes unfold the domain and consequently impair NPM1 targeting to the nucleolus (Grummitt et al. 2008).

The expression of NPM1 fused to GFP in mammalian cells allowed the direct observation of its sub-cellular localization (Dundr et al. 2000). This approach proved to be extremely powerful to evaluate the effect of leukaemic mutations on the cellular distribution of the protein. In particular, two designed point-mutants, W288A and W290A, were shown to lead to loss of nucleolar localization and to the accumulation of the protein in the nucleoplasm (Falini et al. 2006; Nishimura et al. 2002). Indeed, it was later shown that the NPM1 C-terminal domain with W288 and W290 mutated to alanine lacked any distinct tertiary structure and did not undergo a cooperative unfolding transition when heated (Grummitt et al. 2008). This suggested that mutation of these residues prevents nucleolar localization simply by preventing the correct folding of the C-terminal domain of NPM1. In order to test this hypothesis other mutations in the C-terminal domain that

would be expected to disrupt its structure whilst retaining the tryptophans were produced (Grummitt et al. 2008). The sub-cellular localization of these mutants transfected into NIH-3T3 cells was indeed nucleoplasmic and not nucleolar, suggesting the fundamental role of correctly folded C-terminal domain for nucleolar localization (Falini et al. 2006).

Cytoplasmic accumulation of NPM1 leukaemic mutants also depends upon the NES motifs. Two of them are physiological and are located at the N-terminus (corresponding to residues 42–49 and 94–102) (Wang et al.2005). The third motif is introduced by the mutational event at the NPM1 C-terminus (residues 287–296) (Falini et al. 2005). However, it is not yet clear how much each of the three NES motifs contributes to mutant protein nuclear export (Bolli et al. 2009) (Fig. 1.4).

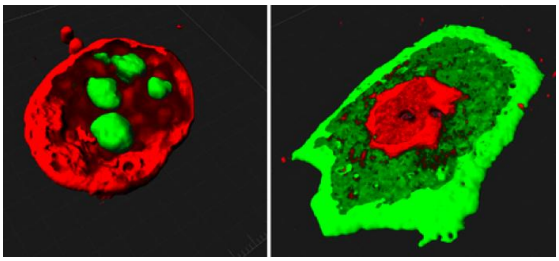

Fig. 1.5 Ectopic aberrant localization of mutated NPM1 in the cytoplasm. Left panel: GFP-tagged wild-type NPM1 localizes in nucleoli inside the nucleus. Right panel: GFP-tagged mutated NPM1 aberrantly localizes outside the nucleus in the cytoplasm. Red stain highlights the nuclear membrane.

One important feature to be highlighted is that, although *NPM1* mutations are always heterozygous, the mutated protein has a dominant negative effect on wild-type localization. In fact, mutated NPM1 is still capable to interact with

the wild-type protein, through its unchanged N-terminal domain, and to form mixed oligomers that are, in great part, localized in the cytoplasm.

1.7 Nucleophosmin interacts with nucleic acids

The NPM1 C-terminal domain is responsible for the nucleic acid binding activity. NPM1 was shown to bind both DNA and RNA oligonucleotides with a preference for single-stranded structures over double-stranded DNA B structures, in a sequence-unrelated manner (Wang et al.1994; Dumbar et al. 1989). Accordingly, a role for NPM1 as a single-stranded binding protein (ssb) was proposed (Dumbar et al. 1989), a property that may be linked to the export of ribosome subunits from the nucleus (Falini et al. 2009). NPM1 was also reported to have endoribonuclease activity on ribosomal RNA (rRNA) (Savkur and Olson 1998; Herrera et al. 1995). Recently, the protein was also identified as a cofactor in the transcriptional activation of the mitochondrial superoxide dismutase 2 (*SOD2*) gene; in particular it was shown that full-length NPM1 binds a G-rich region at the *SOD2* promoter (Xu et al. 2007), shedding new light on NPM1 possible role in transcriptional regulation through its interaction with G-rich promoter regions.

G-rich sequences have the potential to form non-canonical DNA secondary structures, called G-Quadruplexes, under physiological conditions (Gellert et al. 1962). Different studies are increasing our attention on non-canonical DNA conformations, describing their structures, interacting partners and potential biological roles (Phan et al. 2006, Sen and Gilbert 1988). As well as on the non-canonical secondary structures (with particular attention to G-quadruplexes), different studies in the last few decades underlined the

increasingly important role of DNA G-rich binding proteins focusing attention on how they interact, and how this interaction may regulate different cellular processes (Fry 2007). In fact experimental evidence corroborates the reasonable assumption that regulated interactions of G-rich DNA sequences with different proteins may alternatively cause formation, stabilization, destabilization, unwinding or selective nucleolytic digestion of G-quadruplex sequences, suggesting key roles for these interactions in the cell (Fry 2007).

1.8 G-Quadruplex

Nowadays, it is becoming increasingly evident that non-Watson-Crick interactions between bases and non-canonical DNA secondary structures have biological roles, with particular reference to Guanine-rich nucleic acid sequences that are able to form higher order secondary structures: the so called guanine-quadruplexes (G-Quadruplexes). Observations starting from 1960s established that a synthetic guanosine 5'-monophosphate (GMP) has the ability to form helical aggregates under physiological-like *in-vitro* conditions of pH, temperature and salt concentrations (Gellert et al. 1962). These initially inexplicable finding was explained in the late 1980s when biochemical experiments demonstrated that oligonucleotides containing runs of three or four adjacent guanines (G-tracts), spontaneously can fold in a valet of four-stranded formations called tetraplex or quadruplex (Sen and Gilbert 1988, Williamson et al. 1989).

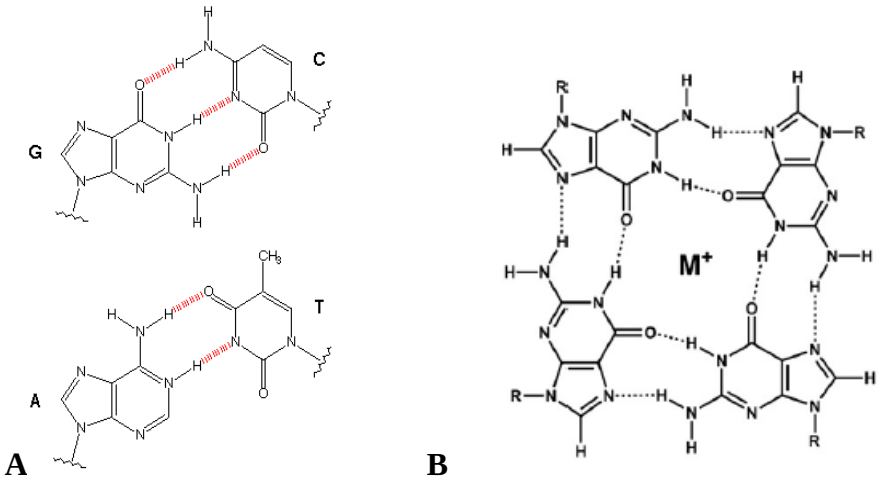

Fig.1.6 (A) Watson–Crick base-pairing. (B) Four guanines can hydrogen bond in a square arrangement to form a guanidine tetrad. There are two hydrogen bonds on each side of the square.

These non-canonical structures are composed of four or more Guanine-tracts (G-tracts), which are able to form stacking guanine tetrads, where the tetrads stack on top of each other. The guanines within the tetrad interact via Hoogsteen hydrogen bonds between the N1, N7, O6 and N2 guanine atoms and are stabilized by monovalent cations such as potassium or sodium (Qin et al. 2008) (Fig.1.6B). These guanine tetrads are the basic building block of G-Quadruplexes, and contribute to their secondary structure and stability by guanidine tetrads' large π-surfaces, that tend to stack on each other due to π-π stacking interactions, with stabilizing monovalent cation intercalated between tetrads (Gilbert et al. 1999) (Fig.1.7).

By definition G-Quadruplex is a DNA secondary structure that consist of multiple vertically stacked guanidine tetrads stabilized by alkali ion, most commonly K^+ or Na^+ (Sen et al. 1990). K^+ and Na^+ effectively bind to and

stabilize many G-Quadruplex structures and in particular K^+ shows a higher stabilizing activity, with respect to Na^+, due to the better coordination of K^+ with the eight carbonyl oxygen atoms present in the adjacent stacked tetrads (Sen et al. 1990). Moreover K^+ and Na^+ are the prevalent alkali ions in the cell environment underlining that physiological conditions are favorable for G-Quadruplex formation. A variety of G-Quadruplex structures have been determined using X-ray crystal diffraction and NMR analysis (Kang et al. 19992, Smith and Feigon 1992, Wang and Patel 1993, Kettani et al. 1995). These studies led to classify quadruplexes into three majors groups: unimolecular, bimolecular and four-molecular Quadruplexes. Furthermore, depending on direction of the strands, may assume different topologies described as parallel, antiparallel or mixed.

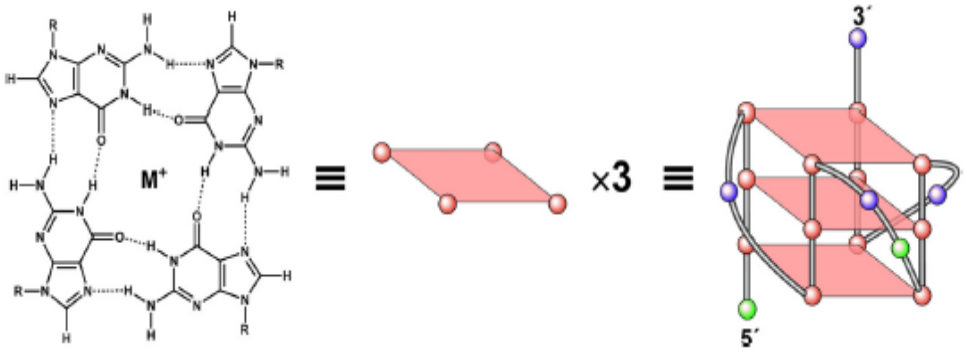

Fig. 1.7 Structure of a G-tetrad and an example of the folding pattern of intramolecular G-quadruplex

1.9 G-Quadruplexes: biological role

Data gathered in last decades are focusing attention on G-Quadruplexes, and in particular mounting evidence indicates that these secondary structures exist in vivo and play a number of biological functions.

For many years there was minimal evidence pointing towards biological role of G-Quadruplex structures, until it was discovered that the ends of human chromosomes, the telomeres, are composed of tandem repeats of G-rich DNA sequence: d(TTAGGG) (Moyzis et al. 1988). Subsequent work demonstrated the ability of these sequences to form intramolecular G-Quadrupexes, and how these structures can regulate Telomerase activity (Wang and Patel 1993, Zahler et al. 1991). In fact formation of quadruplexes at telomere's ends directly inhibits Telomerase activity, preventing telomeres elongation and impacting on cellular senescence and cell cycle regulation. Telomerase activity is directly implicated in more then 85% of cancer processes, where it is over-expressed contributing to cancer cells immortalization. These data describe a direct regulating role of a G-Quadruplex structure in pathological process, shedding a new light and suggesting a possible functional role for this non-canonical secondary structure.

Starting from studies of G-Rich 3'overhanging in human telomere and Telomerase relationship, both in pathological and physiological conditions, extensive research effort has been devoted to investigate the relevance of potential function of DNA G-Quadruplex in biology.

Many computational studies have been carried out to identify G-rich sequences with potential to form G-Quadruplex secondary structures (Putative Quadruplex Sequence, PQS) in the human genome and other

genomes. These studies showed that the human genome contains as many as 376000 PQS (Putative Quadruplex Sequence; Huppert and Balasubramanian 2005). Although the occurrence of PQS in genomes is a third less than expected by chance (Huppert and Balasubramanian 2007), their location seems to be non-random (Zhao et al. 2007). The highest occurrence of PQS is in non-coding repetitive DNA regions such as telomeres, hyper-recombining sites and, interestingly, in promoter regions of eukaryotic genes with over 40% of human gene promoters containing at least one PQS, emphasizing G-Quadruplex potential regulatory role (Huppert and Balasubramanian 2007).

Different studies identified PQS that are able to form G-Quadruplex *in vitro*, in promoter regions of different human genes, with particular attention to important proto-oncogenes like c-Myc (Suddiqui et al. 2002), VEGF (Sun et al. 2005), HIF-1α (De Armound et al. 2005), Ret (Guo et al. 2007), KRAS (Cogoi et al. 2006), Bcl-2 (Dexheimer et al. 2006), c-Kit (Rankin et al. 2007), PDGF-A (Qin et al. 2007) and c-Myb (Lee et al. 2008). Generally the proximal regions of these promoters are GC-rich regions and usually are hypersensitive to nuclease and may form an altered structure with a single stranded character which is often a feature of transcriptionally active genes (Sun et al. 2005).

This association of G-Quadruplex with gene promoters was consistent with the hypothesis of linking G-Quadruplexes formation with transcriptional control of genes (Simonson et al. 1998). Subsequently, different studies showed that mutations that destabilize G-Quadruplex structure in promoter regions led to an increased in transcription activity (Siddiqui-Jain 2002). This has led to suggestions that G-Quadruplexes play a key-role in the regulation

of gene expression through formation or resolution of specific Quadruplex structures, thus describing these DNA non-canonical secondary structures as a *cis*-acting regulatory elements.

One of the better characterized G-Quadruplex forming sequence is the one in *c-Myc* oncogene promoter region. The *c-Myc* protooncogene encodes a multifunctional transcription factor that plays a critical role in different cellular processes such as regulation of cell cycle progression, cell growth, differentiation, transformation, angiogenesis, and apoptosis (Oster et al. 2002). c-MYC is able to activate a number of genes by forming complexes with other transcription factors such as MAX or MAD that interact with specific DNA sequences, such as the E-box sequence (Grandori et al. 2000). Overexpression of c-MYC is associated with a significant number of human malignancies, including breast, colon, cervix, and small-cell lung cancers, osteosarcomas, glioblastomas, and myeloid leukemias (Facchini et al. 1998). *c-Myc* transcription is usually subject to tight transcriptional regulation and it is under the complex control of multiple promoters. The promoter region of c-*Myc* contains a number of *cis*-elements that have been shown to assume either a single-stranded or a non–B-DNA conformation; in particular there are seven nuclease hypersensitive elements (NHEs) in the c-*Myc* promoter, and one of these, the NHE III_1 region, has been shown to have the ability to form non–B-DNA structures. This region is located from –1 to –115 base pairs upstream of the P1 promoter and has been shown to control up to 90% of the total c-*MYC* transcription (Berberich et al. 1995). The NHE III_1 in *c-Myc* contains a 27-base pair G-rich sequence (c-Myc-27-mer) at the non-coding strand that is capable of engaging in a slow equilibrium between B-

form duplex DNA, single-stranded DNA, and G-Quadruplex (Gonzàles et al. 2010). Quadruplex formation in NHE III_1 induce a complete silencing of *c-Myc* transcription (Siddiqui-Jain et al. 2002), suggesting the importance of this regulatory mechanism based on equilibrium between formation and resolution of this DNA non-canonical secondary structures. Biological significance of the intramolecular G-Quadruplex structure formed in *c-Myc* NHE III_1 has been evaluated also in two Burkitt's lymphoma cell lines, confirming G-Quadruplex transcriptional inhibitory activity (Siddiqui-Jain et al. 2002).

Furthermore, different proteins contribute to transcriptional regulation favoring G-Quadruplex formation or resolution. Different studies identified many proteins that could functionally interact with non-canonical DNA conformation shedding light on the importance of these structures with particular attention to G-Quadruplex (Fry 2007).

2 Aim of the thesis

The Nucleophosmin C-terminal domain is the portion of the protein that is altered by *NPM1* gene mutations in Acute Myeloid Leukemia. This domain, which consists of a three-helix bundle, as recently revealed by NMR spectroscopy (Grummitt et al. 2008), is largely destabilized or totally unfolded in the leukogenic variant protein, due to a change in the gene reading frame that results in a protein that is longer by four residues and with a different sequence in the last seven. Two main variations result from the mutational event: i) the loss of one or both of two critical tryptophan residues in the hydrophobic core of the three helix bundle, which accounts for the dramatic loss of stability of the domain (Scaloni et al. 2010) and ii) the appearance of a new Nuclear Export Signal at the C-terminus of the protein. Both variations concur to the aberrant translocation of the protein in the cytosol, which is the hallmark of this kind of leukemia.

The NPM1 C-terminal domain has been known for a long time to be responsible for the nucleic acid binding properties of the protein. It was shown that NPM1 binds both RNA and DNA, with a preference for single stranded nucleic acids and with no specific sequence requirements (Hingorani et al., 2000).

Because leukemic mutant NPM1 is stably localized in the cytoplasm, we decided to further investigate the DNA binding properties of the protein, assuming that some important function related to nucleic acid binding in the nucleus might be impaired in these mutants.

This task, which was the initial aim of my Ph.D. work, was tackled through the combination of different spectroscopic techniques and probing NPM1 C-terminal constructs of different lengths and a variety of site-directed mutants.
Importantly these studies revealed that, although capable of binding any DNA oligo with low affinity, NPM1 C-terminal domain displays higher binding affinity for oligos capable of assembling as G-quadruplexes. Since the first rather serendipitous observation of this feature, my experimental work was directed towards the full characterization of NPM1 as a novel G-quadruplex binding protein.
Having established the preference of NPM1 for G-quadruplex DNA regions, the next step of my experimental work was to structurally characterize the complex formed by the NPM1 C-terminal domain and a representative G-Quadruplex from the oncogene *c-MYC* promoter. This aim was particularly intriguing to us, both because of its possible implications in understanding NPM1 biology, but also because structural information about how proteins recognize G-quadruplex DNA is very limited. This task was achieved with the use of several NMR techniques and working in collaboration with Prof. Ivano Bertini and Prof. Lucia Banci, at CERM-University of Florence, where I spent the last year of my Ph.D. period.

In summary this thesis reports a detailed characterization of NPM1 DNA binding properties and a structural investigation of a representative NPM1-G-quadruplex complex.

I believe that these studies may provide a useful conceptual framework to better understand some of the biological functions played by

this highly pleiotropic protein and their loss in the aberrantly cytoplasmic localized variant protein.

3 Results

3.1 NPM1 interaction with DNA

It is well known that NPM1 C-terminal domain is able to bind both DNA and RNA, with a preference for single stranded oligonucleotides over duplexes and with no specific sequence requirements (Hingorani et al., 2000). This behavior suggested that NPM1 nucleic acid binding activity may serve for NPM1-mediated ribosome transfer to the cytosol.

Recently, however, it was demonstrated that full-length NPM1 can interact with a specific G-rich sequence present in *SOD2* promoter region and it was suggested that, through this recognition event, the protein is directly involved in the transcriptional activation machinery, functioning as a positive regulator of the *SOD2* gene, by mediating the interaction between enhancer and promoter regions (Xu et al. 2007). This promoter region is predicted to fold in a hairpin with a five paired-bases stem and a stretch of ten consecutive Guanines that form a loop (hereby named G-10 loop; Fig. 3.1B). By interfering with base-pairing in the wild type sequence or eliminating the loop, NPM1 is unable to bind the promoter region, preventing gene transcription, and thus indicating a direct connection between promoter region structure, NPM1 binding and gene transcription (Xu et al. 2007).

Our first aim was to confirm that the DNA binding activity of the G-10 loop could be mapped to the C-terminal domain of the protein, as previously reported in literature for unstructured oligos (Hingorani et al., 2000), and to better define the boundaries of the DNA binding domain. To this purpose, a biotynilated version of the G10-loop was immobilized on a streptavidin chip and used as bait in Surface Plasmon Resonance analysis. Two different

constructs of the NPM1 C-terminal domain were used as the analytes: NPM1-C70 comprising the last 70 C-terminal residues, and NPM1-C53 comprising the last 53 residues, (Fig. 3.1 A). The first construct was chosen because previous binding data on truncated proteins mapped the NPM1 nucleic acid binding activity to this segment (Wang et al. 1994; Hingorani et al. 2000), whereas NPM1-C53 was chosen because it is the minimal folding unit of the C-terminal domain (Scaloni et al. 2009), which adopts a stable three helix bundle structure (Grummitt et al. 2008).

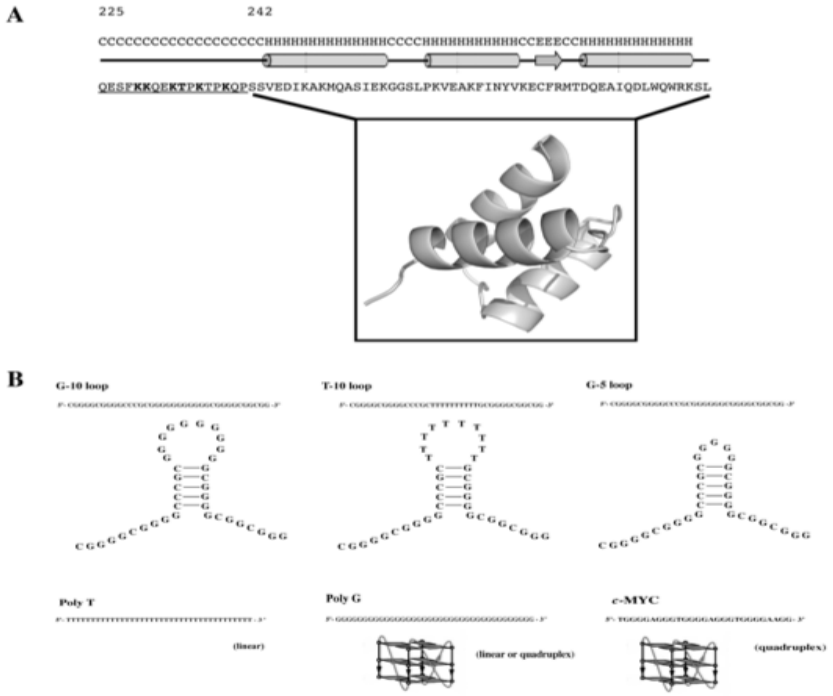

Fig. 3.1 Protein constructs and DNA oligos used in this study. A) Sequence of the C-terminal domain of NPM1. The region differentiating NPM1-C70 from NPM1-C53 (whose 3D structure is shown in the inset) is underlined. Mutated lysines are shown in bold. The PSIPRED prediction is also shown (C=coil, H=α-helix, E=β-strand).B) Oligos used in this study are shown both in sequence and in putative

structure.

When testing NPM1-C70 as the analyte we obtained a K_D = 7.2 µM (Fig. 3.2 A, Table 3.1) while with NPM1-C53 we obtained a K_D = 169 µM (Fig. 3.2 B, Table 3.1). Interestingly, both the association and dissociation rate constants could be determined when NPM1-C70 was the analyte (Fig. 3.2 A), while they were too rapid to be determined with NPM1-C53 (Fig. 3.2 B).

This result suggests that the nucleic acid binding site is altered in the shorter protein construct and that, in this case, binding is mainly dictated by electrostatic interactions of the positively charged protein domain (Fig. 3.1) with the negatively charged G10-loop.

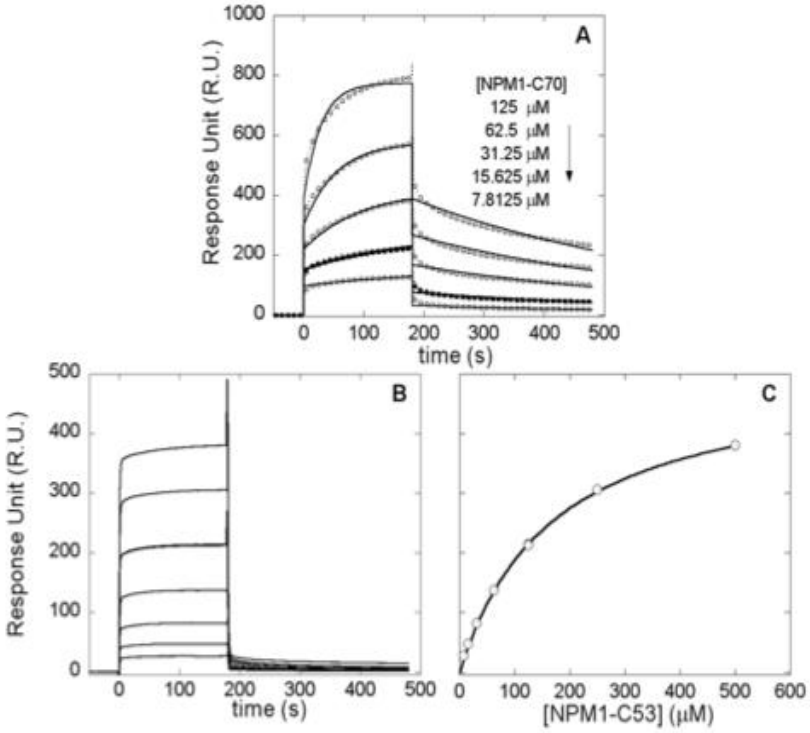

Fig. 3.2 SPR sensorgrams of the interaction between G10-loop and NPM1 constructs. The oligo, biotynilated at its 5' end, was immobilized on a SA (streptavidin) sensor chip. A) experimental curves (dashed lines with circles) represent different concentrations of NPM1-C70 used as the analyte and were fitted according to a single exponential binding model with 1:1 stoichiometry (continuous lines). B) experimental curves represent different concentrations of NPM1-C53. Both on and off rates were too rapid to be resolved, C) Scatchard plot used to determine the dissociation constant with NPM1-C53.

Next, we wanted to determine the structural properties of the DNA hairpin necessary for binding NPM1. To this end we first tested a so-called T10-loop, where the ten guanines at the loop of the G10-hairpin are replaced by ten thymines, while all the other bases are conserved (Fig. 3.1 B). With NPM1-C70 we obtained a K_D = 307 µM that increased to K_D = 1.17 mM with NPM1-C53 (Table 3.1). This experiment indicates that a sequence that forms a hairpin structure resembling that of the native G10-loop but with thymines instead of guanines at the loop is poorly recognized, with a 42-fold lower affinity. NPM1-C53 is indeed poorly competent for binding.

To explore the dimensional requirements of the hairpin for high affinity recognition, we next immobilized a G5-loop, which maintains the same hairpin arrangement of the G10-loop but with only five guanines in the loop (Fig. 3.1 B). With this oligonucleotide, we obtained a K_D = 40 µM with NPM1-C70 and of 224 µM with NPM1-C53. These experiments indicate that both the presence of guanines at the hairpin loop and their number contribute, to the global affinity, suggesting that the 3D-structure of the G10-loop plays an important role.

To further assess the DNA binding properties of NPM1, we next tested an oligo made only of T-bases (38-mer), hereby named poly-T (Fig. 3.1 B), and

measured a K_D = 120 µM with NPM1-C70 and a K_D = 520 µM with NPM1-C53 (Table 3.1). This suggests that the increased flexibility of the polyT linear oligo with respect to the T10-loop can partly compensate for the absence of guanines.

A further experiment was designed to establish the role played by guanines in a DNA sequence that does not necessarily form hairpin structures. We reasoned that if the protein preferentially recognizes with high affinity a hairpin loop made of guanines, it might also be able to induce such loop formation in a poly-G oligo and recognize it with good affinity (Figure 3.1 B). Thus with poly-G we expected to find a K_D for NPM1-C70 higher than that found with the G10-loop but lower than that found with the poly-T oligo.

Interestingly enough, with NPM1-C70 we obtained a K_D = 5.8 µM, comparable to that of the physiological substrate G10-loop; with NPM1-C53 we obtained instead a K_D = 31 µM, the lowest measured so far with this protein construct (Table 3.1). These results suggest that the recognition of a poly-G oligo is far more specific than we might have anticipated. This may be rationalized by hypothesizing that this oligo is not linear but has the potential to form a structure that resembles that of the G10-loop. Indeed, by investigating the literature, we realized that a polyG oligo, if long enough, such as our 38-mer, has the potential to form 3D-structures known as G-quadruplexes (Huppert and Balasubramanian, 2005; Neidle, 2009).

Ligand	**Analyte**	**K_D** (μM)	**k_{on} ($M^{-1}s^{-1}$)**	**k_{off} (s^{-1})**
G10-loop	NPM1-C70	7.2 ± 0.2	2.8 ± 0.1 x 10^2	1.98 ± 0.03 x 10^{-3}
	NPM1-C53	169 ± 5	-	-
T10-loop	NPM1-C70	307 ± 7	-	-
	NPM1-C53	1170 ± 40	-	-
G5-loop	NPM1-C70	40 ± 3	-	-
	NPM1-C53	224 ± 6	-	-
Poly T	NPM1-C70	120 ± 26	-	-
	NPM1-C53	520 ± 30	-	-
Poly G	NPM1-C70	5.8 ± 0.8	-	-
	NPM1-C53	31 ± 4	-	-
cMyc	NPM1-C70	1.9 ± 0.1	7.4 ± 0.3 x 10^3	1.37 ± 0.03 x 10^{-2}
	NPM1-C53	82 ± 15	-	-

Tab. 3.1 Dissociation constants and kinetic parameters as determined by SPR analysis

3.2 NPM1 C-terminal domain binds a sequence known to form a G-quadruplex

G-quadruplexes are formed by sequences displaying at least four stretches of at least three guanines, with no sequence requirements for the intervening loops that are usually one to seven nucleotides long ($G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$) (Huppert, 2008). The guanines interact with each other in an arrangement different from the classical B DNA pairing, forming planar tetraeds stabilized by the so-called Hoogsteen type H-bonds (Neidle, 2009; Huppert, 2008). These structures are greatly stabilized by Na^+ or K^+ ions that intercalate in the rings formed by the four guanines in the tetraed. Repetitive G-rich sequence stretches are highly represented in the human genome and are clustered at gene promoters, suggesting their functional importance.

We decided to investigate whether one well-characterized G-quadruplex might interact with our NPM1 constructs. We focussed our attention on a sequence contained in the NHE_{III} (nuclease hypersensitive element III) region of the *c-Myc* promoter that is known to regulate up to 90% of total c-MYC expression. This is a well-characterized example of a parallel G-quadruplex forming region, both *in vitro* and *in vivo* (Gonzales et al., 2010). An oligo representing the G-quadruplex forming region of this promoter (hereby c-MYC oligo) was therefore immobilized on a chip for SPR analysis (Fig. 3.1 B).

Interestingly, we found that NPM1-C70 binds the c-MYC quadruplex with a K_D = 1.9 μM (Fig. 3.3 and Table 3.1), confirming that NPM1-C70 has high affinity for a sequence that adopts a G-quadruplex structure. Similarly to the

G10-loop, and contrary to the other oligos, both the association and dissociation rates could be calculated, suggesting that c-Myc G-quadruplex recognition is specific. Moreover, the experiment performed using NPM1-C53 as the analyte led to a K_D = 82 μM, confirming reduced affinity with this shorter domain (Fig. 3.3 and Table 3.1).

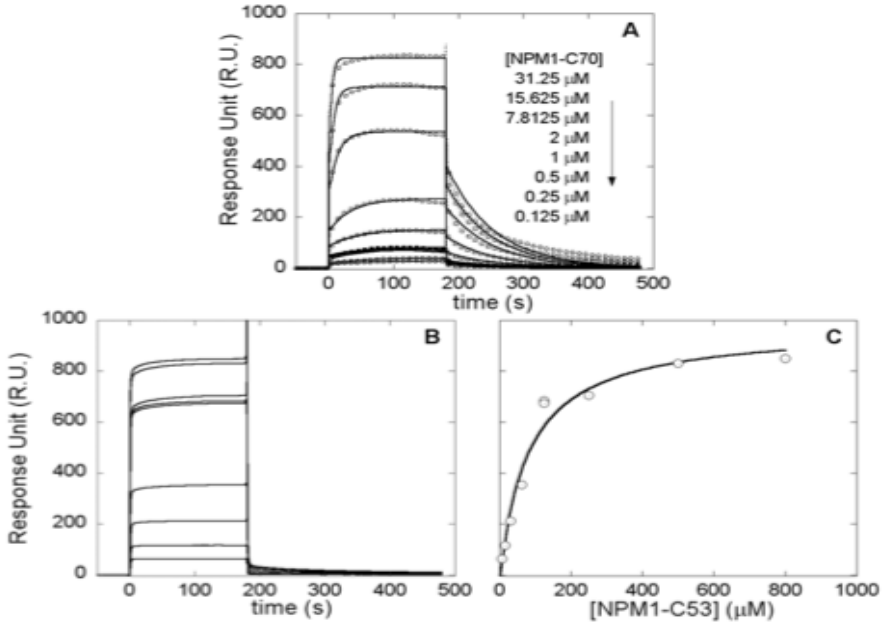

Fig. 3.3 SPR sensorgrams of the interaction between c-MYC and NPM1 constructs. A) experimental curves (dashed lines with circles) represent different concentrations of NPM1-C70 used as the analyte and were fitted according to a single exponential binding model with 1:1 stoichiometry (continuous lines). B) experimental curves represent different concentrations of NPM1-C53. C) Scatchard plot referring to curves in panel B for dissociation constant determination.

3.3 The G10-loop forms a G-quadruplex structure in Vitro

The results reported so far suggest that the NPM1 C-terminal domain recognizes with particularly high affinity sequences known to form G-quadruplex three-dimensional arrangements. Under this light, it is interesting to note that the G10-loop sequence, which is predicted to form a hairpin according to conventional Watson-Crick pairing (Fig. 3.1B), also matches the above mentioned folding rule for G-quadruplexes and, accordingly, is predicted to form a G-quadruplex by the Quadparser algorithm (Huppert and Balasubramanian 2009). Therefore, when annealing, this oligo has the potential to form at least two alternative structures. To infer which of these two structures is the most likely to be populated, we first collected the CD spectra of G10-loop and c-MYC, for comparison. It is well known that circular dichroism is diagnostic of G-quadruplex formation (Huppert 2008). In particular, by comparing the spectra of the c-MYC oligo annealed in the absence or presence of 100mM NaCl or 100mM KCl, respectively, we observe a red-shift and increase in intensity of the peak at around 260 nm and the formation of a through at 240nm(Fig. 3.4*A*). Importantly the same features are observed in the case of the G10-loop spectra (Fig. 3.4*B*). These variations are both considered hallmarks of parallel G-quadruplex formation (Huppert 2008, Gonzàlez et al. 2009). A second indication of a G quadruplex structure for the G10-loop is derived from denaturation experiments. In Fig. 3.4*C* the thermal melting profiles of G10-loop in the 25–105 °C interval are shown, whereas their corresponding static spectra are reported in Fig. 3.4*D*. These data indicate that the melting transition is still not complete at 105 °C for the G10-loop in the presence of 100mM KCl or NaCl, whereas a poorly

cooperative transition with a *Tm* centered at around 65–70 °C is observed when the same experiment is performed in the absence of monovalent cations. The predicted melting temperature for a G10-loop adopting hairpin structure is *Tm* =69.1 °C in 100 mM monovalent cations, according to the mfold server. Conversely the Quadpredict algorithm predicts higher *Tm* values, *i.e.* 94.9 and 77.1 °C, in the presence of 100 mM KCl or NaCl, respectively. These higher values are in better agreement with our experimental data, suggesting a G-quadruplex structure in these conditions.

Thermal melting profiles were also collected for the c-MYC oligo (Fig. 3.4*E*), as a control, and found to be similar to those obtained with the G10-loop oligo (Fig. 3.4*C*).

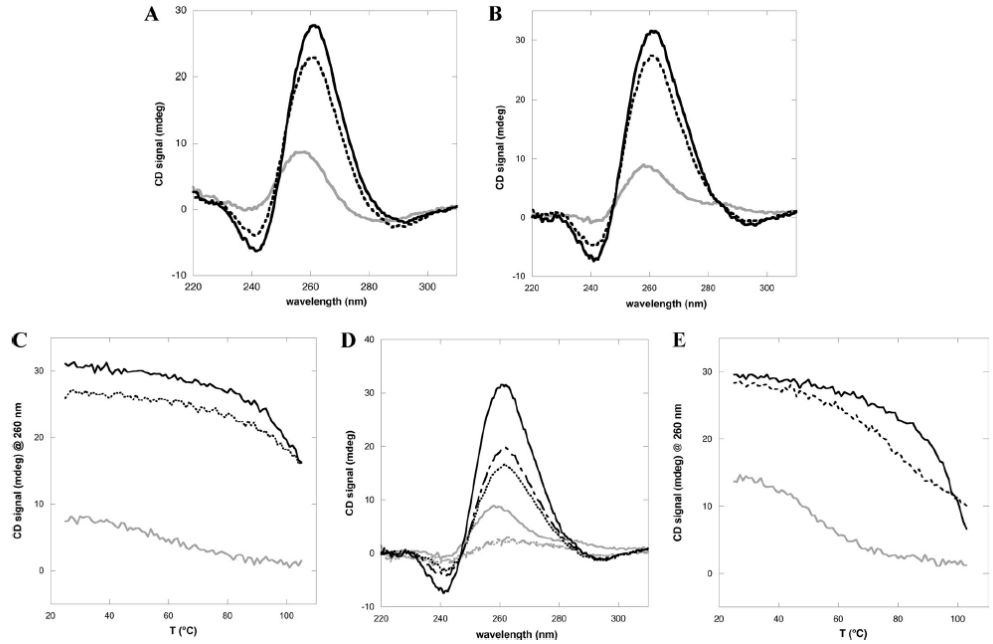

Fig. 3.4 G-10 loop adopts a G-Quadruplex fold in vitro. *A*, CD spectra of the c-MYC oligo annealed in the absence (*gray continuous line*) or presence of 100 mM

NaCl (*dashed line*) or 100mM KCl (*black continuous line*), respectively. *B*, CD spectra of the G10-loop oligo. The line format is the same as in *A*. *C*, thermal denaturation profiles of G10-loop in the 25–105 °C interval and in the absence (*gray continuous line*) or presence of 100mM NaCl (*dashed line*) or 100mM KCl (*black continuous line*). *D*, CD spectra of the G10-loop in the absence (*gray continuous*, *gray dashed dotted*, and *gray dotted* at 25, 95, and 105 °C, respectively) or presence of 100 mM KCl (*black continuous*, *black dashed dotted*, and *black dotted* at 25, 95, and 105 °C, respectively). *E*, thermal denaturation profiles of c-MYC oligo in the 25–105 °C interval, collected in the absence (*gray continuous line*) or presence of 100mM NaCl (*dashed line*) or 100mM KCl (*black continuous line*).

Finally we analyzed the binding of the porphyrin TmPyP4 to our oligos, by means of SPR. This molecule is known to bind G-quadruplex structures with high affinity and a complex stoichiometry involving at least two binding sites with different affinities (Freyer et al. 2007, Wei et al. 2006, Wei et al. 2010). The interaction of TmPyP4 with poly(T) oligo (Fig. 3.5*A*) determines a series of sensorgrams that, in the concentration range explored, always reach equilibrium before the end of the contact time between analyte and ligand, allowing both kinetic analysis, with a simple 1:1 model interaction (Fig. 3.5*A*), and the construction of a Scatchard plot (Fig. 3.5*B*). The average value for this double determination is K_D=345±15 nM and is compatible with the presence of a single binding site (Table 2). When analyzing the T10-loop data, again we observed a single binding site with a lower dissociation constant (K_D=35±15 nM) (Fig. 3.5, *C* and *D*, and Table 2). Conversely, when analyzing TmPyP4 binding to the c-MYC oligo (Fig. 3.5, *E* and *F*) a more complex behavior was observed, depending on the TmPyP4 concentration. At low concentrations (*bottom curves* of Fig. 3.5*E*) the dissociation phase is characterized by a single process corresponding to the off-rate constant of a high affinity binding site. By increasing the TmPyP4 concentration, the off-

rate becomes clearly biphasic (*upper curves* of Fig. 3.5*E*) reflecting the titration of a second binding site with lower affinity.

Moreover, not all the traces reach the equilibrium, so that the corresponding Scatchard plot is determined for a subpopulation of the porphyrin concentrations studied. As a result, the equilibrium analysis performed via the Scatchard plot is not compatible with a single binding site (see the *dashed line* in Fig. 3.5*F*), whereas an excellent agreement is obtained assuming two independent

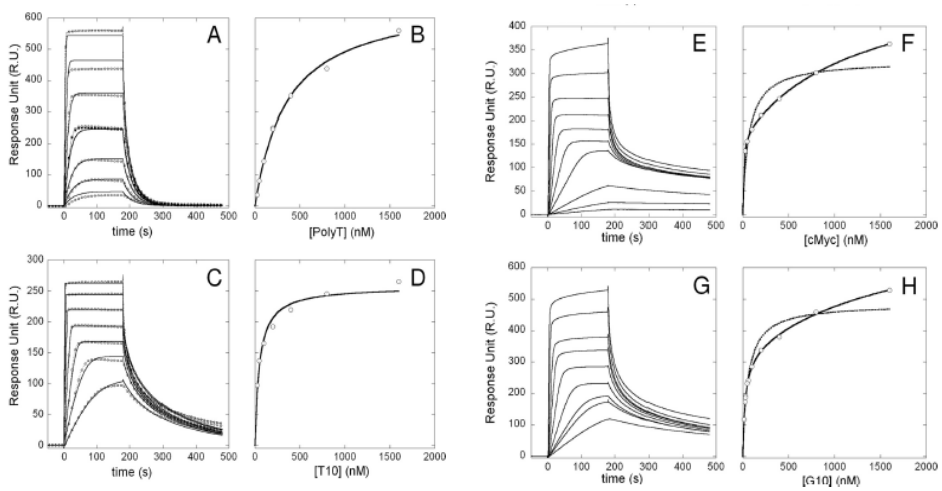

Fig. 3.5 **SPR sensorgrams of the interaction between the porphyrin TmPyP4 and different oligos (*panels A*, *C*, *E*, and *G* for poly(T), T10-loop, c-MYC, and G10-loop, respectively) and their corresponding Scatchard plots (*panels B,D, F*, and*H*).** *A*, poly(T), from bottom to top of the panel [TmPyP4]=1600, 800, 400, 200, 100, 50, and 25 nM. Experimental data are reported as *dashed lines with circles* (where the concentration was measured twice, also as *dashed line with squares*); best fit assuming a 1:1 interaction is reported as a *solid line*. *B*, poly(T), Scatchard plot, *open circles* are experimental data points and the *solid line* is the fit assuming a single binding site. *C*, T10-loop, from bottom to top of the panel [TmPyP4]_1600, 800, 400, 200, 100, 50, and 25 nM. Symbols used are the same as in *panel A*. *D*, T10-loop, Scatchard plot, the *continuous line* represents the best fit according to 1:1 interaction model. *E*, c-MYC, from bottom to top of the panel [TmPyP4]=1600, 800, 400, 200, 100, 50, 25, 12.5, 6.25, and 3.125 nM. Experimental data are reported as

solid lines. *F*, c-MYC, concentrations used in the Scatchard plot are [TmPyP4]=1600–25 nM. *Open circles* are experimental data points and the *dashed line* represents the best fit assuming a single binding site, whereas the *solid line* is the best fit assuming two independent binding sites. *G*, G10-loop, from bottom to top of the panel [TmPyP4]=1600, 800, 400, 200, 100, 62.5, 31.25, 25, and 15.625 nM. Experimental data are reported as *solid lines*. *H*, G10-loop, the concentrations used in the Scatchard plot are [TmPyP4] =1600–31.25 nM. Symbols used are the same as in *panel F*.

binding sites (*solid line* in Fig. 3.5*F* and see Table 3.2 for the corresponding *KD* values). Importantly, the experiments performed with the G10-loop indicated that TmPyP4 binding to this oligo follows the same behavior as observed with c-MYC, with the same stoichiometry and similar dissociation constants for the high and low affinity sites, respectively (Fig. 3.5, *G* and *H*, and Table 3.2). In conclusion, TmPyP4 binding data obtained with the G10-loop are in agreement with those obtained with c-MYC and with previous work on G-quadruplex forming oligos (Freyer et al. 2007, Wei et al. 2006, Wei et al. 2010), whereas data obtained with the hairpin T10-loop are not. Taken together, our results suggest that, in the presence of physiological amounts of monovalent cations, the G10-loop folds as a G-quadruplex *in vitro* and might also have this structure when recognized *in vivo* by NPM1 (Xu et al. 2007).

Ligand	**K_D** (M)	**k_{on} ($M^{-1}s^{-1}$)**	**k_{off} (s^{-1})**
Poly T	$330.9 \pm 0.4 \times 10^{-9}$ (Kinetics)	$1.013 \pm 0.001 \times 10^{8}$	33.52 ± 0.02
	$360 \pm 30 \times 10^{-9}$ (Scatchard)	-	-

	345 ±15 x 10^{-9} (Average value)		
T10-loop	20 ± 10 x 10^{-9} (Kinetics)	2.9 ± 0.8 x 10^{10}	6.2 ± 2 x 10^{2}
	50 ± 8 x 10^{-9} (Scatchard)	-	-
	35 ± 8 x 10^{-9}(Average value)		
c-Myc	Site 1: 8 ± 14 x 10^{-9}	-	-
	Site 2: 1.3 ± 0.2 x 10^{-6}	-	-
G10-loop	Site 1: 26 ± 3 x 10^{-9}	-	-
	Site 2: 1.9 ± 0.9 x 10^{-6}	-	-

Tab. 3.2 Dissociation constants and kinetic parameters of TmPyP4 binding to selected DNA, determined by SPR analysis

3.4 NPM1-C70 induces the Formation of G-quadruplex Structures

Having established that NPM1-C70 binds with high affinity preformed G-quadruplex structures, we next investigated whether it is able to induce G-quadruplex formation *in vitro*. In Fig. 3.6*A* we report CD spectra of the not annealed c-MYC oligo titrated with NPM1-C70, in the absence of monovalent ions. By progressively increasing the amount of NPM1-C70 we observe, once again, a red-shift and increase in signal of the 260 nm peak and the progressive formation of a through at 240 nm, which indicates that a G-quadruplex is formed upon protein binding. The same effect, albeit less evident, is obtained when using the G10-loop oligo (Fig. 3.6*B*). In a mirror experiment, using pre-annealed c-MYC or G10-loop oligos (Fig. 3.6, *C* and

D, respectively) we checked whether NPM1-C70 might have the property to destabilize pre-formed G-quadruplex structures. In both cases increasing amounts of protein had no effect on the CD signal of oligos pre-annealed in the presence of monovalent cations. Thus NPM1-C70 is able to induce G-quadruplex formation in unstructured oligos, whereas it does not unwind pre-structured oligos.

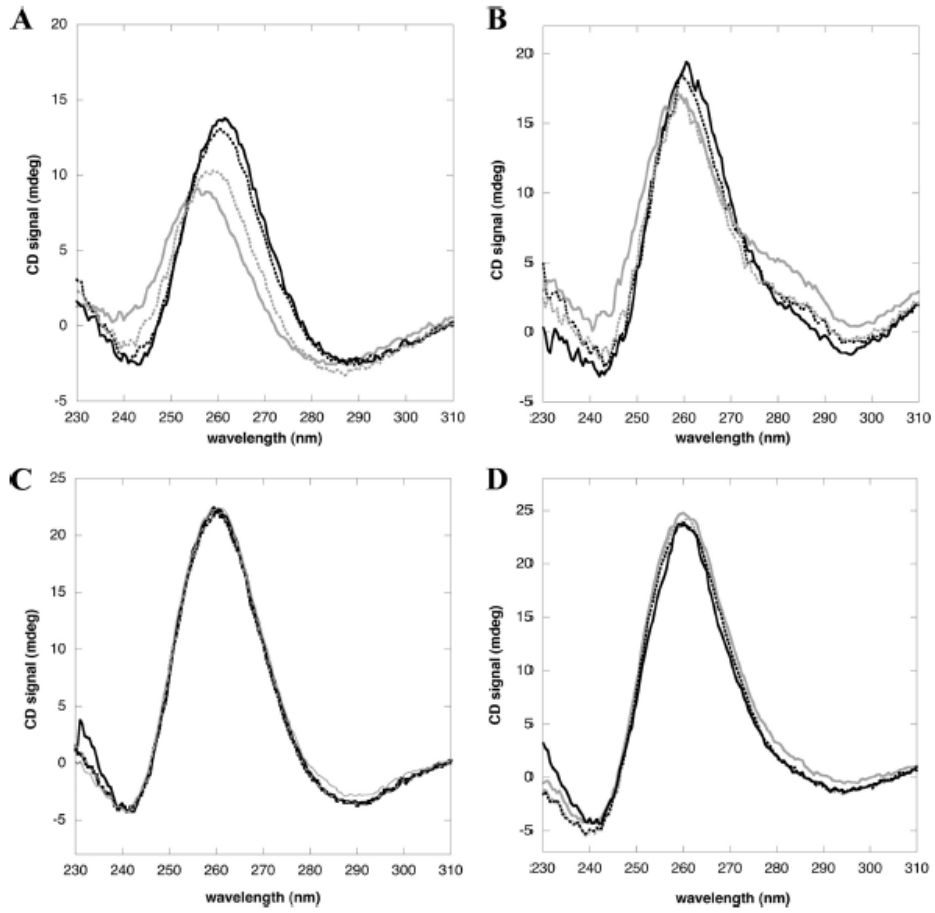

Fig. 3.6 NPM1-C70 is able to promote the formation of G-quadruplex arrangement in unstructured oligos, whereas it does not unfold G-

quadruplexes. *A*, not annealed c-MYC oligo (20μM) in the absence (*gray solid line*) and presence of increasing amounts of NPM1-C70 (*gray dots*, *black dots*, and *black solid line* for 10, 20, and 30 μM, respectively); the red shift and increase in amplitude at 260 nm and the formation of a through at 240 are indicative of G-quadruplex formation. *B*, not annealed G10-loop oligo (20 μM) in the absence (*gray solid line*) or presence of increasing amounts of NPM1-C70 (*gray dots*, *black dots*, and *black solid line* for 70, 100, and 150 μM, respectively). *C*, spectra of c-MYC oligo (20 μM), pre-annealed in the presence of 100 mM NaCl, and incubated with increasing amounts of NPM1-C70 (same amounts as in *panel A*). *D*, same experiment as in *panel C* performed using the G10-loop (20 μM). Amounts of NPM1-C70 are the same as in *panel B*.

3.5 Identification of key residues in the NPM1-quadruplex interaction

Data reported so far strongly suggest that the NPM1 region comprised between the longer NPM1-C70 construct and the shorter NPM1-C53 construct (aa 225-241) is necessary for high affinity binding. This region contains five lysine residues that we hypothesized might play a role in the specific recognition played by NPM1-C70 *vs* NPM1-C53 (Fig. 3.1). To test this hypothesis we mutated each of them into alanine. The K229A-K230A double mutant was also prepared. With these variant proteins we performed a complete set of SPR experiments to measure the binding affinities with the G10-loop and c-MYC oligos. Interestingly, none of the single mutants significantly affected the interaction when the variant proteins were tested against both oligos (Table 2.5). However, when testing the double mutant K229A-K230A, we observed a marked reduction in affinity and obtained a K_D= 135 μM for the G10-loop and K_D= 78 μM for the c-MYC oligo. These values are very close to those obtained when testing the two oligos with NPM1-C53 (compare Table 3.1 with Table 3.3) and suggest that residues

K229 and K230 cooperate to the specific and high affinity recognition played by NPM1-C70 on both oligos. Besides that, they also suggest that the mode of recognition of the c-MYC and G10-loop is likely to be similar, pointing again to a G-quadruplex structure for the G-loop oligo.

	G10-loop			**cMYC**		
	K_D (µM)	**k_{on} ($M^{-1}s^{-1}$)**	**k_{off} (s^{-1})**	**K_D (µM)**	**k_{on} ($M^{-1}s^{-1}$)**	**k_{off} (s^{-1})**
NPM-C70	7.2 ± 0.2	2.8 ± 0.1 x 10^2	1.98 ± 0.03 x 10^{-3}	1.9 ± 0.1	7.4 ± 0.3 x 10^3	1.37 ± 0.03 x 10^{-2}
K229A	14.7 ± 0.8	1.6 ± 0.2 x 10^2	2.30 ± 0.04 x 10^{-3}	10.3 ± 0.9	7.3 ± 0.4 x10^2	7.5 ± 0.3 x10^{-3}
K230A	10.9 ± 0.3	1.50 ± 0.02 x 10^2	1.60 ± 0.02 x 10^{-3}	15 ± 1	2.8 ± 0.1 x10^2	4.3 ± 0.1 x10^{-3}
K233A	7.5 ± 0.5	4.5 ± 0.2 x 10^2	3.4 ± 0.1 x 10^{-3}	17.2 ± 0.6	3.0 ± 0.1 x10^2	5.2 ± 0.1 x10^{-3}
K236A	8.1 ± 0.2	3.3 ± 0.1 x 10^2	2.70 ± 0.02 x 10^{-3}	10 ± 1	2.8 ± 0.2 x10^2	2.7 ± 0.2 x10^{-2}
K239A	19.8 ± 0.5	1.04 ± 0.02 x 10^2	2.05 ± 0.02 x 10^{-3}	6.8 ± 0.9	8.7 ± 0.7 x10^2	5.9 ± 0.3 x10^{-3}
K229A-K230A	135 ± 8	-	-	78 ± 7	-	-

Tab. 3.3 Dissociation constants and kinetic parameters of different NPM1-C70 mutants binding as determined by SPR analysis

3.6 Structural characterization of NPM1 C-terminal binding domain

By the experiments showed above we have investigated the NPM1 C-terminal domain DNA binding activity trying to identify specific requirements for high affinity recognition both at the protein and DNA level. Our data suggest that the two protein constructs we analyzed, i.e. NPM1-C70 and NPM1-C53, have a markedly different DNA binding activity. The shortest one, NPM1-C53, that was shown by Grummitt et al. to fold in a three helix bundle, is poorly competent for DNA binding, compared to the longest construct, NPM1-C70. Experimental evidence also suggests that NPM1-C70 binds with high affinity specific DNA sequences that are able to fold in G-Quadruplexes. This feature was initially characterized with a well-known a G-quadruplex forming region, both *in vitro* and *in vivo*, present in *c-Myc* promoter region (Gonzales et al., 2010) and later extended to the physiological NPM1 target found at the *SOD2* promoter region used for initial experiments fold as a G-Quadruplex. Finally to definitively establish NPM1 as a *bona fide* G-quadruplex binding protein, we also showed that NPM1-C70 is able to induce G-quadruplex folding in the absence of stabilizing cations, and started to map residues important for G-quadruplex recognition in the N-terminal tail of the NPM1-C70 construct.

Therefore NPM1 can be added to the increasing list of G-Quadruplex binding proteins. Even though the importance of G-quadruplex recognition for a variety of physiological processes is now clear, only very little structural information is available concerning the molecular recognition mechanism of complex formation (Sissi et al. 2011). To the best of our knowledge, only two structures are available: i) thrombin in complex with a synthetic aptamer that

folds as a G-quadruplex (Padmanabhan et al. 1993), and ii) the *Oxytricha nova* TEBP heterodimer bound to its telomeric sequence (Horvath et al. 2001).

Therefore we decided to analyze the interaction of NPM1 DNA binding domain with the G-quadruplex sequence from the NHE_{III} region of the *c-MYC* promoter, using NMR techniques. This work was performed in collaboration with Prof. Ivano Bertini and Prof. Lucia Banci, at CERM laboratories, University of Florence.

For NMR experiments we used the NPM1-C70 protein construct, having previously shown that the shorter NPM1-C53 is poorly competent for DNA binding and that the 17-residue lysine-rich region at its N-terminus (residues 225-241) that differentiate NPM1-C70 from NPM-C53 is necessary for high affinity.

As a first step we determined, by a combination of standard NMR techniques, the structure of NPM1-C70, since only the structure of NPM1-C53 was known (Grummitt et al., 2008). The protocol used for structure determination is detailed in the Materials and Methods section while statistics about structure determination are shown in Table 3.4.

The structure depicted in Figure 3.7A comprises a well-defined 3-helix bundle, similar to the NPM1-C53 construct (Grummitt et al. 2008), in terms of length, relative orientation and hydrophobic interactions between all the helices in pair (Fig. 3.7A). On the other hand, the lysine-rich region (residues 225-241) that enhances DNA binding is unstructured, as indicated by high values of ^{15}N-^{1}H R_1 (Fig. 3.7B) and low values for ^{15}N-^{1}H R_2 (Fig. 3.7C) and heteronuclear NOEs (Fig. 3.7D), with no secondary structure elements, nor

any propensity to fractionally take them as clearly shown from the chemical shift index analysis (Fig. 3.7B-C). Conversely, the folded part of the NPM1-C70 construct shows ^{15}N-^{1}H R_1 and ^{15}N-^{1}H R_2 values typical of an 8 kDa protein (Fig. 3.7B-C) and positive values for heteronuclear ^{15}N{^{1}H} NOEs values (Fig. 3.7D), even though the latter are lower than those of a rigid protein, thus indicating the occurrence of fast internal motions.

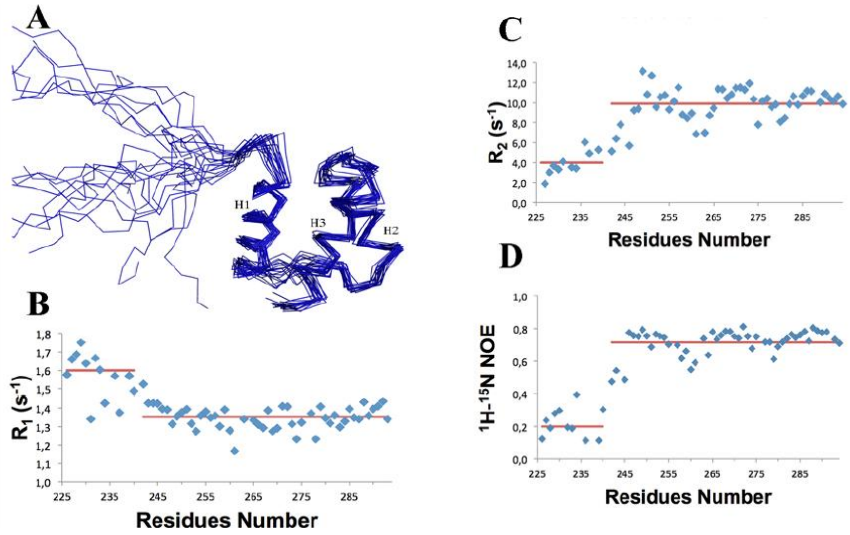

Fig. 3.7 NPM1-C70 quadruplex-binding C-terminal domain, encompassing residues 225-294. A) NMR solution structure of NPM1-C70 showing the twenty lower-energy structures. A lysine-rich natively unstructured segment (aa 225-242) precedes the terminal three-helix bundle. B) ^{15}N-^{1}H R_1 NOE values and C) ^{15}N-^{1}H R_2 NOE values. Low R_1 and high R_2 values for segment 225-242 are consistent with the N-terminal tail being unstructured. D) Heteronuclear ^{15}N{^{1}H} NOEs are positive but smaller than expected for a 8 kDa protein, indicating fast internal motion in the three-helix bundle. Red lines indicate average values for the 225-242 and the 243-294 segments, respectively.

	AMBER 10.0[a]
RMS deviations per meaningful distance constraint (Å)[b]:	(20 structures)
Intraresidue (189)	0.0065 ± 0.0063
Sequential (279)	0.0208 ± 0.0062
Medium range (220)[c]	0.0198 ± 0.0031
Long range (40)	0.0106 ± 0.0122
Total (728)	0.0178 ± 0.0015
RMS violations per meaningful dihedral angle constraints (deg)[b]:	
Phi (42)	1.20 ± 1.01
Psi (42)	6.17 ± 4.19
Average no. of NOE violations larger than 0.3 Å	0.05 ± 0.22
Average NOE deviation ($Å^2$)[d]	0.01 ± 0.03
Average angle deviation (rad^2)[d]	0.66 ± 0.10
RMSD to the mean structure (Å) (BB)[e] RMSD to the mean structure (Å) (HA)[e]	0.80 ± 0.21 1.40 ± 0.15
Structural analysis	
% of residues in most favorable regions[f, g]	96.5%
% of residues in allowed regions[f, g]	3.3%
% of residues in generously allowed regions[f, g]	0.1%
% of residues in disallowed regions[f, g]	0.0%
G-factor[f, g]	0.03
Structure Z-scores[g, h]	
1st generation packing quality	1.27
2nd generation packing quality	4.51
Ramachandran plot appearance	-1.50
$\chi1/\chi2$ rotamer normality	-3.04
Backbone conformation	-0.50

[a]AMBER indicates the energy minimized family of 20 structures.
[b]The number of meaningful constraints for each class is reported in parenthesis.
[c]Medium range distance constraints are those between residues (i,i+2), (i,i+3), (i,i+4) and (i,i+5).
[d]NOE and torsion angle constraints were applied with force constants of 20 kcal mol^{-1} $Å^{-2}$, 20 kcal mol^{-1} rad^{-2}, respectively.
[e]The RMSD to the mean structure is reported considering the segment 20-68.
[f]As it results from the Ramachandran plot analysis. For the PROCHECK statistics, an overall G-factor larger than -0.5 are expected for a good quality structure.
[g]The statistic analysis is reported considering the segment 20-68.
[h]Values based on WHAT-IF output. A Z-score is defined as the deviation from the average value for this indicator observed in a database of high-resolution crystal structures, expressed in units of the standard deviation of this database-derived average. Typically, Z-scores below a value of -3 are considered poor, those below -4 are considered bad.

Table 3.4 Statistical analysis of the energy minimized family of conformers of NPM1-C70

3.7 The complex of NPM1-C70 with c-MYC G-quadruplex DNA

We have already shown that NPM1-C70 binds a DNA oligonucleotide resembling a specific sequence found at the NHE$_{III}$ region of the *c-MYC* promoter and that this 27-mer region (c-MYC Pu27) of sequence, 5'-

T**GGGG**A**GGG**T**GGGG**A**GGG**T**GGGG**AAGG-3', is known to form a parallel G quadruplex structure in the presence of K^+ at physiological concentration (100- 150 mM) (Gonzàlez and Hurley 2010). Pu27 contains five runs of three or more consecutive guanines and therefore can, in principle, populate several G-quadruplex structures with different topologies (Huppert 2008). These multiple conformations are indeed observed in the free state, as monitored by multiple sets of NMR signals in slow exchange in the 1D ^{1}H NMR spectra (Fig. 3.8A, blue line). Moreover, Pu27 maintains different G-quadruplex topologies also when bound to NPM1-C70 (Fig. 3.8A, red line). The structural heterogeneity of the Pu27 sample complicates the analysis of its interaction with NPM1-C70.

However, recently, Patel and coworkers showed that a *c-MYC* derived shorter oligonucleotide of 24 residues, containing a guanine to inosine substitution in one of the guanine runs, populates only one of the possible G-Quadruplex conformations (called Pu24I: 5'-T**GA**G**GG**T**GGIG**A**GGG**T**GGGG**AAGG-3') (Phan et al. 2005). Therefore we decided to use this oligonucleotide for further studies that were greatly facilitated by the availability of the ^{1}H nuclei assignment for the Pu24I oligonucleotide, gently provided to us by courtesy of Anh Tuan Phan and Vitaly Kuryavyi (Phan et al. 2005).

First, by comparing the 1D ^{1}H NMR spectra of Pu24I in the free state and bound to NPM1-C70, we confirmed that Pu24I, differently from Pu27, displays only a single G-quadruplex topology that is retained after NPM1-C70 binding (Fig 3.8B). Then, to assess whether Pu24I undergoes major conformational changes upon NPM1-C70 binding, we performed intramolecular NOEs experiments for the bound state of Pu24I and compared

it to its free state. To characterize the bound state a ω_1 ^{13}C-filtered, ω_2-^{13}C-filtered NOESY experiment was recorded in a 2D plane (^{1}H-^{1}H plane) while, in the case of Pu24I free state, a standard 2D NOESY was performed. As shown in Figure 3.8C, no major variations are visible in the superimposition of the two spectra, indicating that the G-quadruplex structure of Pu24I is maintaining its conformation when bound to NPM1-C70.

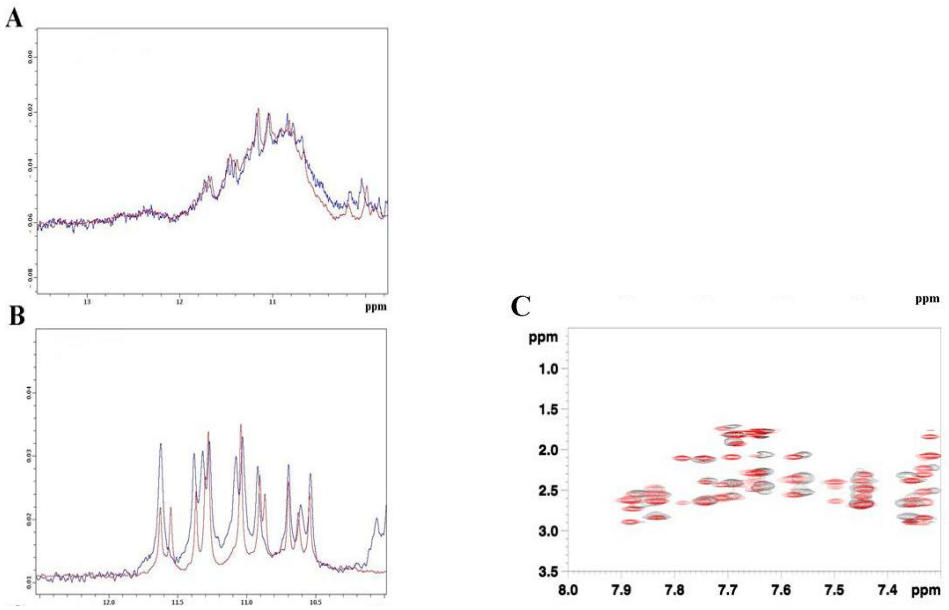

Fig. 3.8 Analysis of Pu27 and Pu24I G-quadruplex conformations. A) Superimposition of the 1D ^{1}H NMR spectra of Pu27 in the free state (blue) and bound (red) to C70-NPM1. Both spectra were acquired at 700MHz and at 290K. B) Superimposition of the 1D ^{1}H NMR spectra of Pu24I in the free state (blue) and bound (red) to C70-NPM1. Both spectra were acquired at 700MHz and 290K. C) Details of the superimposition between ω_1-^{13}C-filtered, ω_2-^{13}C-filtered NOESY experiments in a 2D plane (^{1}H-^{1}H plane) of Pu24I-NPM1-C70 complex (black) and a classical 2D NOESY of Pu24I in its free state (red). Both spectra were acquired at 700 MHz and 290K.

To further investigate the interaction of NPM1-C70 with Pu24I, we titrated ^{15}N-labeled NPM1-C70 with increasing amounts of unlabeled Pu24I, following ^{1}H-^{15}N HSQC spectra of labeled NPM1-C70. (Fig. 3.9).

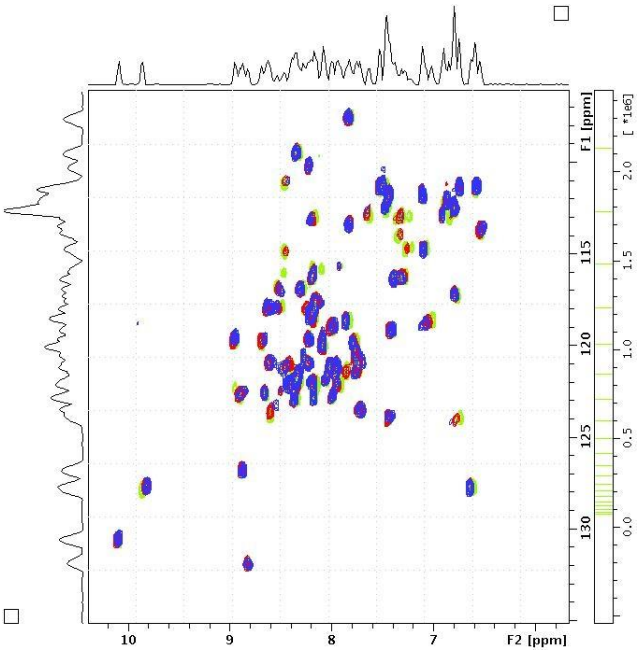

Fig. 3.9 ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled NPM1-C70. ^{1}H-^{15}N HSQC spectra of labeled NPM1-C70 before (green) and after addition of unlabeled Pu24I, respectively 50% (red) and 100% (blue) of NPM1-C70 concentration. Spectra were acquired at 700MHz and 290K.

The NMR titration results confirm the interaction between the NPM1-C70 and Pu24I and, by analyzing ^{1}H-^{15}N HSQC spectra, we observed both the appearance of new peaks and the disappearance of others, indicating the formation of a new species that exchanges slowly with the free protein on the NMR time scale, i.e. $< 10^{-2}$ s^{-1} (Fig. 3.9 and 3.10A). The ratio of the intensity of the signals, for a given resonance in the two species, changes linearly with

the amount of Pu24I, and reaches a plateau upon addition of stoichiometric amounts of Pu24I, consistently with a 1:1 protein- Pu24I complex (Fig. 3.9).

In particular some signals of the protein experience a significant change in chemical shift (δ_{HN}). To define which residues are interacting residue the CSP (Chemical Shifts Perturbation) should be higher than the average variations plus one standard deviation ($<\delta_{HN}> = 0.043 \pm 0.025$ ppm, Black line in Fig. 3.10B). Also signals that broaden their linewidth or disappear during the titration were taken under consideration.

This analysis shows that NPM1-C70 interacts with Pu24I mainly *via* hydrophilic residues, located in helices H1 and H2 and exposed on the same side of the protein (Fig. 3.10B and Fig 3.10C). Interestingly, also a few residues located on helix H3 display chemical shifts variations (Fig. 3.10B and Fig. 3.10C). Since the latter are buried and interact with helices H1 and H2, we conclude that a strong coupling is present among the 3 helices. In the protein-Pu24I complex, no chemical shift variations are detected for residues belonging to the NPM1-C70 N-terminal unstructured segment, indicating that this tail is not directly involved in the complex (Fig. 3.10B). This surprising result is confirmed by analysis of the ^{15}N-heteronuclear relaxation data in the complex, where the N-terminal tail is still characterized by high values for heteronuclear ^{15}N-^{1}H R_1 (Fig. 3.10D) and low values for heteronuclear ^{15}N-^{1}H R_2 (Fig. 3.10E) and heteronuclear NOEs (Fig. 3.10F), consistent with a natively unstructured state. Conversely, the folded C-terminal region of NPM1-C70 shows ^{15}N-^{1}H R_1 and ^{15}N-^{1}H R_2 values typical of a complex of 16 kDa molecular weight (in agreement with the sum of the molecular weights

of NPM1-C70 and Pu24I) and positive values for heteronuclear $^{15}N\{^{1}H\}$ NOE values (Fig. 3.10E-D-F).

Both results suggest that the complex between NPM1-C70 and Pu24I is stable, consistent with the high affinity described in previous observation for the Pu27 oligonucleotide. The relaxation data analysis also pointed out that the heteronuclear $^{15}N\{^{1}H\}$ NOE values increase in the folded region of NPM1-C70, indicating an increasing rigidity for the three-helix bundle of NPM1-C70 upon binding of Pu24I (Fig. 3.10F).

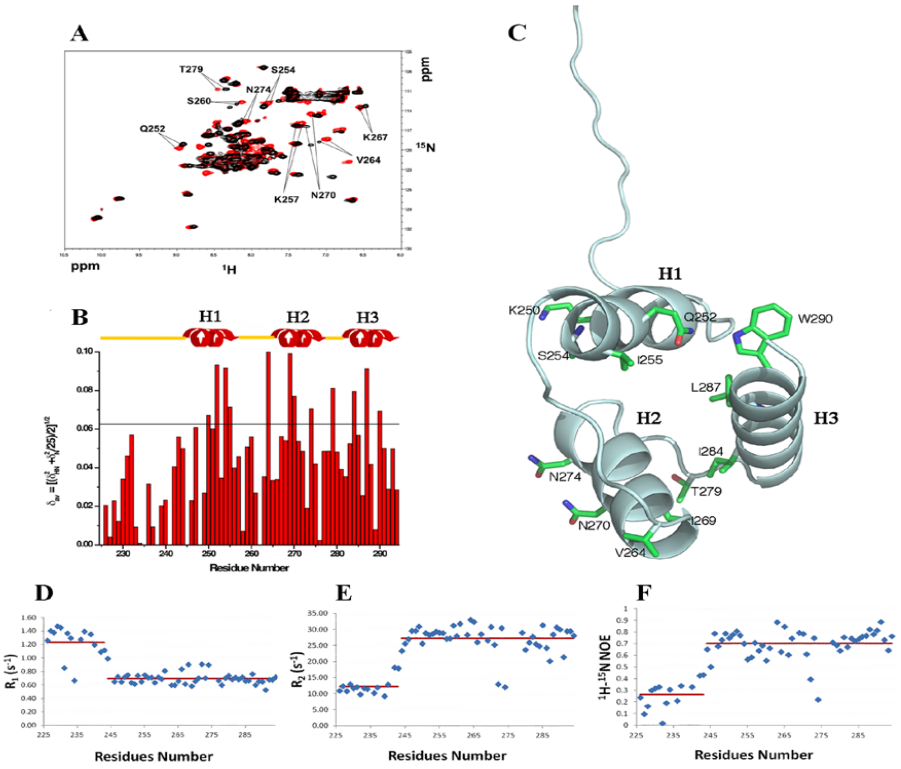

Fig. 3.10 Interaction of NPM1-C70 with the Pu24I G-quadruplex. A) ^{15}N HSQC spectra of the protein before (black) and after addition of stoichiometric amounts of unlabeled Pu24I (red). Representative chemical shift variations are labeled

indicating relevant residues. B) Chemical shift variations cluster in the three-helix bundle while they are not found at the N-terminal 225-242 segment. The horizontal black line indicates the average chemical shift variation plus one standard deviation upon Pu24I addition. C) Residues experiencing chemical shift variations higher than the average plus one standard deviation are highlighted on the structure of the protein. Residues belonging to helices H1 and H2 are solvent exposed. A few hydrophobic residues belonging to helix H3 are also affected indicating coupling between the helices upon Pu24I binding. D) Heteronuclear 15N-1H R1 and E) heteronuclear 15N-1H R2 values for NPM1-C70 in complex with Pu24I indicate that the N-terminal region flanking the three-helix bundle remains unstructured after Pu24I binding. F) The increase of heteronuclear 15N{1H} NOE values in the three helixbundle upon Pu24I binding (see Figure 3.7D for comparison) suggests increate rigidity.

3.8 Experimentally restrained molecular docking of the NPM1-C70 – Pu24I complex

To gain additional information on the structure of the complex, ^{13}C-filtered NOESY experiments were performed. Six intermolecular NOEs were detected between NPM1-C70 and Pu24I in the complex (Table 3.5). An example of assignment of NOE cross-peaks is shown in Figure 3.11A. They involve Lys and Asn residues of the protein (whose sidechains are let free to rotate during the subsequent docking calculation) and sugar backbone protons on Pu24I. These distance restraints, together with the chemical shifts perturbations defined as Ambiguous Interaction Restraints (AIRs; also listed in Table 3.5), were used to calculate a structural model of the NPM1-C70/Pu24I complex, within the data-driven HADDOCK docking program (see the Experimental section for details about the protocol used).

Intermolecular NOEs		
NPM1-C70 atoms	***Pu24I atoms***	***Distance (Å)***
Lys 250-HG3/HG2	G13-H3'	4.01
Lys 250-HE2/HE3	G11-H5'	5.83
Lys 257-HD2/HD3	A12-H1'	5.53
Lys 257-HD3/HD2	T1-H72	4.03
Asn 270-HD22/HD21	G15-H5''	4.52
Asn 274-HD22/HD21	G23-H21	3.12
Ambiguous Interaction Restraints		
Active Residues of NPM1-C70		***Active Residues of Pu24I***
Glu245		Gua5
Lys248		Gua13
Lys250		Gua17
Gln252		Gua20
Ser254		
Ile255		
Lys257		
Val264		
Phe268		
Ile269		
Asn270		
Asn274		
Phe276		
Thr279		
Ile284		
Leu287		
Trp290		

Table 3.5 Intermolecular NOEs and AIRs used for HADDOCK docking calculations

Importantly, HADDOCK calculations identified a single cluster of docking poses. In particular, 184 final complex structures were obtained at the end of the procedure with a RMSD from the lowest energy solution of 1.5 + 0.9 Å. As an example, the twentieth lower Energy docking poses are reported in Figure 3.11B. As shown in Figure 3.11C the interaction surface involves one side of the three-helix bundle, the interacting residues being all located on helices H1 and H2, with a buried area of 1358.6 +/- 92.2 Å2. Based on the assignment of the intermolecular NOEs (Figure 3.11A and Table 3.5), the interaction involves side chains of NPM1-C70 residues Lys250, Lys257, Asn270 and Asn274. On the Pu24I side, intermolecular NOEs indicate the

involvement of protons of the backbone (mainly the deoxyribose ring) of nucleotides T1, G11, A12, G13, G15 and G23. From the structure model analysis, a linear stretch of nucleotides (from G11 to G16) located on one side of

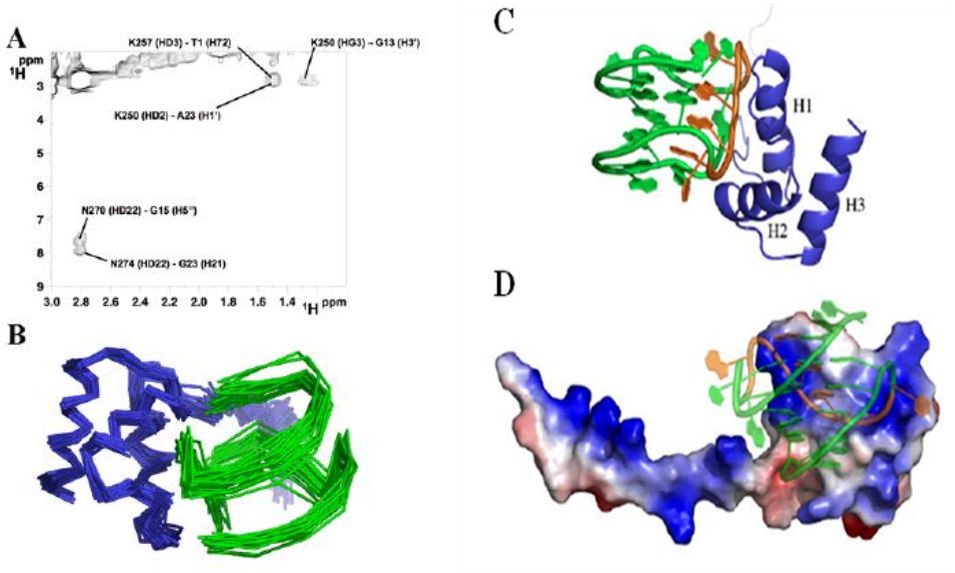

Fig. 3.11 Intermolecular NOEs and HADDOCK calculations and structural model of the NPM1-C70/Pu24I complex. A) An example of NOE cross peaks assignment from ω_1-^{13}C-edited, ω_2-^{13}C-filtered NOESY experiments in a 2D plane (1H-1H plane) of the NPM1-C70-Pu24I complex acquired on a ^{13}C,^{15}N NPM1-C70/unlabeled Pu24I sample. B) The twentieth lower energy complex structures obtained by NMR-data restrained molecular docking calculations using the HADDOCK protocol. NPM1-C70 is represented in blue and Pu24I in green. C) Ribbon representation of the lowest energy model, showing helices H1 and H2 of NPM1-C70 approaching the G-quadruplex "laterally" and interacting with a specific segment of the backbone (in orange). D) NPM1-C70 is represented with its electrostatic surface (blue for positive and red for negative, respectively), while Pu24I is in ribbon. The Pu24I structure is shown in transparency to highlight the small positively charged groove in between helices H1 and H2 that accommodates a stretch of Pu24I nucleotides (G11-G16, colored in orange). The long unstructured tail is also positively charged and may play a role in long range electrostatic interactions with the approaching oligonucleotide.

the G-quadruplex scaffold (Fig 3.11C), intercalates into a groove formed by helices H1 and H2 (Fig. 3.11D). Interestingly, this stretch contributes to the formation of each of the three stacked guanine tetrads in the G-quadruplex scaffold. By analyzing the twentieth lower energy docking solutions it appears that residues Lys250, Lys257, Asn270 and Asn274 always participate to the formation of salt bridges or hydrogen bonds with backbone phosphates belonging to the G11-G16 linear stretch. In addition, residues Lys267 and Cys275 are always found at the interface, where they contact phosphate groups in the G11-G16 stretch, in all twentieth lower energy docking poses.

4 Discussion

The initial aim of this work was to better understand the DNA binding features of NPM1, starting from foundations already established by others, that described the NPM1 C-terminal domain as responsible of DNA binding with a increased affinity for single-stranded structures over duplexes and no sequence requirements (Wang et al. 1989, Dumbar et al. 1994).

Since binding experiments reported so far were rather qualitative, we resorted to develop binding assays with the Surface Plasmon Resonance (SPR) technology in order to quantitatively describe the NPM1-DNA binding reaction.

Our first SPR experiments indicate that, besides the shorter NPM1-C53 construct being able to recognize any oligonucleotide tested with low affinity, high affinity recognition is achieved only with NPM1-C70, thus establishing that the lysine rich tail located at the N-terminus of the C-terminal three-helix bundle of the protein plays a role in DNA recognition (Hingonari et al., 2000).

Moreover, rather than just being a single stranded binding protein as initially suggested, NPM1's preferential binding is for oligonucleotides with defined 3D structure. Indeed this is the case of the G10-loop, a sequence found at the promoter of the *SOD2* gene, which is bound by NPM1 *in vivo* and was predicted to form a hairpin based on conventional Watson-Crick pairing (Xu et al., 2007).

However, a predicted hairpin structure is not sufficient to confer high binding affinity, as shown by the results obtained with an alternative oligonucleotide, (T-10 loop), or modifying the oligonucleotide in the hairpin by reducing the

number of guanines in the loop (G-5 loop). We showed that in both cases affinity is reduced (Fig. 3.1, Tab. 3.1). These experiments suggest that both the presence of guanines and their number (table 3.1) in the loop contribute to the global affinity, indicating that the 3D structure adopted by the G-10 loop oligo plays a critical role in recognition.

Further investigations suggested that not only the predicted hairpin structure of the G10-loop is not *per se* sufficient to confer protein high affinity recognition but also indicated that the expected hairpin might not be the real structure adopted by the G10-loop *in vitro* and possibly also *in vivo.*

First, we showed that NPM1-C70 is able to recognize with high affinity oligos characterized by G-rich sequences, and specifically those obeying the G-Quadruplex folding rule (Huppert and Balasumbramanian 2005). This was shown first using a poly(G) oligo of the same length of the G-10 loop and, importantly, further confirmed with the well-characterized G-Quadruplex forming sequence located at the Nuclease Hypersensitive Element III_i ($NHEIII_1$) of the c-MYC promoter (Pu27), a 27-mer region that is known to adopt a G-quadruplex conformation both *in vitro* and *in vivo*.

Since the G-10 loop sequence obeys itself to the G-quadruplex folding rule, we started investigating whether this might be the real structure adopted by the oligonucleotide.

First, CD spectroscopy shows that the G10-loop presents all the hallmarks of parallel G-Quadruplex formation in the presence of physiological concentrations of K^+ or Na^+ ions. Moreover we found that the thermal unfolding of this oligo in the presence of monovalent cations is still not complete at temperatures as high as 105°C, reflecting the enhanced stability

conferred to G-Quadruplex structures by K^+ and Na^+. Finally we showed that the G10-loop, contrary to the T10-loop, is capable of binding the porphyrin TmPyP4 with the same stoichiometry and similar dissociation constants of the Pu27 c-MYC oligo, behaving as expected for a G-Quadruplex forming oligo.

Taken together, these data strongly suggest that the G10-loop is indeed folded as a G-Quadruplex *in vitro*. Moreover they suggest that NPM1 rather that recognizing specific sequences, shows a preference for specific 3D structures and in particular for G-quadruplexes. This was further confirmed by experiments showing that NPM1-C70, besides binding G-quadruplex forming regions, is also capable of promoting G-quadruplex formation *in vitro*. This property may be relevant for NPM1 regulatory functions in the cell and needs to be further investigated, especially in light of a putative more general role of the protein, than originally expected, in transcriptional regulation. For instance, it was shown that NPM1, through its interaction with the G-10 loop, promotes the transcriptional activation of the *SOD2* gene (Xu et al. 2007) and it is thus possible to hypothesize that other genes, containing G-quadruplexes in their promoter genes, might be regulated by NPM1.

Under this light it is interesting to note that a large number of promoter regions have been investigated and shown to contain regions capable of forming fully folded G-Quadruplexes at least *in vitro* (Qin and Hurley 2008). In particular these regions are frequently associated to oncogene promoters while they are generally absent in tumor suppressor genes promoters. This association of G-Quadruplex with gene promoters was consistent with the

hypothesis of linking G-Quadruplexes formation with transcriptional control of genes (Simonson et al. 1998).

One of the best characterized G-Quadruplex regulating activity is represented by the $NHEIII_1$ region at the c-*MYC* promoter, that is able to suppress up to 90% of c-*MYC* gene *in vivo* transcription when adopting a G-Quadruplex structure, thus inhibiting tumor growth in xenograft tumor mouse models (Grand et al. 2002). In other cases G-quadruplex stabilization may be associated to gene activation, possibly through protein binding (Huppert 2008), as already shown for the *SOD2* promoter (Xu et al. 2007).

The selective localization at oncogene promoters and telomeres makes of G-quadruplex DNA an attractive target for tumor treatment, and the structures of several G-quadruplex regions alone and in complex with drugs have been reported (Neidle 2009). Conversely, although the list of proteins that bind G-quadruplex DNA is rapidly increasing and the importance of such interactions for a variety of physiological processes is now clear, very little structural information is available concerning the molecular recognition mechanism (Sissi et al. 2011). To the best of our knowledge, only two structures are available: (i) thrombin in complex with a synthetic aptamer that folds as a G-quadruplex (Padmanabhan et al. 1993) and (ii) the Oxytricha nova telomere-binding protein heterodimer bound to its telomeric sequence (Horvath et al. 2001).

Thus, in the second part of my Ph.D. work, we decided to characterize the NPM1-G-quadruplex interaction from a structural point of view, both because such data could shed new light on the roles played by NPM1 in the cell that are lost in the leukemic blasts, and also, more in general, because

they would contribute to better understand protein-G-quadruplex interaction properties. Among the several oligonucleotides tested, we focused our attention on the interaction of NPM1-C70 with the G-quadruplex at the NHEIII region of the *c-MYC* promoter, because this is the DNA sequence bound with the highest affinity whose structure is known (Linge et al. 2003).

The structure of NPM1-C70 alone and bound to the DNA fragment was investigated by a combination of NMR experiments and docking calculations guided by experimental restraints. As shown in Figure 3.11, NPM1-C70 binds the Pu24I G-Quadruplex through a specific surface at the interface between helices H1 and H2. Several positively charged and polar residues establish interactions mainly with phosphate groups of a linear stretch of nucleotides, which fits the small groove at the H1-H2 interface. Interestingly, this stretch contributes to the formation of the main G-Quadruplex scaffold while the interactions with nucleotides belonging to the connecting loops appear marginal. This may explain why NPM1 recognizes, with comparable affinity, several G-Quadruplexes, which differ in loop length and distribution.

Analyzing NPM1-C70 residues involved in the interaction, much interest was generated by Lys257 and Lys267, because they are acetylated *in vivo* by p300 and deacetylated by SIRT1 (Shandilya et al. 2009). NPM1 acetylation results in dislocation of the protein from nucleoli to the nucleoplasm, where NPM1 interacts with transcriptionally active RNA Polimerase II. Our finding that these two residues are at the interface with Pu24I DNA suggests that loss of nucleolar localization may be due to impaired DNA binding at the nucleoli coupled to acetylation.

Interestingly, we also found from docking simulations that Cys275 is always located at the center of the surface buried by Pu24I. Cys275 is targeted and alkylated by the natural antitumoral compound (+)-avrainvillamide (Wulf et al. 2007). Treatment of LNCaP or T-47D cells with (+)-avrainvillamide leads to an increase in cellular p53 concentrations and promotes apoptosis. It is therefore conceivable that these effects may be linked to (+) avrainvillamide mediated impairment of NPM1 nucleic acid binding efficiency.

When we investigated the structure of NPM1-C70, both alone and in complex with the c-MYC quadruplex, we were particularly interested in analyzing the arrangement of the N-terminal tail of NPM1-C70 construct, the lysine-rich 17-residue region that differentiate NPM1-C70 from NPM1-C53. Indeed we had previously shown that this region is i) required for high affinity binding and ii) does not form any particular secondary structure element, neither before nor after binding. To our surprise we showed that this region is completely unstructured both before binding and in complex with Pu24I and, more importantly, it does not contribute to the final complex.

By inspecting the literature, however, we realized that this finding parallels previous observations in other systems. For instance, the affinity of the cAMP- responsive element-binding protein KID domain for the KIX domain of the cAMP-responsive element-binding protein- binding protein is significantly reduced when an unstructured portion of the domain that does not participate directly to the complex contact surface is deleted (Zor et al. 2002). Similarly, the interaction of splicing factor 1 with the large subunit of the U2 small nuclear RNA auxiliary factor (U2AF65) is affected by flanking unstructured regions that do not physically contact the partner (Selenko et al.

2003). This phenomenon, which is more frequent than previously anticipated, has been termed "flanking fuzziness" (Tompa and Fuxreiter 2008). Furthermore, it was shown recently that a large fraction of transcription factors are characterized by the presence of unstructured regions that flank the DNA-binding domain, at one or both ends, and that these regions may impact the affinity for specific or nonspecific DNA sequences (Gao et al. 2012).

What could the role of this unstructured N-terminal tail of NPM1-C70 be? Although this issue is still under investigation, it is conceivable that the presence of an unstructured segment adjacent to the interacting domain may provide a larger platform for long range electrostatic interactions or even transient physical contacts that facilitate the fine tuning of binding (Tompa and Fuxreiter 2008). This is supported by our observation that two concomitant Lys to Ala substitutions (Lys229-Lys230) in the unstructured segment of NPM1-C70 result in a dramatic decrease of global affinity.

Furthermore, exposed unstructured regions may be modified by post-translational modifications, driving regulatory changes. Several residues in the unstructured segment may be modified, such as Lys229 and Lys230 that are among the lysines acetylated by p300 and deacetylated by SIRT1 (Shandilya et al. 2009). A number of putative phosphorylation sites are also present in the tail, including Ser227, Thr234, Thr237, and Ser242; among these, the phosphorylations of Ser227 by PKC and of Thr234-Thr237 by the Cdk1-cyclinB complex were experimentally validated (Beckmann et al. 1992, Okuwaki et al. 2002). Both acetylation and phosphorylation may

therefore interfere with NPM1 nucleic acid binding and play a role in NPM1 activ- ity and trafficking throughout the cell cycle.

In conclusion, in this work we have undertaken a rigorous analysis of the DNA binding activity of the NPM1 C-terminal domain, which is largely destabilized in Acute Mieloyd Leukemia patients bearing mutations at the *NPM1* gene. We initially focused on defining both the domain boundaries necessary for high affinity recognition and on the DNA sequences and structures that could be recognized efficiently. These studies led to the identification of NPM1 as a novel G-quadruplex binding protein. Then we analyzed the structure of NPM1 C-terminal domain in complex with a representative G-quadruplex and i) identified residues directly involved in the binding, and ii) provided a structural explanation for the higher affinity displayed by NPM1 for G-quadruplex sequences.

Leukemia-associated NPM1 mutations cause dramatic destabilization up to total unfolding of the terminal three-helix bundle that is responsible for the aberrant cytosolic transloca- tion of the protein. Scaloni et al. (Scaloni et al. 2009, Scaloni et al. 2010) showed that a malleable “native-like” structure that accelerates folding is retained in the denatured state of the wild-type NPM1-C53 three-helix bundle and involves helices H2 and H3 (H3 is also the site of leukemic mutations), whereas helix H1 is totally unfolded in the denatured state. A possible strategy to rationally target NPM1 for the treatment of this type of leukemia might be that of developing a drug able to stabilize in the leukemic variant a native-like structure by altering through binding the folding-unfolding equilibrium in favor of the native state. Our folding studies suggested that such a drug should preferentially target helix H1. Because we

have shown here that G-Quadruplex DNA specifically binds a region in between helices H1 and H2 of NPM1- C70, we may attempt to rationally design aptamers or other smaller molecules that by mimicking the binding properties of G-quadruplexes to NPM1 C-terminal domain might stabilize a native-like state in the leukemic variant.

5 Experimental Procedures and Methods

5.1 Cloning and expression of Cter-NPM1 construct: NPM1-C53 and NPM1-C70.

NPM1-C53 (Nucleophosmin ΔN-242) gene was obtained by gene synthesis (GeneART, Ragensburg, Germany) and was cloned into a pET28+(a) vector (Novagen, San Diego, CA) between restriction sites NdeI and BamHI (New England Biolabs, Hitchin, UK).

NPM1-C70 (Nucleophosmin ΔN-225) gene was obtained by gene synthesis (GeneART, Ragensburg, Germany) and was cloned into a pET28+(a) vector (Novagen, San Diego, CA, USA) between restriction sites NdeI and BamHI (New England Biolabs, Hitchin, UK).

Both protein construct were expressed fused to thrombin cleavable His6-tag at N-terminus, using *E.coli* BL21(DE3) strain as expression system.

Expression of NPM1-C53 and NPM1-C70 is similar and provide transformed cells growth A_{600}=0.5 in LB medium supplemented with Kanamycin at 37 °C. At this point 1mM isopropyl 1-thio-β-D-galactopyranoside (Inalco, Milan, Italy) was added and cells were further grown at 20 °C for 16 h under vigorous flask shaking.

5.2 Protein purification.

Also purification protocols of NPM1-C53 and NPM1-C70 are similar and begin with cells harvesting and resuspending in lysis buffer (Hepes 20 mM pH 7.0, 150 mM NaCl, 25 mM Imidazole, 5 mM β-mercaptoethanol). Cells were supplied with a protease-inhibitor cocktail (Roche, Basel, Swiss).

Sonicated cells were centrifuged and supernatants were recovered.

Proteins were purified using a 5ml Ni^{2+} chelating affinity column (GE-Healthcare, Boston, USA) loaded at 1 ml/min and washed with three column volumes with lysis buffer. Elutions were performed with a linear gradient from 0 to 100% elution buffer (Hepes 20 mM pH 7.0, 150 mM NaCl, 1 M Imidazole, 5 mM β-mercaptoethanol) with a flow rate of 2 ml/min. Peak fractions were pooled and fractions containing the protein, as judged from SDS-PAGE, were collected, diluted 10-fold with lysis buffer, and incubated with thrombin .

The fusion proteins were cleaved by thrombin digestion with 1 U enzyme per 2 mg fusion protein, at 4°C for 2h.

After thrombin cleavage the reaction mixture was supplemented with protease inhibitors (Roche Applied Science) and loaded on a Ni^{2+} affinity column, pre-equilibrated with lysis buffer, to remove the thrombin clave N-terminal His tag and uncleaved protein. Protein was recovered from the flow-through, diluted 10-fold with 20mM Hepes, pH 7.0, 5 mM β-mercaptoethanol, and loaded on a SP-Sepharose column pre-equilibrated with the same buffer. Protein was eluted using a NaCl linear gradient, from 0 to 100% elution buffer (Hepes 20 mM pH 7.0, 1 M NaCl, 5 mM β-mercaptoethanol) with a flow rate of 2 ml/min. Peak fractions were pooled and buffer was exchanged to remove the salt, concentrated up to 600mM, with 20 mM Hepes buffer, pH 7.0, 5mM β-mercaptoethanol, by using concentration tubes with cut-off of 3000 Da according to the manufacturer instructions (Millipore, Billerica, MA, USA), and stored at -20 °C.

For isotope enrichment of protein sample, cells were grown in a minimal medium with $(^{15}NH_4)_2SO_4$ and [^{13}C]glucose, using same unlabeled protocol both for expression and purification.

5.3 Site-direct mutagenesis

NPM1-C70 protein variants were obtained through site-directed mutagenesis and using the QuikChange II kit (Stratagene, La Jolla, CA), following the manufacturer's instructions. Forward oligos used for PCR are the following:

K229A,

5'

CATATGCAGGAAAGCTTC**GCG**AAACAGGAAAAAACGCCGAAAAC C- 3';

K230A, 5'-ATATGCAGGAAAGCTTCAAA**GCG**CAGGAAAAAACGCCGAAAACC - 3';

K229A/ K230A, 5'-CATATGCAGGAAAGCTTC**GCGGCG**CAGGAAAAAACGCCGAAAAC CCCGAAAGGCCCG- 3';

K233A, 5'-CAGGAAAAAACGCCG**GCG**ACCCCGAAAGGCCCGAGC-3';

K236A, 5'-CCGAAAACCCCG**GCG**GGCCCGAGCAGCGTG-3';

and

K239A,

5’-
GAAAGCTTCAAAAAACAGGAA**GCG**ACGCCGAAAACCCCGAAAGG
C-3’.

NPM1-C70 protein variants were expressed and purified as the wild-type construct.

5.4 Oligonucleotides

Oligonucleotides used (Fig.3.1*B*) were purchased from PRIMM s.r.l (Milan, Italy) and purified by HPLC. Oligonucleotides used for SPR analysis were biotinylated at the 5’-end. Prior to use, lyophilized oligos were resuspended in the appropriate buffer (20 mM Hepes, pH 7.0, with or without 100mM NaCl or KCl (according to the different experiments), quantified spectrophotometrically and annealed. For annealing, oligos were heated to 95 °C for 15 min and slowly cooled down at RT, overnight.

Oligonucleotides used in NMR experiments studies are Pu27 of sequence 5’-T**GGGG**A**GGG**T**GGGG**A**GGG**T**GGGG**AAGG-3’ and Pu24I of sequence 5’- **TG**A**GGG**T**GGIG**A**GGG**T**GGGG**AAG G-3’. Pu27 and Pu24I were purchased from PRIMM (Milano, Italy) and from IDT (San Jose, CA), respectively, and were both HPLC-purified. Lyophilized oligos were dissolved in Phosphate buffer 20 mM pH 7.0, KCl 100 mM and annealed, as described for other Oligos.

The parallel G-quadruplex assembly of both oligos, after annealing, was assessed by inspecting CD spectra collected with a Jasco J-710 spectropolarimeter.

5.5 Surface Plasmon Resonance

The interactions between biotynilated DNA constructs (ligands) with the purified proteins NPM1-C70 or NPM1-C53 (analytes), were all measured by SPR technique using a Biacore X100 instrument (Biacore, Uppsala, Sweden). Each biotynilated DNA construct was immobilized on a Sensor Chip SA, pre-coated with streptavidin from Biacore AB (Uppsala, Sweden). Capturing procedure on biosensor surface was performed according to manufacturer's instructions and setting the aim for ligand immobilization to 1000 RU. Running buffer was HEPES-buffered saline-EP (HBS-EP), which contains 10 mM HEPES (pH 7.4), 0.15 M NaCl, 3 mM EDTA, 0.005% (v/v) Surfactant P20 (Biacore AB, Uppsala, Sweden). Analytes were dissolved in running buffer and binding experiments were performed at 25°C with a flow rate of 30 ml/minute. The association phase (k_{on}) was followed for 180 s, while the dissociation phase (k_{off}) was followed for 300 s. The complete dissociation of active complex formed was achieved by addition of 10 mM Hepes, 2 M NaCl, 3 mM EDTA, 0.005% (v/v) P20, pH 7.4 for 60 s before each new cycle start. Analytes were tested in a wide range of concentrations, in order to reach at least a two-fold increase from lower concentration tested. When experimental data met quality criteria, kinetic parameters were estimated according to a 1:1 binding model using Biacore X100 Evaluation Software. Conversely, an affinity steady state model was applied to fit the data. In this latter case two possible situations were exploited:

a) single binding site, using the equation y = Rmax*[Analyte] / ([Analyte] + KD)

b) Double independent binding site, using the equation

y = Rmax*[Analyte] /([Analyte] + KD) + Rmax1*[Analyte] /([Analyte] + KD1)

5.6 Circular Dichroism

All circular dichroism experiments were performed using a Jasco J710 instrument (Jasco Inc., Easton, MD) equipped with a Peltier apparatus for temperature control. Static spectra of G10-loop and c-MYC oligos were collected at 25 °C, using oligos annealed in the appropriate buffer and concentrated to 20 µM. Spectra were collected using a quartz cell with 1-mm optical path length (Hellma, Plainview, NY) and a scanning speed of 100 nm/min. The reported spectra are the average of five scans. To measure NPM1-C70 induced G-quadruplex formation, oligos dissolved in buffer (without additional salts) but not annealed, and concentrated to 20 µM, were incubated for 1 h with appropriate amounts of NPM1-C70 and spectra collected as above. To monitor NPM-C70 induced unfolding, oligos were annealed in Hepes buffer supplemented with 100 mM NaCl and incubated with NPM1-C70 increasing amounts. The spectral contribution of buffers and proteins was subtracted as appropriate. To measure the effect of the c-MYC and G10-loop oligos on NPM1-C70 structure, CD spectra of the protein (20µM) in 20mM Hepes, pH 7.0, alone or incubated with pre-annealed oligos (20 µM) were recorded. The spectral contribution of the oligo alone was then recorded and subtracted as appropriate. Thermal denaturation experiments were performed using a quartz cell with 1-mm optical path length and

monitoring the variation of CD signal at 260 nm. Temperature was progressively increased, in 1 °C/min steps, from 25 to 105 °C. The Kaleidagraph software was used for CD spectra analysis and representation.

5.7 Structure calculations of the free protein

The 1H, 13C, and 15N resonance frequencies of NPM1-C70 were assigned using all classical NMR experiments. NMR experiments used for resonance assignment and structure calculations were performed on 13C, 15N labeled NPM1-C70 sample containing 10% D2O. NMR spectra were collected at 298 K, processed using standard Bruker software (TOPSPIN 2.1), and analyzed with CARA (Keller 2004). Structure calculations were performed with the software package UNIO, using as input the amino acid sequence, the chemical shift lists, and three 1H-1H NOE experiments: two dimensional NOESY, three-dimensional 13C-edited NOESY, and three-dimensional 15N-edited NOESY, recorded at 900 MHz with a mixing time of 100 ms. The standard protocol, included in UNIO, with seven cycles of peak picking using ATNOS (Herrmann et al. 2002), NOE assignment with CANDID (Herrmann et al. 2002), and structure calculation with CYANA-2.1 (Gunter 2004) was used. φ and ψ dihedral angles were obtained from the chemical shift analysis using TALOS+ software (Wishart et al. 1994, Eghbalnia et al. 2005, Shen et al. 2009).

In each ATNOS/CANDID cycle, the angle constraints were combined with the updated NOE upper distance constraints in the input for subsequent CYANA-2.1 structure calculation cycle.

The 20 conformers with the lowest target function values were subjected to restrained energy minimization in explicit water with AMBER 11.0 (Bertini et al. 2011, Case et al. 2010). NOE and torsion angle constraints were used. The quality of the structures was evaluated using the programs PROCHECK, PROCHECK NMR (Laskowski et al. 1996) and WHATIF (Vriend 1990). Statistics about the energyminimized family of conformers are reported in Table 3.4. The atomic coordinates and structural restraints for NPM1-C70 have been deposited in the Protein Data Bank with accession code 2llh. Resonance assignments are also available at the BioMagResBank (accession number 18048).

5.8 Structure calculations of the complex

To identify intermolecular NOEs in the NPM1-C70/Pu24I complex, a ω_1-^{13}Cedited, ω_2-^{13}C-filtered NOESY experiment was recorded in a 2D plane (^{1}H-^{1}H plane) on ^{13}C,^{15}N NPM1-C70/unlabelled Pu24I (Zwahlen et al. 1997). The selected temperature was 290 K and the mixing time used was 120 ms. ^{1}H, ^{13}C, ^{15}N backbone resonances of NPM1-C70 in the complex were assigned performing all the typical experiments for backbone assignment.

To calculate a structural model for the interaction between NPM1-C70 and Pu24I, a data-driven molecular docking was performed using the HADDOCK protocol. HADDOCK comprises a series

of Python scripts that run on top of the structure determination programs ARIA and CNS (Dominguez et al. 2003, Brunger et al. 1998, Linge et al. 2001, Linge at al. 2003, de Vries et al. 2010). The method relies on the

definition of Ambiguous Interaction Restraints (AIRs) derived from experimental data. For our docking calculations we defined as unambiguous restraints residues resulting from intermolecular NOEs cross-peaks on both protein and DNA. Furthermore we defined as AIRs: i) residues experiencing chemical shifts variations above the average variations plus one standard deviation (<δHN> = 0.043 ± 0.025 ppm, Black line in Fig. 3.10B), ii) Residues whose signals broadened their linewidth (Glu245, Lys248, Phe268) and/or disappeared during the titration (Phe276) and iii) DNA atoms interacting with the protein, identified also by chemical shift perturbation, as detected in homonuclear experiments (1D and 2D). Residues used as AIRs are listed in Table 3.5. The HADDOCK docking protocol consists of: (i) randomization of orientation and rigid body minimization, (ii) semirigid simulated annealing in torsion angle space, (iii) final refinement in Cartesian space with explicit solvent. The rigid body docking step was performed five times, with 1000 structures generated at each stage, the best 200 of which were refined in the semiflexible stage and subsequently in explicit water. Electrostatic and van der Waals terms were calculated with a 8.5 Å distance cutoff. (Dominguez et al. 2003). For the docking procedure, the structures of NPM1-C70 and Pu24I (PDB code 2A5P) were used as starting points.

Heteronuclear relaxation experiments were performed on ^{15}Nlabeled samples of NPM1-C70 at 700 MHz. The ^{15}N backbone longitudinal (R_1) and transverse (R_2) relaxation rates as well as heteronuclear ^{15}N{^{1}H} NOEs were measured using a standard protocol (Orekhov et al. 1994, Farrow et al.1994).

6 References

Ahmad, Y., Boisvert, F.M., Gregor, P., Cobley, A., Lamond, A.I. (2009) NOPdb: nucleolar proteome database-2008 update. *Nucleic Acids Res.* 37, 181–184.

Ahn, J.Y., Liu, X., Cheng, D., Peng, J., Chan, P.K., Wade, P.A., Ye, K. (2005) Nucleophosmin/B23, a nuclear PI(3,4,5)P(3) receptor, mediates the antiapoptotic actions of NGF by inhibiting CAD. *Mol. Cell* 18, 435–445.

Beckmann R., Buchner K., Jungblut P.R., Eckerskorn C., Weise C., Hilbert R., and Hucho F. (1992) Nuclear substrates of protein kinase C. *Eur. J. Biochem.* 210, 45-51.

Berberich SJ, Postel EH. 1995. PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element. *Oncogene* 10:2343–47

Berger, R., Busson, M., Baranger, L., Hélias, C., Lessard, M., Dastugue, N., Speleman, F. (2006) Loss of the *NPM1* gene in myeloid disorders with chromosome 5 rearrangements. *Leukemia* 20, 319–321.

Bertini I., Case D.A., Ferella L., Giachetti A., and Rosato A. (2011) A Grid-enabled web portal for NMR structure refinement with AMBER. *Bioinformatics* 27, 2384-2390.

Bertwistle, D., Sugimoto, M. & Sherr, C.J. (2004) Physical and functional interactions of the ARF tumor suppressor protein with nucleophosmin/B23. *Mol. Cell. Biol.* 24, 985–996.

Bischof, D., Pulford, K., Mason, D.Y. & Morris, S.W. (1997) Role of the nucleophosmin (NPM) portion of the non-Hodgkin's lymphoma-associated NPM-anaplastic lymphoma kinase fusion protein in oncogenesis. *Mol. Cell. Biol.* 17, 2312–2325.

Bolli, N., De Marco, M.F., Martelli, M.P., Bigerna, B., Pucciarini, A., Rossi, R., Mannucci, R., Manes, N., Pettirossi, V., Pileri, S.A., Nicoletti, I., Falini, B. (2009) A dose-dependent tug of war involving the NPM1 leukaemic mutant, nucleophosmin, and ARF. *Leukemia* 23, 501-509.

Bolli, N., Nicoletti, I., De Marco, M.F., Bigerna, B., Pucciarini, A., Mannucci, R. Martelli, M.P., Liso, A., Mecucci, C., Fabbiano, F., Martelli, M.F., Henderson, B.R., Falini, B. (2007) Born to be exported: COOH-terminal nuclear export signals of different strength ensure cytoplasmic accumulation of nucleophosmin leukemic mutants. *Cancer Res.* 67, 6230–6237.

Bonetti, P., Davoli, T., Sironi, C., Amati, B., Pelicci, P.G., Colombo, E. (2008) Nucleophosmin and its AML-associated mutant regulate c-Myc turnover through Fbw7 gamma. *J. Cell. Biol.* 182, 19-26.

Boon, K., Caron, H.N., van Asperen, R., Valentijn, L., Hermus, M.C., van Sluis, P., Roobeek, I., Weis, I., Voûte, P.A., Schwab, M., Versteeg, R. (2001) N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. *EMBO J.* 20, 1383–93.

Borer, R.A., Lehner, C.F., Eppenberger, H.M., Nigg, E.A. (1989) Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell* 56, 379–390.

Brunger A.T., Adams P.D., Clore G.M., DeLano, W.L., Gros P., Grosse-Kunstleve R. W., Jiang J.S., Kuszewski J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T., and Warren G.L. (1998) *Crystallography & NMR System*: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D Biol. Crystallogr*. 54, 905−921.

Case D.A., Darden T.A., Cheatham T.E. III, Simmerling C.L., Wang J., Duke R.E., Luo R., Walker R.C., Zhang W., Merz K.M., Roberts B., Hayik S., Roitberg A., Seabra G., Swails J., Goetz A.W., Kolossvai I., Wong K.F., Paesani F., Vanicek J., Wolf R.M., Liu J., Wu X., Brozell S.R., Steinbrecher T., Gohlke H., Cai Q., Ye X., Wang J., HsiehM.-J., Cui G., Roe D.R.,

Mathews D.H., Seetin M.G., Salomon-Ferrer R., Sagui C., Babin V., Luchko T., Gusarov S., Kovalenko A., and Kollman P.A. (2010) AMBER 11, University of California, San Francisco.

Chan, W.Y., Liu, Q.R., Borjigin, J., Busch, H., Rennert, O.M., Tease, L.A., Chan, P.K. (1989) Characterization of the cDNA encoding human nucleophosmin and studies of its role in normal and abnormal growth. *Biochemistry* 28, 1033–1039.

Cogoi S., L.E. Xodo, Colombo, E., Martinelli, P., Zamponi, R., Shing, D.C., Bonetti, P., Luzi, L., Volorio, S., Bernard G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription, Nucleic Acids Res. 34 (2006) 2536e2549.,

Colombo E, Alcalay M, Pelicci PG. (2011)Nucleophosmin and its complex network: a possible therapeutic target in hematological diseases. Oncogene.;30(23):2595-609

Cordell, J.L., Pulford, K.A., Bigerna, B., Roncador, G., Banham, A., Colombo, E., Pelicci, P.G., Mason, D.Y., Falini, B. (1999) Detection of normal and chimeric nucleophosmin in human cells. *Blood* 93, 632–642.

Dalenc, F., Drouet, J., Ader, I., Delmas, C., Rochaix, P., Favre, G. et al. (2002) Increased expression of a COOH-truncated nucleophosmin resulting from alternative splicing is associated with cellular resistance to ionizing radiation in HeLa cells. *Int. J. Cancer* 100, 662–668.

De Armond R., S. Wood, D. Sun, L.H. Hurley, S.W. Ebbinghaus,Evidence for the presence of a guanine quadruplex forming regionwithin a polypurine tract of the hypoxia inducible factor 1alpha promoter, Biochemistry 44 (2005) 16341e16350.

de Vries S.J., van Dijk M., and Bonvin A.M.J.J. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols* 5, 883-897.

den Besten, W., Kuo, M.L., Williams, R.T. & Sherr, C.J. (2005) Myeloid leukemia-associated nucleophosmin mutants perturb p53-dependent and independent activities of the Arf tumor suppressor protein. *Cell Cycle* 4, 1593–1598.

Dexheimer T.S., D. Sun, L.H. Hurley, Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter, J. Am. Chem. Soc. 128 (2006) 5404e5415.

Di Fiore, P.P. (2008) Playing both sides: nucleophosmin between tumor suppression and oncogenesis. *J. Cell. Biol.* 182, 7-9.

Dominguez C., Boelens R., and Bonvin A.M.J.J. (2003) HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* 125, 1731–1737.

Dumbar, T.S., Gentry, G.A. & Olson, M.O. (1989) Interaction of nucleolar phosphoprotein B23 with nucleic acids. *Biochemistry* 28, 9495–9501.

Dundr, M., Misteli, T., and Olson, M.O. (2000) *J. Cell. Biol.* 150, 433-446.

Eghbalnia H.R., Wang L., Bahrami A., Assadi A., and Markley J.L. (2005) Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. *J. Biomol. NMR* 32, 71-81.

Facchini L.M., L.Z. Penn, The molecular role of Myc in growth and transformation: recent discoveries lead to new insights, FASEB J. 12 (1998) 633e651.

Falini, B., Bolli, N., Liso, A., Martelli, M.P., Mannucci, R., Pileri, S., Nicoletti, I. (2009) Altered nucleophosmin transport in acute myeloid leukaemia with mutated NPM1: molecular basis and clinical implications. *Leukemia* 23, 1731-1743.

Falini, B., Martelli, M. P., Bolli, N., Bonasso, R., Ghia, E., Pallotta, M. T., Diverio, D., Nicoletti, I., Pacini, R., Tabarrini, A., Galletti, B. V., Mannucci, R., Roti, G., Rosati, R., Specchia, G., Liso, A., Tiacci, E., Alcalay, M., Luzi, L., Volorio, S., Bernard, L., Guarini, A., Amadori, S., Mandelli, F., Pane, F., Lo-Coco, F., Saglio, G., Pelicci, P.G., Martelli, M.F., and Mecucci, C. (2006). Immunohistochemistry predicts nucleophosmin (NPM) mutations in acute myeloid leukemia. *Blood* 108, 1999-2005.

Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., Bigerna, B., Pacini, R., Pucciarini, A., Liso, A., Vignetti, M., Fazi, P., Meani, N., Pettirossi, V., Saglio, G., Mandelli, F., Lo-Coco, F., Pelicci, P. G., and Martelli, M. F. (2005). Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N. Engl. J. Med.* 352, 254-266.

Falini, B., Nicoletti, I., Martelli, M. F., and Mecucci, C. (2007). Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. *Blood* 109, 874-885.

Farrow N.A., Muhandiram R., Singer A.U., Pascal S.M., Kay C.M., Gish G., Shoelson S.E., Pawson T., Forman-Kay J.D., and Kay L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* 33, 5984–6003.

Feuerstein, N., Chan, P. K. & Mond, J. J. (1988) Identification of numatrin, the nuclear matrix protein associated with induction of mitogenesis, as the nucleolar protein B23. Implication for the role of the nucleolus in early transduction of mitogenic signals. *J. Biol. Chem.* 263, 10608–10612.

Frehlick, L.J., Eirin-Lopez, J.M., Ausio, J. (2007) New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Bioessays* 29, 49–59.

Freyer, M. W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L. H., and Lewis, E. A. (2007) *Biophys. J.* 92, 2007–2015

Fry M., Tetraplex DNA and its interacting proteins. Front Biosci. 2007 May 1;12:4336-51.

Gao X., Bulyk M.L., and Hartemink A.J. (2012) Intrinsic disorder within and flanking the DNA-binding domains of human transcription factors. *Pac. Symp. Biocomput.* 104-115.

Gellert, M. et al. (1962) Helix formation by guanylic acid. Proc. Natl.Acad. Sci. U. S. A. 48, 2013–2018.

Gilbert D.E, Feigon J., Multistranded DNA structures, Curr. Opin. Struct. Biol. 9 (1999) 305-314.

González V, Hurley LH,The c-MYC NHE III(1): function and regulation. Annu Rev Pharmacol Toxicol. 2010;50:111-29

Gonzalez V., and Hurley L.H. (2010) The cMYC NHEIII (1): function and regulation. *Annu. Rev. Pharmacol. Toxicol.* 50, 111-129.

Gonzàlez, V., Guo, K., Hurley, L., and Sun, D. (2009) *J. Biol. Chem.* 284,23622–23635

G-quadruplexes. *Chem. Soc. Rev.* 37, 1375–1384

Grand, C. L., Han, H.,Munoz, R. M., Weitman, S., Von Hoff, D. D., Hurley, L. H., and Bearss, D. J. (2002) *Mol. Cancer Ther.* 1, 565 573

Grandori C, Cowley SM, James LP, Eisenman RN. 2000. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu. Rev. Cell Dev. Biol.* 16:653–99

Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K., Pandolfi, P.P. (2005) Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* 437, 147–153.

Grisendi, S., Mecucci, C., Falini, B., and Pandolfi, P. P. (2006) Nucleophosmin and cancer. *Nat. Rev. Cancer* 6, 493-505.

Grummitt, C.G., Townsley, F.M., Johnson, C.M., Warren, A.J., and Bycroft, M. (2008). Structural consequences of nucleophosmin mutations in acute myeloid leukemia. *J. Biol. Chem.* 283, 23326-23332.

Guntert P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* 278, 353–378.

Guo K., A. Pourpak, K. Beetz-Rogers, V. Gokhale, D. Sun, L.H. Hurley, Formation of pseudosymmetrical G-quadruplex and i-motif structures inthe proximal promoter region of the RET oncogene, J. Am. Chem. Soc. 129 (2007) 10220e10228

Herrera, J.E., Correia, J.J., Jones, A.E. & Olson, M.O. (1996) Sedimentation analyses of the salt- and divalent metal ion-induced oligomerization of nucleolar protein B23. *Biochemistry* 35, 2668–2673.

Herrera, J.E., Savkur, R. & Olson, M.O. (1995) The ribonuclease activity of nucleolar protein B23. *Nucleic Acids Res.* 23, 3974–3979.

Herrmann T., Güntert P., and Wüthrich K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 319, 209–227.

Herrmann T., Güntert P., and Wüthrich K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* 24, 171–189.

Hingorani, K., Szebeni, A. & Olson, M.O. (2000) Mapping the functional domains of nucleolar protein B23. *J. Biol. Chem.* 275, 24451–24457.

Horvath M.P., and Schultz S.C. (2001) DNA G-quartets in a 1.86 Å resolution structure of an Oxytricha nova telomeric protein-DNA complex. *J. Mol. Biol.* 310, 367-

Huppert, J. L. (2008) Four-stranded nucleic acids: structure, function and targeting of

Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 35,406–413

Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res., 33, 2908–2916.

Kang, C., X. Zhang, R. Ratliff, R. Moyzis & A. Rich:Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature*, 356, 126-31 (1992)

Keller, R. (2004) The Computer Aided Resonance Assignment Tutorial (CANTINA Verlag, Goldau).

Kettani, A., R. A. Kumar & D. J. Patel: Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J Mol Biol*, 254, 638-56 (1995)

Kurki, S., Peltonen, K., Latonen, L., Kiviharju, T.M., Ojala, P.M., Meek, D., Laiho, M. (2004) Nucleolar protein NPM interacts with HDM2 and protects tumor suppressor protein p53 from HDM2-mediated degradation. *Cancer Cell* 5, 465–475.

Laskowski R.A., Rullmann J.A.C., MacArthur M.W., Kaptein R., and Thornton J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486.

Lee S.Palumbo, R.M. Memmott, D.J. Uribe, Y. Krotova-Khan, L.H. Hurley, S.W. Ebbinghaus, A novel G-quadruplex forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity, Nucleic Acids Res. 36 (2008) 1755e1769.

Lee, H.H., Kim, H.S., Kang, J.Y., Lee, B.I., Ha, J.Y., Yoon, H.J., Lim, S.O., Jung, G., Suh, S.W. (2007) Crystal structure of human nucleophosmin-core reveals plasticity of the pentamer-pentamer interface. *Proteins* 69, 672-678.

Leong, S.M., Tan, B.X., Bte Ahmad, B., Yan, T., Chee, L.Y., Koh, L.P., Yeoh, A.E., Koay, E.S., Mok, Y.K., Lim, T.M. (2010) Mutant nucleophosmin deregulates cell death and myeloid differentiation through excessive caspase-6 and -8 inhibition. *Blood*, epub ahead of print.

Li, J., Sejas, D.P., Burma, S., Chen, D.J., Pang, Q. (2007) Nucleophosmin suppresses oncogene-induced apoptosis and senescence and enhances oncogenic cooperation in cells with genomic intability. *Carcinogenesis* 28, 1163-1170.

Li, J., Zhang, X., Sejas, D. P., Bagby, G.C. & Pang, Q. (2004) Hypoxia-induced nucleophosmin protects cell death through inhibition of p53. *J. Biol. Chem.* 279, 41275–41279.

Linge J. P., O'Donoghue S. I., and Nilges M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol.* 339, 71−90.

Linge J.P., Habeck M., Rieping W., and Nilges M. (2003). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 19, 315-316

Liso, A., Bogliolo, A., Freschi, V., Martelli, M.P., Pileri, S.A., Santodirocco, M., Bolli, N., Martelli, M.F., Falini, B. (2008) In human genome, generation of a nuclear export signal through duplication appears unique to nucleophosmin (NPM1) mutations and is restricted to AML. *Leukemia* 22, 1285–1289.

Moyzis,R.K., Buckingham,J.M., Cram,L.S., Dani,M.,Deaven,L.L., Jones,M.D., Meyne,J., Ratliff,R.L. and Wu,J.R. (1988) A highly conserved repetitive DNA sequence,(TTAGGG)n, present at the telomeres of human chromosomes.Proc. Natl Acad. Sci. USA, 85, 6622–6626.

Nakagawa, M., Kameoka, Y. & Suzuki, R. (2005) Nucleophosmin in acute myelogenous leukemia. *N. Engl. J. Med.* 352, 1819–1820.

Namboodiri, V.M., Akey, I.V., Schmidt-Zachmann, M.S., Head, J.F. & Akey, C.W. (2004) The structure and function of *Xenopus* NO38-core, a histone chaperone in the nucleolus. *Structure* 12, 2149–2160.

Naoe, T., Suzuki, T., Kiyoi, H., Urano, T. (2006) Nucleophosmin: a versatile molecule associated with hematological malignancies. *Cancer Sci.* 97, 963-969.

Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* 19, 239–250

Nishimura, Y., Ohkubo, T., Furuichi, Y. & Umekawa, H. (2002) Tryptophans 286 and 288 in the C-terminal region of protein B23.1 are important for its nucleolar localization. *Biosci. Biotechnol. Biochem.* 66, 2239–2242.

Nozawa, Y., Van Belzen, N., Van der Made, A.C., Dinjens, W.N. & Bosman, F.T. (1996) Expression of nucleophosmin/B23 in normal and neoplastic colorectal mucosa. *J. Pathol.* 178, 48–52.

Okuwaki M., Tsujimoto M., and Nagata K. (2002) The RNA binding activity of a ribosome biogenesis factor, nucleophosmin/B23, is modulated by phosphorylation with a cell cycle-dependent kinase and by association with its subtype. *Mol. Biol. Cell* 13, 2016– 2030.

Okuwaki, M., Matsumoto, K., Tsujimoto, M. & Nagata, K. (2001) Function of nucleophosmin/B23, a nucleolar acidic protein, as a histone chaperone. *FEBS Lett.* 506, 272–276.

Olson, M.O., Wallace, M.O., Herrera, A.H., Marshall-Carlson, L. & Hunt, R.C. (1986) Preribosomal ribonucleoprotein particles are a major component of a nucleolar matrix fraction. *Biochemistry* 25, 484–491

Orekhov V.Y., Pervushin K.V., and Arseniev A.S. (1994) Backbone dynamics of (1-71) bacterioopsin studied by two-dimensional 1H-15N NMR spectroscopy. *Eur. J. Biochem.* 219, 887-896.

Oster SK, Ho CS, Soucie EL, Penn LZ. (2002). The myc oncogene: MarvelouslY Complex. *Adv. Cancer Res.* 84:81–154

Padmanabhan K., Padmanabhan K.P., Ferrara J.D., Sadler J.E., and Tulinsky A. (1993) The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* 268, 17651-17654.

Phan A.T., Kuryavyi V., Gaw H.Y., and Patel D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* 1, 167-173.

Phan, A. T., V. Kuryaviy & D. J. Patel (2006): DNA architecture: from G to Z. Curr. Opin. Struc. Biol., 16, 288-98.

Pruneri L.,, G., Alcalay, M., Pelicci, P.G. (2006) Delocalization and destabilization of the Arf tumor suppressor by the leukemia-associated NPM mutant. *Cancer Res.* 66, 3044–3050.

Qin Y., E.M. Rezler, V. Gokhale, D. Sun, L.H. Hurley, Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4, Nucleic Acids Res. 25 (2007) 7698e7713.

Qin Y., Hurley L., Structures, folding patterns and functions of intramolecular DNA G-Quadruplex foun in eukaryotic promoter regions. Biochimie 90 (2008) 1149–1171.

Raimondi, S.C., Dubé, I.D., Valentine, M.B., Mirro, J. Jr., Watt, H.J., Larson, R.A., Bitter, M.A., Le Beau, M.M., Rowley, J.D. (1989) Clinicopathologic manifestations and breakpoints of the t(3;5) in patients with acute non-lymphocytic leukemia. *Leukemia* 3, 42–47.

Rankin S., A.P. Reszka, J. Huppert, M. Zloh, G.N. Parkinson, A.K. Todd, S. Ladame, S. Balasubramanian, S. Neidle, Putative DNA quadruplex formation within the human c-kit oncogene, J. Am. Chem. Soc. 127 (2005) 10584e10589

Redner, R.L., Rush, E.A., Faas, S., Rudert, W.A. & Corey, S.J. (1996) The t(5;17) variant of acute promyelocytic leukemia expresses a nucleophosmin-retinoic acid receptor fusion. *Blood* 87, 882–886.

Savkur, R.S. & Olson, M.O. (1998) Preferential cleavage in pre-ribosomal RNA byprotein B23 endoribonuclease. *Nucleic Acids Res.* 26, 4508–4515.

Scaloni, F., Federici, L., Brunori, M., and Gianni, S. (2010) *Proc. Natl. Acad. Sci. U.S.A.* 107, 5447–5452

Scaloni, F., Gianni, S., Federici, L., Falini, B., and Brunori, M. (2009) *FASEB J.* 23, 2360–2365

Selenko P., Gregorovic G., Sprangers R., Stier G., Rhani Z., Krämer A., Sattler M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell* 11, 965–976.

Sen D & Gilbert W. Formation of parallel four-stranded complexes by guanine rich motifs in DNA and its implications for meiosis. (1988) Nature, 334, 364-366.

Sen D., Gilbert W., A sodium-potassium switch in the formation of four-stranded G4-DNA, Nature 344 (1990) 410-414.

Shandilya J., Swaminathan V., Gadad S.S., Choudhari R., Kodaganur G.S., and Kundu T.K. (2009) Acetylated NPM1 localizes in the nucleoplasm and regulates transcriptional activation of genes implicated in oral cancer manifestation. *Mol. Cell Biol.* 29, 5115-5127.

Shen Y., Delaglio F., Cornilescu G., and Bax A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44, 213-223.

Shields, L.B., Gerçel-Taylor, C., Yashar, C.M., Wan, T.C., Katsanis, W.A., Spinnato, J.A., Taylor, D.D. (1997) Induction of immune responses to ovarian tumor antigens by multiparity. *J. Soc. Gynecol. Investig.* 4, 298–304.

Siddiqui-Jain A., C.L. Grand, D.J. Bearss, L.H. Hurley, Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 11593e11598

Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. Nucleic Acids Res., 26, 1167–1172.

Sissi C., Gatto B., and Palumbo M. (2011) The evolving world of protein-Gquadruplex recognition: A medicinal chemist's perspective. *Biochimie* 93, 1219-1230.

Smith, F. W. & J. Feigon: Quadruplex structure of*Oxytricha* telomeric DNA oligonucleotides. *Nature*, 356, 164-8 (1992)

Spector, D.L., Ochs, R.L., Busch, H. (1984) Silver staining, immunofluorescence, and immunoelectron microscopic localization of nucleolar phosphoproteins B23 and C23. *Chromosoma* 90,139–148.

Spector, D.L., Ochs, R.L., Busch, H. (1984) Silver staining, immunofluorescence, and immunoelectron microscopic localization of nucleolar phosphoproteins B23 and C23. *Chromosoma* 90,139–148.

Subong, E.N., Shue, M.J., Epstein, J.I., Briggman, J.V., Chan, P.K., Partin, A.W. (1999) Monoclonal antibody to prostate cancer nuclear matrix protein (PRO:4–216) recognizes nucleophosmin/B23. *Prostate* 39, 298–304.

Sun D., K. Guo, J.J. Rusche, L.H. Hurley, Facilitation of a structuraltransition in the polypurine/polypyrimidine tract within the proximalpromoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents, Nucleic Acids Res. 33 (2005) 6070e6080.

Szebeni, A. & Olson, M.O. (1999) Nucleolar protein B23 has molecular chaperone activities. *Protein Sci.* 8, 905–912.

Tanaka, M., Sasaki, H., Kino, I., Sugimura, T. & Terada, M. (1992) Genes preferentially expressed in embryo stomach are predominantly expressed in gastric cancer. *Cancer Res.* 52, 3372–3377.

Tompa P., and Fuxreiter M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33, 2-8.

Tsui, K.H., Cheng, A.J., Chang, P.L., Pan, T.L. & Yung, B.Y. (2004) Association of nucleophosmin/B23 mRNA expression with clinical outcome in patients with bladder carcinoma. *Urology* 64, 839–844.

Vriend G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8, 52–56.

Wang AHJ, Quigley GJ, Kolpak FJ, Crawford JL, van Boom JH, Van der Marel G, Rich A (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. 282 (5740): 680–686.

Wang G, Vasquez KM (2006). Non-B DNA structure-induced genetic instability. Mutat Res 598 (1–2): 103–119.

Wang, D., Baumann, A., Szebeni, A., Olson, M.O. (1994) The nucleic acid binding activity of nucleolar protein B23. 1 resides in its carboxylterminal end. Interaction of nucleolar phosphoprotein B23 with nucleic acids. *J. Biol. Chem.* 269, 30994–30998.

Wang, D., Umekawa, H., Olson, M.O. (1993) Expression and subcellular locations of two forms of nucleolar protein B23 in rat tissues and cells. *Cell. Mol. Biol. Res.* 39, 33–42.

Wang, W., Budhu, A., Forgues, M., Wang, X.W. (2005) Temporal and spatial control of nucleophosmin by the Ran-Crm1 complex in centrosome duplication. *Nat. Cell. Biol.* 7, 823–830.

Wang, Y. & D. J. Patel: Solution structure of a parallel-stranded G-quadruplex DNA. *J Mol Biol*, 234, 1171-83 (1993)

Wang,Y. and Patel,D.J. (1993) Solution structure of the humantelomeric repeat d[AG3(T2AG3)3] G-tetraplex. Structure, 1, 263–282.Nucleic Acids Research, Vol. 40, No. 11 4737

Wei, C., Jia, G., Yuan, J., Feng, Z., and Li, C. (2006) *Biochemistry* 45, 6681–6691

Wei, C., Wang, J., and Zhang, M. (2010) *Biophys. Chem.* 148, 51–55

Wells R.D., Blakesley R.W., Hardies S.C., Horn G.T., Larson J.E., Selsing E., Burd J.F., Chan H.W., Dodgson J.B., Jensen K.F., Nes I.F., Wartell R.M., The role of DNA structure in genetic regulation,CRC Crit. Rev. Biochem. 4 (1977) 305-340

Williamson, J.R. et al. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell 59, 871–880.

Wishart D.S., and Sykes B.D. (1994) The 13C chemical shift index: a simple method for the identification of protein secondary structure using 13C chemical shift data. *J. Biomol. NMR* 4, 171-180.

Wu, M.H., Chang, J.H. & Yung, B.Y. (2002) Resistance to UV-induced cell-killing in nucleophosmin/B23 overexpressed NIH 3T3 fibroblasts: enhancement of DNA repair and up-regulation of PCNA in association with nucleophosmin/B23 over-expression. *Carcinogenesis* 23, 93–100.

Wu, M.H., Chang, J.H., Chou, C.C. & Yung, B.Y. (2002) Involvement of nucleophosmin/B23 in the response of HeLa cells to UV irradiation. *Int. J. Cancer* 97, 297–305.

Wulff J.E., Siegrist R., and Myers A.G. (2007) The natural product avrainvillamide binds to the oncoprotein nucleophosmin. *J. Am. Chem. Soc.* 129: 14444-14451.

Xu, Y., Fang, F., St Clair, W.H., Kasarkis, E.J. and St Clair, D. (2007) *J. Biol. Chem.* 282, 15981-15994.

Yung, B.Y. (2004) c-Myc-mediated expression of nucleophosmin/B23 decreases during retinoic acid-induced differentiation of human leukemia HL-60 cells. *FEBS. Lett.* 578, 211–216.

Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M.(1991) Inhibition of telomerase by G-quartet DNA structures. Nature, 350, 718–720.

Zeller, K.I., Haggerty, T.J., Barrett, J.F., Guo, Q., Wonsey, D.R., Dang, C.V. (2001) Characterization of nucleophosmin (B23) as a Myc target by scanning chromatin immunoprecipitation. *J. Biol. Chem.* 276, 48285–48291.

Zhao,Y., Du,Z. and Li,N. (2007) Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals. FEBS Lett., 581, 1951–1956.

Zor T., Mayr B.M., Dyson H.J., Montminy M.R., and Wright P.E. (2002) Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* 277, 42241–42248.

Zwahlen C., Legault P., Vincent S.J.F., Greenblatt J., Konrat R., and Kay L.E. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* 119, 6711–6721.

7 Appendix

Published OnlineFirst October 6, 2009; DOI:10.1158/0008-5472.CAN-09-1314

Experimental Therapeutics, Molecular Targets, and Chemical Biology

Structural Basis for the Binding of the Anticancer Compound 6-(7-Nitro-2,1,3-Benzoxadiazol-4-Ylthio)Hexanol to Human Glutathione *S*-Transferases

Luca Federici,[1] Carlo Lo Sterzo,[2] Silvia Pezzola,[3] Adele Di Matteo,[2] Flavio Scaloni,[2] Giorgio Federici,[4] and Anna Maria Caccuri[3]

[1]Ce.S.I. Center of Excellence on Aging, "G. D'Annunzio" University Foundation and Department of Biomedical Sciences, University of Chieti, Chieti, Italy; [2]Department of Biochemical Sciences, "Sapienza" University of Rome; [3]Department of Chemical Sciences and Technologies and [4]Departments of Internal Medicine and Laboratory Medicine, Policlinico di Tor Vergata, University of Rome "Tor Vergata", Rome, Italy

Abstract

Glutathione *S*-transferases (GST) constitute a superfamily of enzymes with diversified functions including detoxification from xenobiotics. In many human cancers, Pi class GST (GSTP1-1) is overexpressed and contributes to multidrug resistance by conjugating chemotherapeutics. In addition, GSTP1-1 displays antiapoptotic activity by interacting with c-*Jun* NH_2-terminal kinase, a key regulator of apoptosis. Therefore, GSTP1-1 is considered a promising target for pharmaceutical treatment. Recently, a potent inhibitor of GSTs, 6-(7-nitro-2,1,3-benzoxadiazol-4-ylthio)hexanol (NBDHEX), was identified and tested on several tumor cell lines demonstrating high antiproliferative activity. To establish the structural basis of NBDHEX activity, we determined the crystal structure of NBDHEX bound to either GSTP1-1 or GSTM2-2 (mu class). NBDHEX in both cases binds to the H-site but occupies different positions. Furthermore, the compound is covalently attached to the GSH sulfur in the GSTM2-2 crystal, forming a σ-complex, although it is bound but not conjugated in the GSTP1-1 crystal. Several differences in the H-sites of the two isozymes determine the higher affinity of NBDHEX for GSTM2-2 with respect to GSTP1-1. One such difference is the presence of Ile^{104} in GSTP1-1 close to the bound NBDHEX, whereas the corresponding position is occupied by an alanine in GSTM2-2. Mutation of Ile^{104} into valine is a frequent GSTP1-1 polymorphism and we show here that the Ile^{104}Val and Ile^{104}Ala variants display a 4-fold higher affinity for the compound. Remarkably, the GSTP1-1/Ile^{104}Ala structure in complex with NBDHEX shows a considerable shift of the compound inside the H-site. These data might be useful for the development of new anticancer compounds. [Cancer Res 2009;69(20):8025–34]

Introduction

Multidrug resistance in cancer treatment is commonly defined as the capacity of a cancer cell to develop simultaneous resistance to several chemotherapeutic agents. One common mechanism through which cancer cells may achieve multidrug resistance is the overexpression of glutathione *S*-transferases (GST; refs. 1, 2).

Requests for reprints: Luca Federici, Ce.S.I. Center of Excellence on Aging, University of Chieti "G. D'Annunzio", Via Colle dell'Ara, 66100 Chieti, Italy. Phone: 39-87154-1414; Fax: 39-87154-1598; E-mail: lfederici@unich.it.

doi:10.1158/0008-5472.CAN-09-1314

GSTs constitute a superfamily of enzymes with diversified functions (3, 4). These enzymes are known to decrease the pharmacologic activity of a wide range of structurally unrelated drugs through their conjugation with glutathione (GSH). Furthermore, the GS-conjugated compounds may be actively extruded from the cell through specialized pumps; principally, the multidrug resistance proteins MRP-1 and MRP-2 (ABCC1, ABCC2; ref. 1).

In humans, several GST isoforms are differentially expressed and are subdivided into different classes (5). The Pi class GSTP1-1 is overexpressed in a variety of different human malignancies including lung (6), colon (7), stomach (8), kidney (9), ovary (10), mouth (11), and testis (12) cancers. Furthermore, in some cases, GSTP1-1 overexpression has been linked to acquired multidrug resistance to chemotherapeutic agents including cisplatin, adriamycin, etoposide, thiotepa, chlorambucil, and ethacrynic acid (1, 13). This phenotype may arise from the conjugating activity of the enzyme but also because GSTP1-1 displays a "nonenzymatic" antiapoptotic activity through its interaction with the c-*Jun* NH_2-terminal kinase (JNK), a key enzyme in the apoptotic cascade (14, 15). Because GSTP1-1 binds JNK through its COOH-terminal region, which contributes to shaping the hydrophobic substrate binding site (H-site), inhibitors that target this site, together with inactivating the enzyme, might also dissociate its complex with JNK and activate the JNK apoptotic pathway (16, 17). Therefore, GSTP1-1 is considered as a promising target for inactivation in cancer treatment and numerous groups have spent considerable effort finding potent inhibitors of this enzyme (18, 19).

Recently, a new class of nonpeptidomimetic inhibitors of human GSTs has been identified based on 7-nitro-2,1,3-benzoxadiazole ring derivatives (20). Among these, the compound 6-(7-nitro-2,1,3-benzoxadiazol-4-ylthio)hexanol (NBDHEX) has been shown to inhibit GST isoforms at micromolar or submicromolar amounts and to induce apoptosis in several tumor cell lines by dissociating the GSTP1-1/JNK complex (20, 21). Moreover NBDHEX is able to overcome the P-glycoprotein and MRP1 associated resistance in leukemia (22, 23) and small cell lung cancers (24) and also to overcome the GSTP1-1–related cisplatin resistance in osteosarcoma (25), suggesting that it might be used to treat a number of cancers in combination with other drugs.

A detailed functional analysis of the mode of action of NBDHEX indicated that this compound behaves like a suicide substrate of GSTs through the formation of a σ-complex with GST-bound GSH. In this complex, the GSH sulfur is thought to conjugate the NBDHEX benzoxadiazole ring at its C4, which assumes a tetrahedral conformation by retaining the hexanol

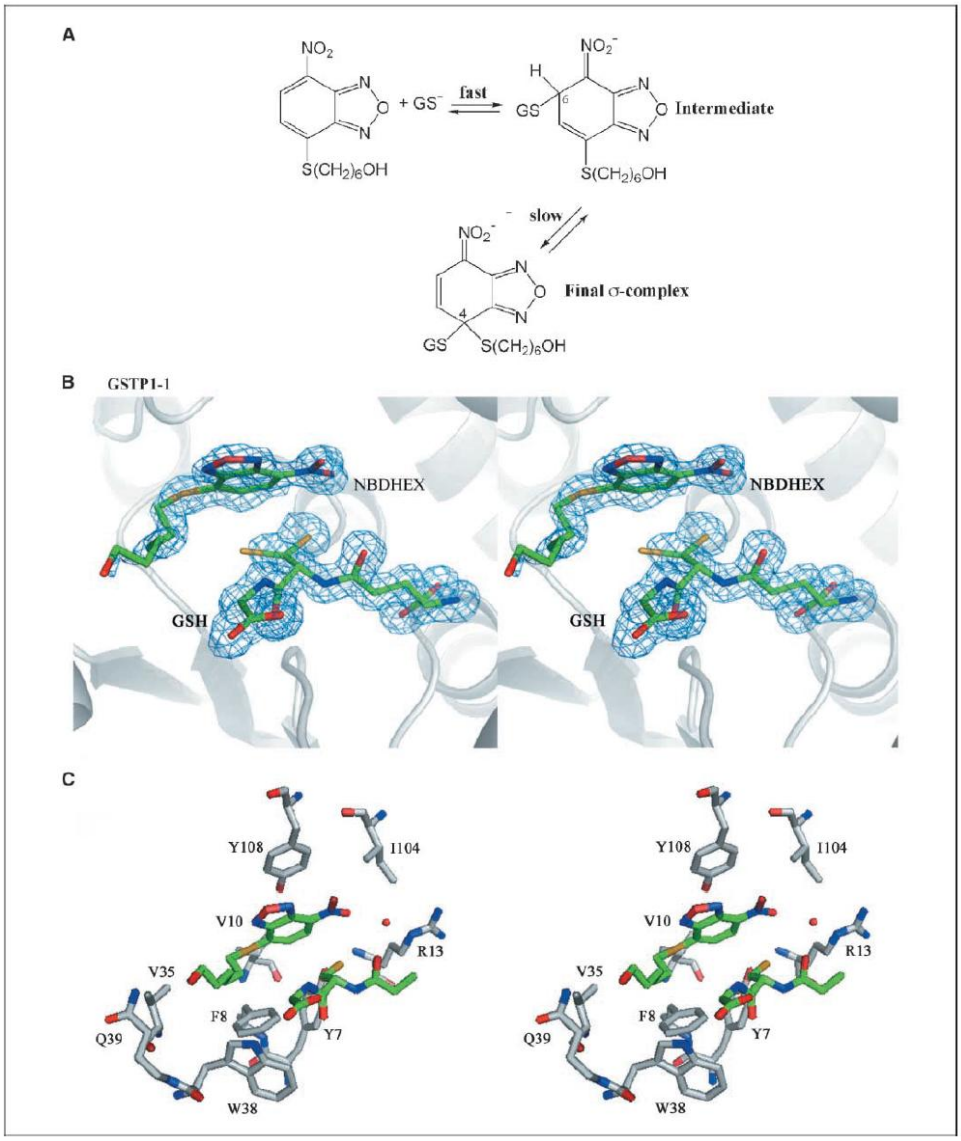

Figure 1. *A*, proposed mechanism of σ-complex formation between NBDHEX and the GST-bound GSH on the basis of kinetics data (20). *B*, crystal structure of the GSTP1-1/GSH/NBDHEX complex. Stereoview of the GSTP1-1 active site (chain A). GSH and NBDHEX (*sticks*); the σ_A-weighted 2Fo-Fc map, contoured at 1.0 σ (*blue mesh*). No continuous electron density was observed between the NBDHEX molecule and the GSH sulfur atom. Furthermore, electron density accounts for a double conformation of the GSH sulfur atom. *C*, detailed view of the active site architecture. NBDHEX and GSH (*sticks with green carbons*). The GSH sulfur atom is represented only in its main conformation (75% of occupancy) for clarity. Residues in the proximity of NBDHEX (*sticks with white carbons*). The water molecule bridging Arg^{13} with the benzoxadiazole nitro group (*red sphere*).

Published OnlineFirst October 6, 2009; DOI:10.1158/0008-5472.CAN-09-1314

moiety (Fig. 1*A*; ref. 20). This intermediate is stable into the GSTP1-1, but also into the mu class GSTM2-2 active sites and thus inactivates both enzymes (20).

One of the issues raised by the use of GST inhibitors is their specificity (18). In principle, because the major role in cancer is played by GSTP1-1, a specific inhibitor for this particular isozyme would be desirable. Unfortunately, NBDHEX seems to be less specific for GSTP1-1 (IC_{50} = 0.80 μmol/L) than for GSTM2-2 (IC_{50} < 0.01 μmol/L; 20). In this work, we investigated how NBDHEX binds to both GSTP1-1 and GSTM2-2, and we present the crystal structures of the GSTP1-1/GSH/NBDHEX and GSTM2-2/GSH/NBDHEX complexes.

NBDHEX binds tightly to the enzymes' H-sites, but occupies different positions due to the molecular architecture of the H-site in the two isozymes. Furthermore, whereas a σ-complex was observed in the GSTM2-2 isozyme, a preconjugation state was observed with GSTP1-1. Finally, we investigated the role played in GSTP1-1 by the common $Ile^{104}Val$ polymorphism (26) and by the $Ile^{104}Ala$ mutation, that mimics the residue occupying this position in GSTM2-2. We determined the structure of the GSTP1-1/$Ile^{104}Ala$ variant in complex with NBDHEX and we show that this mutation is sufficient to cause a considerable shift of the NBDHEX position within the H-site. Our data provides a rational basis for explaining the differential affinity of NBDHEX towards GST isoforms and paves the way for the rational design of new derivatives with increased affinity and specificity for GSTP1-1.

Materials and Methods

Site-directed mutagenesis, expression, and purification. The GSTP1-1/Val^{104} and GSTP1-1/Ala^{104} variants were obtained using the QuickChange kit (Stratagene) and plasmid pGST1 (27). The following forward primers were used:

I104V Fw 5′-CCTCCGCTGCAAATACGTCTCCCTCATCTACAC-3′
I104A Fw 5′-GACCTCCGCTGCAAATACGCCTCCCTCATCTACACCAAC-3′

Human GSTM2-2, GSTP1-1/Ile^{104}, GSTP1-1/Val^{104}, and GSTP1-1/Ala^{104} were expressed in *Escherichia coli* and purified as previously described (28).

Crystallization. GSTP1-1 and GSTM2-2 were concentrated to 10 mg/mL in 10 mmol/L of Hepes (pH 7.0). Experiments were performed after pretreating the protein solution with 5 mmol/L of GSH (Sigma), alone or in combination with NBDHEX (100–500 μmol/L). NBDHEX was synthesized as previously described (20).

GSTM2-2/GSH/NBDHEX cocrystals were obtained by vapor diffusion using the following mother liquor: PEG8000 16% w/v, 100 mmol/L of MES (pH 6.0), and 200 mmol/L of NaCl. Crystals were cryoprotected by transfer into a mother liquor containing the appropriate amounts of GSH and NBDHEX plus 20% ethylene glycol.

GSTP1-1/GSH/NBDHEX cocrystals, grown using PEG8000 as a precipitating agent, were of poor quality and did not diffract satisfactorily. Therefore, GSTP1-1/GSH crystals were grown and then soaked with NBDHEX. GSTP1-1/GSH cocrystals were grown at 4.0°C using 1.8 mol/L of ammonium sulfate as a precipitating agent in 100 mmol/L of MES buffer (pH 6.0). Crystals were then soaked with NBDHEX (100–500 μmol/L). A mother liquor solution containing the appropriate amounts of GSH and NBDHEX plus 20% glycerol was used for cryoprotection. GSTP1-1/Ala^{104} crystals were grown, soaked, and cryoprotected using the same protocols as the wild-type protein.

Data collection and processing. GSTM2-2 complete data were collected to 2.5 Å resolution at the BL14-1 beamline of the BeSSY Synchrotron (Berlin, Germany). The crystals belong to the orthorhombic space group $P2_12_12_1$ and contain two GSTM2-2 dimers in the asymmetric unit.

GSTP1-1 data were collected to 1.55 Å resolution at the ID14-1 beamline of the European Synchrotron Research Facility (ESRF, Grenoble, France). Crystal belongs to the space group C2 and contains one GSTP1-1 dimer in the asymmetric unit. GSTP1-1/Ala^{104} data were collected to 1.8 Å resolution at the ID14-2 beamline of the ESRF Synchrotron. Crystals were isomorphous to wild-type protein crystals. Data were processed with Denzo and scaled with Scalepack (29). Statistics about data collection and processing are reported in Table 1.

Structure determination and refinement. The GSTM2-2/GSH/NBDHEX complex structure was determined by molecular replacement using the program MOLREP (30) and the GSTM2-2 dimer (in the absence of GSH; pdb code:1HNC; ref. 31) as a search model. After initial refinement using the program REFMAC (32), electron density maps were examined using the program COOT (33). Positive electron density in the Fo-Fc map was found in all four subunits, accounting for the presence of GSH. Additional positive electron density was found in the H-site of subunits A and C. This electron density was continuous between the GSH sulfur and the H-site suggesting the presence of a σ-complex between NBDHEX and GSH, which was modeled in chain C with full occupancy and in chain A with half occupancy. The model was refined to 2.5 Å resolution by several stages of restrained refinement in REFMAC, maps inspection, and model adjustments. The final model has an R_{factor} of 21.3% and an R_{free} of 27.3%.

The GSTP1-1/GSH/NBDHEX–soaked crystal was isomorphous with the native GSTP1-1 crystal (pdb code: 6GSS; ref. 34). Therefore, the native GSTP1-1 dimer was used for refinement using REFMAC in the absence of any ligand. Inspection of the Fo-Fc positive difference map indicated the presence of GSH and NBDHEX in both subunits, which were not covalently bound. The structure was refined to 1.55 Å resolution iteratively by using REFMAC and COOT. The final model is refined with an R_{factor} of 16.9% and an R_{free} of 19.5%. The GSTP1-1/Ala^{104} structure in complex with NBDHEX and GSH was obtained using the same procedure as the wild-type protein and refined to a final R_{factor} of 17.9% and an R_{free} of 20.9%.

The quality of all models was assessed using COOT and Procheck (35). Table 1 reports statistics on refinement and model quality. Model coordinates and structure factors have been deposited in the PDB with accession codes: 3GUR (for the GSTM2-2 structure), 3GUS (for GSTP1-1), and 3IE3 (for GSTP1-1/Ala^{104}).

Binding of NBDHEX to GSTs. GST activity was assayed at 25°C as previously reported (36). The affinities of NBDHEX for GSTP1-1/Ile^{104} (wild-type), GSTP1-1/Val^{104}, and GSTP1-1/Ala^{104} were determined in the presence of 1 mmol/L of GSH. The quenching of the intrinsic fluorescence of the protein (excitation at 295 nm and emission at 340 nm) was measured in a single photon counting spectrofluorometer (Perkin-Elmer LS50B) at 25°C after the addition of variable amounts of NBDHEX to 4 μmol/L GST in 0.1 mol/L of potassium phosphate buffer (pH 6.5), containing 1 mmol/L of GSH. Fluorescence data were corrected for inner filter effects and fitted to Eq. A, which yields the apparent dissociation constant (K_D) for the NBDHEX bound to GST,

$$\Delta F = \Delta F_{max}/(1 + K_D/[\text{NBDHEX}]^{nH}) \qquad (A)$$

where ΔF is the protein fluorescence change observed in the presence of a given amount of NBDHEX, ΔF_{max} is the maximum fluorescence change observed at saturating NBDHEX concentration, and nH is the Hill coefficient.

Spectrophotometric analysis. The interaction of NBDHEX with GSTs and GSH was analyzed as previously reported (20). The UV-visible spectrum of NBDHEX (50 μmol/L) in 0.1 mol/L of potassium phosphate buffer (pH 6.5), containing 1 mmol/L of GSH was recorded at 25°C before and after the addition of GSTP1-1/Ile^{104}, GSTP1-1/Val^{104}, or GSTP1-1/Ala^{104} (100 μmol/L).

Results

Structure of the GSTP1-1/GSH/NBDHEX complex. We obtained the crystal structure of the GSTP1-1/GSH/NBDHEX

Table 1. Data collection and refinement statistics

	GSTM2-2	GSTP1-1	GSTP1-1/Ala^{104}
Data collection			
Beamline	BeSSY BL14-1	ESRF ID14-1	ESRF ID14-2
Wavelength (Å)	0.9184	0.9340	0.9330
Resolution (Å)*	29.43–2.5 (2.59–2.5)	30.0–1.55 (1.61–1.55)	29.21–1.80 (1.86–1.80)
Space group	$P2_12_12_1$	C2	C2
Unit cell dimensions (Å)	a = 56.75	a = 78.59	a = 77.87
	b = 78.02	b = 89.35	b = 89.53
	c = 219.36	c = 69.26	c = 68.92
		β = 98.36 degrees	β = 98.04 degrees
Unique reflections	32,108	67,852	42,435
Redundancy	4.8	2.6	4.0
Completeness (%)*	93.2 (83.1)	96.8 (96.5)	99.1 (99.0)
Average (I/σ)*	18.32 (7.01)	9.47 (5.06)	17.38 (8.05)
R_{merge} (%)*,†	8.6 (10.0)	5.3 (26.0)	5.8 (17.8)
Refinement			
Resolution range (Å)	29.43–2.5	30.0–1.55	29.21–1.80
R_{cryst} (%)	21.3	16.9	17.9
R_{free} (%)‡	27.3	19.5	20.9
Mean B factors (Å)^2			
Protein main chain	26.0	12.1	13.0
Protein side chain	26.4	13.5	14.0
Waters	20.6	26.7	22.5
GSH	25.1 (chains B and D)	13.3	13.4
NBDHEX		34.6	32.8
NBDHEX-GS σ-complex	34.1 (chains A and C)		
Rms deviation length (Å)	0.009	0.008	0.007
Rms deviation angle (degrees)	1.45	1.49	1.37
Ramachandran statistics			
Residues in favored regions (%)	91.5	93.3	93.0
Residues in allowed regions (%)	8.5	6.7	7.0
Residues in disallowed regions (%)	0.0	0.0	0.0

*Numbers in parentheses refer to the last resolution shell.
† $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i - \langle I\rangle|/|\langle I\rangle|$, with I_i being the intensity for the ith measurement of an equivalent reflection with indices h,k,l.
‡ R_{free} was calculated on 5% of data excluded before refinement.

complex at 1.55 Å resolution from a crystal that was grown in the presence of protein and GSH only, and subsequently soaked with NBDHEX, because cocrystallization experiments with both NBDHEX and GSH did not lead to well-diffracting crystals. Figure 1*B* shows NBDHEX bound to the protein H-site in proximity to the bound GSH. The crystal asymmetric unit contains a physiologic GSTP1-1 dimer and clear density for NBDHEX was observed in both monomers. The NBDHEX molecule was clearly visible in the electron density, including most of the 6-mercapto-hexanol moiety atoms (Fig. 1*B*). The two NBDHEX benzoxadiazole rings occupy the same position in the H-sites of both chains with rms deviations between equivalent atoms of <0.15 Å, whereas the hexanol moiety atoms have a rms deviation of 0.62 Å.

NBDHEX is oriented in the H-site with the nitro group placed deep inside the interior of the protein and the hexanol moiety facing the active site opening to the solvent (Fig. 1*B*). Several hydrophobic interactions were found involving residues shaping the H-site including Tyr^7, Phe^8, Val^{35}, Trp^{38}, Gln^{39}, Ile^{104}, and above all, Tyr^{108}, which engages an aromatic ring stacking interaction with the NBDHEX benzoxadiazole ring (Table 2; Fig. 1*C*). The main contribution to the molecule stabilization and orientation in the H-site is likely provided by interaction of the benzoxadiazole ring NO_2 group with Arg^{13}. This interaction was not direct but was mediated by a water molecule (*red sphere*, Fig. 1*C*). This water molecule is found in exactly the same position in both monomers and its structural role is confirmed by an average B-factor of 13.05 Å^2, a value in the range of main chain protein atoms' B-factors. The water molecule is hydrogen-bonded both to Arg^{13} Nε, being at a distance of 2.9 Å, and to the benzoxadiazole NO_2 group, being at a distance of 2,6 Å from the closest oxygen.

In our maps, we did not find any trace of continuous electron density connecting the GSH sulfur and NBDHEX. This finding prompted us to exclude the presence of a GSH/NBDHEX σ-complex in this crystal (Fig. 1*B*). Therefore, in our structure, we observe a bound but not conjugated NBDHEX. This was quite surprising because the formation of a GSH/NBDHEX σ-complex in the GSTP1-1 active site can be easily monitored in solution by absorption spectra (20). However, it was previously suggested by kinetic studies that, in the reaction between GSH and either CDNB (37) or 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl; ref. 38), a physical event, i.e., a conformational change in GSTP1-1 structure, follows the binding of the substrates and is crucial for the

σ-complex formation. This conformational transition is strongly dependent on diffusion-controlled motion of active site regions. Because NBDHEX retains the same benzoxadiazole ring as NBD-Cl, a similar conformational change might also be necessary in this case to enable the σ-complex formation, and this might be prevented by a lack of structural flexibility in the crystalline environment.

Nonetheless, the position of NBDHEX with respect to GSH is very interesting. In fact, it has been hypothesized, on the basis of kinetics data, that the formation of the σ-complex, in which the GSH sulfur is bound to the benzoxadiazole C4 (the same carbon that binds the 6-mercapto-hexanol moiety sulfur) is the rate-limiting step of a reaction that first involves the nucleophilic addition of the GSH sulfur to the C6 atom of the benzoxadiazole ring; this covalent complex then evolves to the more stable σ-complex (Fig. 1*A*; ref. 20). Indeed, in our structure, we observe that the GSH sulfur adopts two distinct conformations (Fig. 1*B*). In the main one, which accounts for 75% of total occupancy, the GSH sulfur is at 3.1 Å distance from the NBDHEX C6, whereas it is at 4.9 Å distance from the C4. In the second conformation, which accounts for the remaining 25% of occupancy, the GSH sulfur is closer to the NBDHEX C4 (3.1 Å). Therefore, we can conclude that (*a*) our electron density maps seem to be consistent with the hypothesized mechanism and depict the position of a preconjugation state and (*b*) the formation of the stable σ-complex may be described without invoking major movements of the NBDHEX benzoxadiazole ring from its current position. In fact, a simple rotation of the GSH sulfur is consistent with its nucleophilic attack to the benzoxadiazole C4.

Structure of the GSTM2-2/GSH/NBDHEX complex. The structure of the GSTM2-2/GSH/NBDHEX complex was determined at 2.50 Å resolution. In this case, additional electron density accountable for the presence of the compound was visible only in one monomer for each asymmetric unit dimer (chains A and C). As expected, NBDHEX binds to the H-site. Electron density maps clearly indicate the position of the benzoxadiazole ring, whereas the hexanol moiety is only partially visible (Fig. 2*A*). The two NBDHEX benzoxadiazole rings modeled in chains A and C occupy exactly the same position with rms deviations among equivalent atoms <0.20 Å. NBDHEX is oriented in the same way as in the complex with GSTP1-1, with the benzoxadiazole NO_2 facing the interior of the protein and the hexanol moiety facing the H-site opening. A stacking interaction between the benzoxadiazole ring and Tyr^{115} strongly contributes to the NBDHEX orientation and stabilization (Fig. 2*B*). Several other hydrophobic interactions involve residues shaping the H-site, including Tyr^{6}, Trp^{7}, Gly^{11}, and Leu^{12} (Table 2). As in the GSTP1-1 cocrystal, the main contribution to NBDHEX stabilization is provided by interactions involving the benzoxadiazole NO_2 group. In this structure in fact, the NO_2 group is at close distance from two arginines, Arg^{107} and Arg^{165} (Fig. 2*B*). In particular, the Arg^{165} NH1 atom is at 3.0 Å distance from the closest NO_2 oxygen, establishing a direct H-bond interaction, whereas the same NO_2 oxygen atom is at 3.4 Å distance from the Arg^{107} NH1 atom. Therefore, both residues cooperate to stabilize the NO_2 group that is negatively charged when the σ-complex is formed.

Importantly, in this crystal, we could see continuous electron density spanning from the GSH sulfur to the C4 carbon of the benzoxadiazole ring (Fig. 2*A*), indicating the presence of a σ-complex and thus providing an independent experimental clue to the proposed mechanism of inhibition (20).

Comparison between the two complexes. The electrostatic potential surfaces of GSTP1-1 and GSTM2-2 are shown in Fig. 3*A* and *B*, respectively. The NBDHEX molecule perfectly adapts to the GSTP1-1 H-site profile, with its benzoxadiazole ring being parallel to the active site "ceiling" that is made by Tyr^{108}. The hexanol moiety also follows the active site shape because of its hydrophobic interactions with Phe^{8} and Trp^{38} and of a weak hydrogen bond (3.3 Å distance) between Gln^{39} and the hexanol oxydryl group (Fig. 1*C*). The H-site in the GSTM2-2 structure appears differently shaped from that of GSTP1-1, with a deeper cavity toward the core of the protein, which is occupied by the NBDHEX nitro group (Fig. 3*B*). This different H-site shape mainly results from the presence of residues Ile^{104} and Arg^{13} in the GSTP1-1 H-site, that face the NBDHEX nitro group in the complex (Fig. 1*C*). GSTP1-1 Ile^{104} and Arg^{13} are topologically replaced by an alanine (Ala^{111}) and a

Table 2. Residues within 4.5 Å from NBDHEX

Residue	Distance*	Interaction type
GSTM-2		
Tyr^{6}	2.5	Arom
Trp^{7}	4.3	Arom
Ile^{9}	4.4	Phob
Gly^{11}	3.4	—
Leu^{12}	3.5	H-Phob
Met^{34}	3.9	Phob
Arg^{42}	4.5	H-Phob
Arg^{107}	3.4	HB
Ala^{111}	3.6	H-Phob
Tyr^{115}	3.6	Arom
Arg^{165}	3.0	HB
Phe^{208}	3.4	Arom
Thr^{209}	3.1	Phob
Met^{211}	4.4	Phob
GSTP1-1		
Tyr^{7}	3.9	Arom
Phe^{8}	3.6	Arom
Arg^{13}	4.4	HB (water-mediated)
Val^{35}	4.0	Phob
Trp^{38}	3.6	Phob
Gln^{39}	3.3	HB
Ile^{104}	3.4	H-Phob
Tyr^{108}	3.3	Arom
GSTP1-1/Ala^{104}		
Tyr^{7}	4.1	Arom
Phe^{8}	3.7	Phob
Arg^{13}	2.9	HB (direct)
Val^{35}	4.4	Phob
Trp^{38}	4.3	Phob
Gln^{39}	3.7	HB
Ala^{104}	3.5	—
Tyr^{108}	3.2	Arom
Gly^{205}	3.9	H-Phob

NOTE: Interaction types: HB, hydrophilic-hydrophilic (H-bond); Arom, aromatic-aromatic; Phob, hydrophobic-hydrophobic; H-Phob, hydrophilic-hydrophobic.
*Distances are the average of those relative to the different monomers in the same structure and were calculated using the program LPC/CSU (http://ligin.weizmann.ac.il/cgi-bin/lpccsu/LpcCsu.cgi).

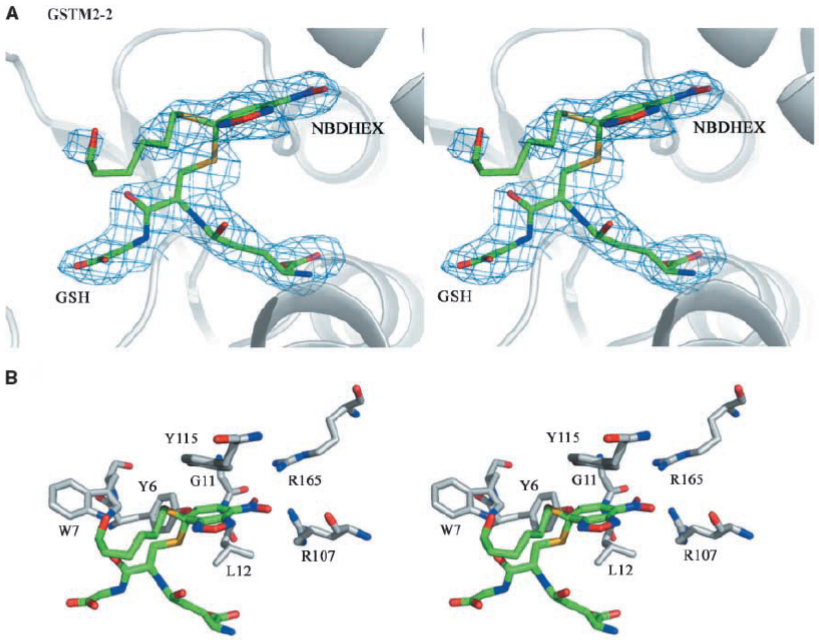

Figure 2. Crystal structure of the GSTM2-2/GSH/NBDHEX complex. *A*, stereoview of the GSTM2-2 active site (chain C). The σ_A-weighted 2Fo-Fc map, contoured at 1.0 σ (*blue mesh*); GSH and NBDHEX (*sticks*). A σ-complex is observed in this case in the electron density map, with the GSH sulfur atom covalently bound to the benzoxadiazole ring C4, which assumes a tetrahedral conformation. *B*, detailed view of the active site architecture. The GSH/NBDHEX σ-complex (*sticks with green carbons*). Residues facing the active site (*sticks with white carbons*). Two positively charged residues (*R107* and *R165*) are at binding distance from the negatively charged nitro group of NBDHEX.

leucine (Leu^{12}), respectively, in the GSTM2-2 structure (Fig. 3*C*), and the presence here of these less cumbersome residues allows the NBDHEX molecule to move further and for its nitro group to interact directly with the positively charged Arg^{107} and Arg^{165} (Figs. 2*B* and 3*C*). Arg^{107} is conserved in GSTP1-1 (Arg^{100}) but its access to the H-site is hindered by Ile^{104}, whereas the GSTM2-2 Arg^{165} is topologically replaced by an isoleucine (Ile^{161}) in GSTP1-1 (Fig. 3*C*).

The superposition between the two structures also highlights that the NBDHEX molecules do not occupy the same position in the two H-sites (Fig. 3*C*). The rms deviation among equivalent atoms of the two NBDHEX benzoxadiazole rings is 4.25 Å in the superimposed structures, and a major shift plus a rotation are required to superpose the free NBDHEX of GSTP1-1 to the one covalently bound to GSH in the GSTM2-2 σ-complex (Fig. 3*C*). This movement, however, is unrealistic because the hydrophobic side chain of Ile^{104} is only 3.4 Å apart from the benzoxadiazole nitro group in the GSTP1-1 complex and provides steric hindrance (Figs. 1*B* and 3*C*). Taken together, our data suggests that the GSH/NBDHEX σ-complex in the GSTP1-1 structure would occupy a different position from the one experimentally observed in the GSTM2-2 cocrystal. This is confirmed by the observation that a simple rotation of the GSH cysteine side chain in GSTP1-1 is sufficient to place the sulfur atom in a position competent for addition to the benzoxadiazole C4 (Fig. 1*B*).

A $Ile^{104}Val$ allelic variation of GSTP1-1 increases the isozyme's affinity for NBDHEX. Residue Ile^{104} of GSTP1-1 faces the benzoxadiazole nitro group and constitutes a physical obstacle to its movement towards the positive charges at the H-site bottom (Figs. 1*C* and 3*C*). This finding immediately captured our attention because mutation of Ile^{104} into valine is a common polymorphism of human GSTP1-1 and is frequent in tumors (26). To better assess the role of Ile^{104} with respect to NBDHEX binding and activity, we mutated it into valine (GSTP1-1/Val) and alanine, which is the residue present at this position in GSTM2-2 (GSTP1-1/Ala; Fig. 3*C*).

The specific activities for CDNB of the GSTP1-1/Val and GSTP1-1/Ala mutant enzymes were 58 $\pm$ 2 and 47 $\pm$ 3 μmol/min/mg, respectively. These values are approximately twice lower than those obtained with wild-type GSTP1-1/Ile^{104} (100 $\pm$ 2 μmol/min/mg; ref. 39). Binding of NBDHEX, in the presence of 1 mmol/L of GSH, followed a hyperbolic behavior (Hill coefficient *nH* 1.0) with GSTP1-1/Ala^{104} and a slight cooperative trend (Hill coefficient *nH* 0.8) with GSTP1-1/Val^{104}.

Remarkably, NBDHEX bound to GSTP1-1/Val^{104} with a dissociation constant of 0.26 $\pm$ 0.07 μmol/L, a value comparable to that of

GSTP1-1/Ala^{104} (0.21 ± 0.06 μmol/L) and approximately four times lower than that of GSTP1-1/Ile^{104} (K_D = 0.90 ± 0.08 μmol/L; ref. 20; Fig. 4*A*). This indicates that replacement of Ile^{104} with a less bulky residue increases the affinity of the enzyme towards NBDHEX.

As previously shown with GSTP1-1/Ile^{104} (20), a remarkable change in the NBDHEX spectrum was observed in the presence of both Val and Ala GSTP1-1 variants. When NBDHEX was incubated with two equivalents of GSTP1-1/Val^{104} or GSTP1-1/Ala^{104} in the presence of 1 mmol/L of GSH, the UV-visible spectrum of NBDHEX, centered at 432 nm, completely disappeared and a new absorption band appeared between 340 and 350 nm (Fig. 4*B*). No spectral changes were observed in the absence of GSH. The extinction coefficient was approximately the same for the Ala^{104}, Val^{104}, and Ile^{104} GSTP1-1 variants, suggesting that NBDHEX forms a stable σ-complex with GSH in all cases. Interestingly, a blue-shift of ~6 nm was observed for the maximum absorption in the GSTP1-1/Ala^{104} spectrum (Fig. 4*B*). This led us to hypothesize that the compound might be slightly shifted in this mutant isozyme with respect to the other two natural variants. To test this hypothesis, we determined the structure of the GSTP1-Ala^{104}/GSH/NBDHEX complex to a resolution of 1.8 Å (Fig. 4*C*). As for the wild-type, we observe a Michaelis complex with the compound bound but not conjugated and a single conformation is found for the GSH sulfur that corresponds to the main conformation observed in the wild-type crystal. The main aspect of this structure is that, as hypothesized, NBDHEX is shifted towards the interior of the site and its nitro group is at hydrogen bond distance (2.9 Å) from the Arg^{13} Nε, whereas the water molecule that bridged the NBDHEX nitro group and Arg^{13} in the wild-type structure, is displaced. A superposition of the GSTP1-1/Ile^{104} and GSTP1-1/Ala^{104} structures (Fig. 4*D*) highlights the different position occupied by the two NBDHEX-bound molecules with an average distance of 1.87 Å between the same atoms of the two benzoxadiazole rings, whereas the rms deviation among Cα in the two superimposed structures is only 0.14 Å.

Discussion

In this work, we tackled the problem of specificity in GST inhibition and focused our attention on NBDHEX, a compound that is currently being investigated for the treatment of several cancers (21–25). Our major goal was to understand the structural basis of the different affinities of NBDHEX for human GST isozymes and gain the knowledge to rationally design molecules that are specific for GSTP1-1. For this purpose, we determined the crystal structure of NBDHEX in complex with GSTP1-1 but also

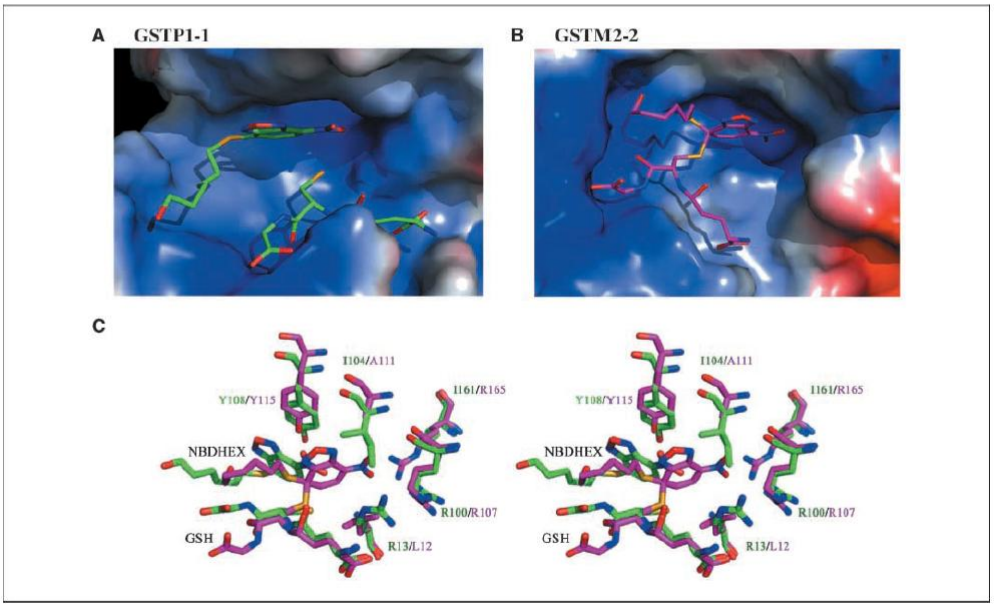

Figure 3. Comparison between the GSTP1-1 and GSTM2-2 active sites. *A,* electrostatic potential surface representation of the GSTP1-1 active site. Positive charge distribution (*blue*), negative charge (*red*), and hydrophobic surface (*white*). NBDHEX and GSH (*sticks with green carbons*). NBDHEX adapts to the shape of the GSTP1-1 H-site ceiling with the benzoxadiazole ring nitro group facing a hydrophobic surface area contributed by GSTP1-1 Ile^{104}. *B,* electrostatic surface representation of the GSTM2-2 active site. Surface color code is the same as in Fig. 4*A*, the GSH/NBDHEX σ-complex (*sticks with magenta carbons*). The H-site in this case seems larger, with a deeper cavity towards the core of the protein that is occupied by the NBDHEX nitro group. *C,* overlay of the GSTP1-1 and GSTM2-2 structures. GSTP1-1 residues (*sticks with green carbons*), GSTM2-2 residues with magenta carbons. The NBDHEX molecule occupies different positions in the two active sites. This is the result of the presence, in the GSTP1-1 H-site, of Ile^{104} and Arg^{13}, which gate the site and are topologically replaced by Ala^{111} and Leu^{12} in the GSTM2-2 active site, respectively.

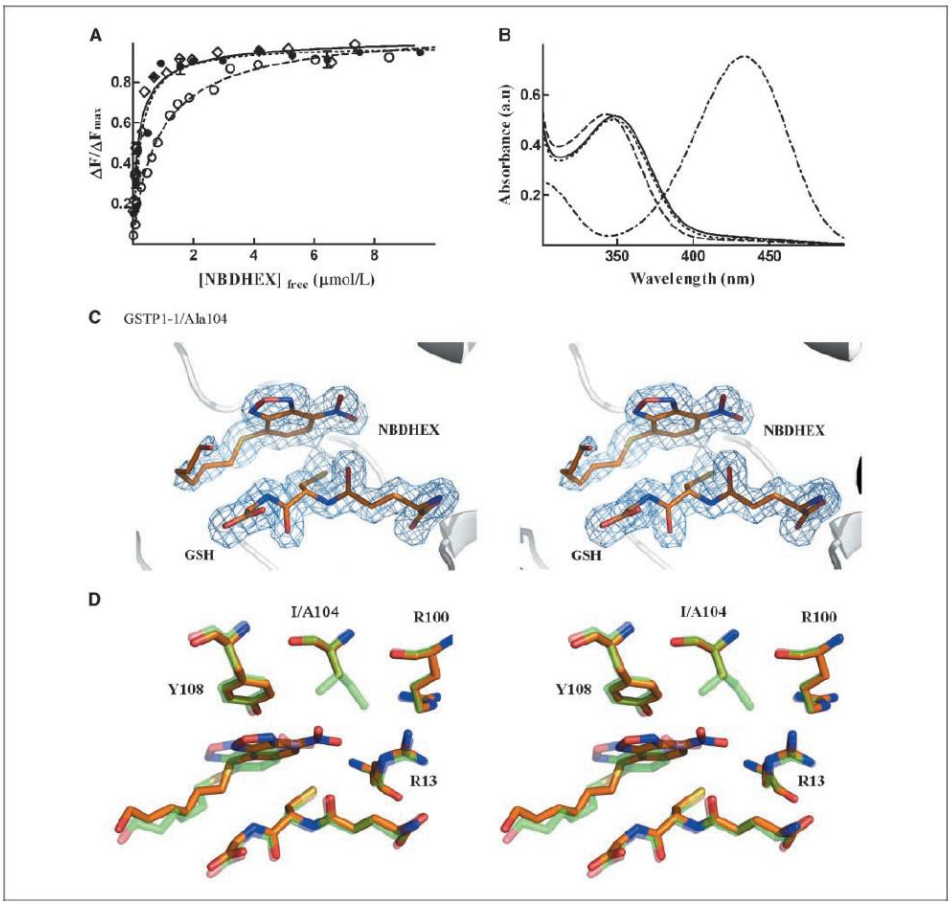

Figure 4. Analysis of the reaction of NBDHEX with GSTP1-1/Ile^{104}, GSTP1-1/Val^{104}, and GSTP1-1/Ala^{104}. *A*, the binding of NBDHEX was studied at 25°C and in the presence of 1 mmol/L of GSH, by following the quenching of the intrinsic fluorescence of the protein. Binding of NBDHEX to GSTP1-1/Val^{104} (—◇—), GSTP1-1/Ala^{104} (· · ●· · ·), and GSTP1-1/Ile^{104} (- - -○- - -). *Points*, mean; *bars*, SD ($n = 4$); *lines*, the best fit of experimental data to Eq. A. *B*, the UV-visible spectrum of NBDHEX (50 μmol/L) in 0.1 mol/L of potassium phosphate buffer (pH 6.5), containing 1 mmol/L of GSH was recorded at 25°C before (- · -) and after the addition of two equivalents of either GSTP1-1/Ile^{104} (—), GSTP1-1/Val^{104} (· · ·), or GSTP1-1/Ala^{104} (- - -). The σ-complex absorption band formed by GSTP1-1/Ala^{104} was centered at 344 nm, whereas that formed by either GSTP1-1/Val^{104} or Ile^{104} (wild-type) variants showed a peak at 350 nm (a representative experiment out of three performed with similar results). *C*, crystal structure of the GSTP1-1/Ala^{104} variant in complex with GSH and NBDHEX. Stereoview of the protein active site. The σ_A-weighted 2Fo-Fc map, contoured at 1.0 σ (*blue mesh*), and GSH and NBDHEX (*sticks*). *D*, overlay of the structures of GSTP1-1 (*green carbons*) and of GSTP1-1/Ala^{104} variant (*orange carbons*). The NBDHEX molecule in the GSTP1-1/Ala^{104} structure is shifted by ~2 Å with respect to the wild-type structure and establishes a direct H-bond interaction with Arg^{13}.

with GSTM2-2, which is the human GST isozyme that displays the higher affinity for the compound (20).

In the case of GSTM2-2, we observe the presence of a σ-complex and thus provide a structural clue to the hypothesized mechanism of inhibition (see Fig. 1*A*; ref. 20). On the other hand, with GSTP1-1, we see a preconjugation state in which the NBDHEX molecule is bound to the enzyme but not conjugated. The reason for this difference might be due to the lack of diffusion-controlled motion of the active site region caused by the crystalline environment. However, we propose that in the GSTP1-1 structure, the σ-complex would occupy the same position of the unbound NBDHEX. This is suggested both by the shape of the H-site and by observing that a simple rotation of the GSH sulfur would be consistent with its addition to the benzoxadiazole ring C4 (see Fig. 1*B*).

A comparative analysis of GSTP1-1 and GSTM2-2 structures indicates that NBDHEX binds the two enzymes in a similar orientation and is stabilized through stacking interactions with aromatic side chains (Tyr^{108} for GSTP1-1 and Tyr^{115} for GSTM2-2). Aromatic stacking interactions with residues in the H-site are also the main contributors to the stabilization of drugs like ethacrynic acid and chlorambucil (40–42). However, unlike what happens with ethacrynic acid and chlorambucil binding to GSTP1-1, in which their polar groups point out of the H-site, in the case of NBDHEX, the benzoxadiazole nitro group constitutes the main driving force to dictate and stabilize binding to both GSTP1-1 and GSTM2-2, and is likely responsible for the higher affinity towards this compound (20, 41, 43). In the case of GSTM2-2, NBDHEX interacts directly with two positively charged arginines located at the bottom of the H-site. These residues provide stabilization when the σ-complex is formed and the nitro group acquires a net negative charge (see Fig. 2*B*). Conversely, in the GSTP1-1 structure, Arg^{13} only provides stabilization but indirectly via a water-mediated contact (see Fig. 1*C*). Of the two GSTM2-2 arginines, only one is topologically conserved in GSTP1-1 (Arg^{100}) but it is too distant from NBDHEX to help stabilize the molecule (see Fig. 3*C*). This is a consequence of the fact that Ile^{104}, a residue that is topologically replaced by an alanine in GSTM2-2, is located between the NBDHEX nitro group and Arg^{100}, and impedes further NBDHEX movement towards this positively charged residue. If this does not prevent the σ-complex from forming, as it is monitored by absorption spectra in solution, it certainly plays a role in diminishing the affinity of the compound for GSTP1-1 (see Fig. 3*C*). These considerations are confirmed by our mutagenesis data on residue Ile^{104}. We show here that replacement of the bulky isoleucine with valine (natural allelic variant) or alanine (the residue that replaces Ile^{104} in GSTM2-2) increases the affinity of GSTP1-1 for NBDHEX by 4-fold. This result may be clinically relevant because it implies that NBDHEX would be equally effective or better against those tumors in which the allelic variant GSTP1-1/Val^{104} is expressed. Moreover, when Ile^{104} is replaced by an alanine, not only does NBDHEX affinity for GSTP1-1 increase, but a blue-shift in the maximum absorption of the σ-complex spectrum is also observed, suggesting that in this mutant, NBDHEX finds a way to move further toward the positive charges and get higher stabilization. We confirmed this hypothesis by determining the structure of the GSTP1-1/Ala^{104} variant in complex with NBDHEX, which showed that the compound is shifted by ~2 Å towards Arg^{13}, displacing the water molecule and establishing a direct H-bond interaction between the benzoxadiazole ring nitro group and the Arg^{13} guanidine group (see Fig. 4*D*).

Can we infer from these data how to modify NBDHEX to inhibit GSTP1-1 with higher affinity? A possible route would be that of synthesizing molecules that retain the benzoxadiazole ring, to take advantage of the aromatic stacking interaction with Tyr^{108} and of the intrinsic stability of its σ-complex within the H-site. At the same time, a newly synthesized inhibitor should interact directly or more strongly with the positively charged Arg^{13}. This might be achieved by substituting the nitro group with other functional groups. Furthermore, because we showed that the NBDHEX hexanol moiety is only weakly involved in protein binding, a concurrent strategy might be that of modifying this part of the molecule with a different leaving group that has the potential to interact strongly with H-site residues. Future work in our laboratories will be aimed at testing several hypotheses suggested by the structural framework provided here.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 4/8/09; revised 7/24/09; accepted 8/10/09; published OnlineFirst 10/6/09.

Grant support: BeSSYII-ELISA grant agreement no. 226716, the Italian Ministero dell'Università e della Ricerca (MIUR), Alleanza Contro il Cancro-ISS, and MIUR, FIRB RBLA03B3KC_004 (A. Di Matteo).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank the beamline scientists at the European Synchrotron Research Facility (Grenoble, France) and at the "Helmholtz Zentrum Berlin fur Materialen und Energie-BeSSYII" (Berlin, Germany) for beamtime allocation and support.

References

1. Meijerman I, Beijnen JH, Schellens JH. Combined action and regulation of phase II enzymes and multidrug resistance proteins in multidrug resistance in cancer. Cancer Treat Rev 2008;34:505–20.
2. Townsend DM, Tew KD. The role of glutathione-S-transferase in anti-cancer drug resistance. Oncogene 2003;22:7369–75.
3. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. Annu Rev Pharmacol Toxicol 2005;45:51–88.
4. Oakley AJ. Glutathione transferases: new functions. Curr Opin Struct Biol 2005;15:716–23.
5. Sheehan D, Meade G, Foley VM, Dowd CA. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochem J 2001; 360:1–16.
6. Inoue T, Ishida T, Sugio K, Maehara Y, Sugimachi K. Glutathione S transferase Pi is a powerful indicator in chemotherapy of human lung squamous-cell carcinoma. Respiration 1995;62:223–27.
7. Ruiz-Gomez MJ, Souviron A, Martinez-Morillo M, Gil L. P-glycoprotein, glutathione and glutathione S-transferase increase in a colon carcinoma cell line by colchicine. J Physiol Biochem 2000;56:307–12.
8. Okuyama T, Maehara Y, Endo K, et al. Expression of glutathione S-transferase-pi and sensitivity of human gastric cancer cells to cisplatin. Cancer 1994;74:1230–36.
9. Grignon DJ, Abdel-Malak M, Mertens WC, Sakr WA, Shepherd RR. Glutathione S-transferase expression in renal cell carcinoma: a new marker of differentiation. Mod Pathol 1994;7:186–9.
10. Green JA, Robertson IJ, Clark A H. Glutathione S-transferase expression in benign and malignant ovarian tumours. Br J Cancer 1993;68:235–9.
11. Zhang L, Xiao Y, Priddy R. Increase in placental glutathione S-transferase in human oral epithelial dysplastic lesions and squamous cell carcinomas. J Oral Pathol Med 1994;23:75–9.
12. Katagiri A, Tomita Y, Nishiyama T, Kimura M, Sato S. Immunohistochemical detection of P-glycoprotein and GSTP1-1 in testis cancer. Br J Cancer 1993;68: 125–9.
13. Ban N, Takahashi Y, Takayama T, et al. Transfection of glutathione S-transferase (GST)-pi antisense complementary DNA increases the sensitivity of a colon cancer cell line to adriamycin, cisplatin, melphalan, and etoposide. Cancer Res 1996;56:3577–82.
14. Adler V, Yin Z, Fuchs SY, et al. Regulation of JNK signaling by GSTp. EMBO J 1999;18:1321–34.
15. Wang T, Arifoglu P, Ronai Z, Tew KD. Glutathione S-transferase P1-1 (GSTP1-1) inhibits c-Jun N-terminal kinase (JNK1) signaling through interaction with the C terminus. J Biol Chem 2001;276:20999–21003.
16. Adler V, Pincus MR. Effector peptides from glutathione-S-transferase-pi affect the activation of jun by jun-N-terminal kinase. Ann Clin Lab Sci 2004;34:35–46.
17. Asakura T, Sasagawa A, Takeuchi H, et al. Conformational change in the active center region of GST P1-1, due to binding of a synthetic conjugate of DXR with GSH, enhanced JNK-mediated apoptosis. Apoptosis 2007;12:1269–80.
18. Mahajan S, Atkins WM. The chemistry and biology of inhibitors and pro-drugs targeted to glutathione S-transferases. Cell Mol Life Sci 2005;62:1221–33.
19. Tew KD. Redox in redux: emergent roles for glutathione S-transferase P (GSTP) in regulation of cell signaling and S-glutathionylation. Biochem Pharmacol 2007;73:1257–69.
20. Ricci G, De Maria F, Antonini G, et al. 7-Nitro-2,1,3-benzoxadiazole derivatives, a new class of suicide inhibitors for glutathione S-transferases. Mechanism of action of potential anticancer drugs. J Biol Chem 2005; 280:26397–405.
21. Turella P, Cerella C, Filomeni G, et al. Proapoptotic activity of new glutathione S-transferase inhibitors. Cancer Res 2005;65:3751–61.
22. Turella P, Filomeni G, Dupuis ML, et al. A strong glutathione S-transferase inhibitor overcomes the P-glycoprotein-mediated resistance in tumor cells. 6-(7-Nitro-2,1,3-benzoxadiazol-4-ylthio)hexanol (NBDHEX) triggers a caspase-dependent apoptosis in

MDR1-expressing leukemia cells. J Biol Chem 2006; 281:23725–32.
23. Ascione A, Cianfriglia M, Dupuis ML, et al. The glutathione S-transferase inhibitor 6-(7-nitro-2,1,3-benzoxadiazol-4-ylthio)hexanol overcomes the MDR1-P-glycoprotein and MRP1-mediated multidrug resistance in acute myeloid leukemia cells. Cancer Chemother Pharmacol 2009;64:419–24.
24. Filomeni G, Turella P, Dupuis ML, et al. 6-(7-Nitro-2,1,3-benzoxadiazol-4-ylthio)hexanol, a specific glutathione S-transferase inhibitor, overcomes the multidrug resistance (MDR)-associated protein 1-mediated MDR in small cell lung cancer. Mol Cancer Ther 2008;7: 371–9.
25. Pasello M, Michelacci F, Scionti I, et al. Overcoming glutathione S-transferase P1-related cisplatin resistance in osteosarcoma. Cancer Res 2008;68:5661–8.
26. Lo HW, Ali-Osman F. Genetic polymorphism and function of glutathione S-transferases in tumor drug resistance. Curr Opin Pharmacol 2007;7:367–74.
27. Battistoni A, Mazzetti AP, Petruzzelli R, et al. Cytoplasmic and periplasmic production of human placental glutathione transferase in *Escherichia coli*. Protein Expr Purif 1995;6:579–87.
28. Lo Bello M, Battistoni A, Mazzetti AP, et al. Site-directed mutagenesis of human glutathione transferase P1-1. Spectral, kinetic and structural properties of Cys-47 and Lys-54 mutants. J Biol Chem 1995;270:1249–53.
29. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–26.
30. Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr 2000;56:1622–4.
31. Raghunathan S, Chandross RJ, Kretsinger RH, Allison TJ, Penington CJ, Rule GS. Crystal structure of human class mu glutathione transferase GSTM2-2. Effects of lattice packing on conformational heterogeneity. J Mol Biol 1994;238:815–32.
32. Murshudov GN, Vagin A, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–55.
33. Emsley P, Cowtan K. COOT: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–32.
34. Oakley AJ, Lo Bello M, Battistoni A, et al. The structures of human glutathione transferase P1-1 in complex with glutathione and various inhibitors at high resolution. J Mol Biol 1997;274:84–100.
35. Laskowsky RA, MacArthur MW, Moss DS, Thornton JR. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26: 283–91.
36. Habig WH, Jakoby WB. Assays for differentiation of glutathione S-transferases. Methods Enzymol 1981;77: 398–405.
37. Ricci G, Caccuri AM, Lo Bello M, et al. Structural flexibility modulates the activity of human glutathione transferase P1-1. Role of helix 2 flexibility in the catalytic mechanism. J Biol Chem 1996;271:16187–92.
38. Caccuri AM, Ascenzi P, Antonini G, et al. Structural flexibility modulates the activity of human glutathione transferase P1-1. Influence of a poor co-substrate on dynamics and kinetics of human glutathione transferase. J Biol Chem 1996;271:16193–8.
39. Johansson A-S, Stenberg G, Widersten M, Mannervik B. Structure-activity relationships and thermal stability of human glutathione transferase P1-1 governed by the H-site residue 105. J Mol Biol 1998;278:687–98.
40. Oakley AJ, Rossjohn J, Lo Bello M, Caccuri AM, Federici G, Parker MW. The three-dimensional structure of the human Pi class glutathione transferase P1-1 in complex with the inhibitor ethacrynic acid and its glutathione conjugate. Biochemistry 1997;36:576–585.
41. Parker LJ, Ciccone S, Italiano LC, et al. The anti-cancer drug chlorambucil as a substrate for the human polymorphic enzyme glutathione transferase P1-1: kinetic properties and crystallographic characterisation of allelic variants. J Mol Biol 2008;380:131–44.
42. Awasthi S, Srivastava SK, Ahmad F, Ahmad H, Ansari GAS. Interactions of glutathione S-transferase-rr with ethacrynic acid and its glutathione conjugate. Biochim Biophys Acta 1993;1164:173–8.
43. Horton JK, Roy G, Piper JT, et al. Characterization of a chlorambucil-resistant human ovarian carcinoma cell line overexpressing glutathione S-transferase mu. Biochem Pharmacol 1999;58:693–702.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 48, pp. 37138–37149, November 26, 2010

Nucleophosmin C-terminal Leukemia-associated Domain Interacts with G-rich Quadruplex Forming DNA*

Received for publication, July 21, 2010, and in revised form, September 9, 2010 Published, JBC Papers in Press, September 27, 2010, DOI 10.1074/jbc.M110.166736

Luca Federici[‡1,2], **Alessandro Arcovito**[§1], **Giovanni L. Scaglione**[§¶], **Flavio Scaloni**[‖], **Carlo Lo Sterzo**[‖], **Adele Di Matteo**[**], **Brunangelo Falini**[‡‡], **Bruno Giardina**[§ §§], **and Maurizio Brunori**[‖]

From the [‡]Ce.S.I. Center of Excellence on Aging and Department of Biomedical Sciences, University of Chieti "G. D'Annunzio," 66013 Chieti, the [§]Institute of Biochemistry and Clinical Biochemistry and [¶]Department of Internal Medicine, Haemostasis Research Center, Università Cattolica del Sacro Cuore, 00168 Rome, the [‖]Istituto Pasteur, Fondazione Cenci Bolognetti and Department of Biochemical Sciences, "Sapienza" University of Rome, 00185 Rome, the [**]Institute of Biology and Molecular Pathology of the CNR, 00185 Rome, the [‡‡]Institute of Hematology, University of Perugia, Perugia, and the [§§]Institute of Chemistry of Molecular Recognition of the CNR, 00168 Rome, Italy

Nucleophosmin (NPM1) is a nucleocytoplasmic shuttling phosphoprotein, mainly localized at nucleoli, that plays a key role in ribogenesis, centrosome duplication, and response to stress stimuli. Mutations at the C-terminal domain of NPM1 are the most frequent genetic lesion in acute myeloid leukemia and cause the aberrant and stable translocation of the protein in the cytoplasm. The NPM1 C-terminal domain was previously shown to bind nucleic acids. Here we further investigate the DNA binding properties of the NPM1 C-terminal domain both at the protein and nucleic acid levels; we investigate the domain boundaries and identify key residues for high affinity recognition. Furthermore, we demonstrate that the NPM1 C-terminal domain has a preference for G-quadruplex forming DNA regions and induces the formation of G-quadruplex structures *in vitro*. Finally we show that a specific sequence found at the *SOD2* gene promoter, which was previously shown to be a target of NPM1 *in vivo*, is indeed folded as a G-quadruplex *in vitro* under physiological conditions. Our data extend considerably present knowledge on the DNA binding properties of NPM1 and suggest a general role in the transcription of genes characterized by the presence of G-quadruplex forming regions at their promoters.

Nucleophosmin (also known as NPM1, B23, numatrin, and hereby termed NPM1)[3] is an abundant phosphoprotein that was originally identified as a non-ribosomal nucleolar protein playing a key role in ribosome biogenesis: NPM1 is able to bind RNA as well as DNA, and has intrinsic RNase activity that preferentially cleaves pre-rRNA (1–4). NPM1 is also involved in the regulation of the important tumor suppressors p53 (5) and p14arf (6, 7) and therefore mutations at the *NPM1* locus or aberrant localization of NPM1 may interfere with p53 or p14arf transcriptional programs and activities. Furthermore, NPM1 plays more than one role outside the nucleolus, including the control of centrosome duplication (8). This function is confirmed by analysis of *NPM1* knock-out mice showing unrestricted centrosome duplication, genomic instability, and mid-gestation embryonic lethality (9). The multiple cellular functions of NPM1 inside and outside the nucleolus are due to its chaperone activity and ability to shuttle between the nucleus and cytoplasm (10, 11). These properties are in turn dictated by its modular organization. In fact, several distinct, although partially overlapping, functional domains and signatures have been identified in NPM1 (12, 13).

NPM1 was first identified in 2005 as the most frequently mutated gene in acute myeloid leukemia, its mutations being observed in about 30% of cases (14) that display distinctive molecular and clinical features (13, 15, 16). Mutations are all localized at the C-terminal end of the protein (exon 12) and invariably result in the aberrant cytoplasmic localization of NPM1 (17). More than 40 alterations have been detected and were found to be mutually exclusive with major chromosomal abnormalities (13, 14). *NPM1* mutations, which are typically heterozygous, have similar consequences on the mutated protein. The reading frame is altered, due to duplication of short base sequences leading to a protein longer by 4 residues and with a sequence different from the wild type in the last 7 residues. As a result, the novel C-terminal sequence in the mutated proteins lacks the wild-type nucleolar localization signal and acquires a newly formed, Crm1-dependent, nuclear export signal (L*XXXVXXVX*L) (17, 18). Furthermore, the protein C-terminal domain appears greatly destabilized and partially or totally unfolded, due to alteration of the protein hydrophobic core (19–21). Both of these consequences at the protein level concur to determine the aberrant and stable cytoplasmic localization of the protein in the cytosol (17, 18).

The NPM1 C-terminal domain is responsible for the nucleic acid binding activity of the protein (Fig. 1*A*). NPM1 was shown to bind both DNA and RNA oligos with a preference for single-stranded structures over double-stranded DNA B structures, in a sequence-unrelated manner (3, 22, 23). Therefore, a role for NPM1 as a single-stranded binding protein was proposed (22), a property that may be linked to the export of ribosome sub-

* This work was supported by grants from the Associazione Italiana Ricerca sul Cancro (AIRC) (to M. B. and B. F.), Research Programs D.1 2009-2010 "Kinetic Study of Nucleophosmin by Surface Plasmon Resonance" and D.2.2 2008-2009 "SPR of Biological Macromolecules" from the Università Cattolica del Sacro Cuore (to A. A.) and grants from the University of Chieti "G. D'Annunzio" (to L. F.).

[1] Both authors contributed equally to this work.

[2] To whom correspondence should be addressed. Tel.: 39-0871541414; Fax: 39-08715410; E-mail: lfederici@unich.it.

[3] The abbreviations used are: NPM1, nucleophosmin; SOD, superoxide dismutase.

units from the nucleus (17). Recently, however, the protein was also identified as a cofactor in the transcriptional activation of the mitochondrial superoxide dismutase 2 (*SOD2*) gene; in particular it was shown that full-length NPM1 binds a G-rich region at the *SOD2* promoter (24). Because leukemic mutant NPM1 is stably localized in the cytoplasm, we decided to further investigate the DNA binding properties of the protein, assuming that some important function related to nucleic acid binding in the nucleus might be impaired in these mutants.

In this work we use different DNA sequences, protein constructs, and site-directed mutants to investigate the NPM1 C-terminal domain boundaries involved in DNA binding, assess the role of individual residues in binding, and analyze the effect of DNA and stabilizing salts on NPM1 folding. Importantly, we show that NPM1, whereas being able to recognize with low affinity any DNA oligo tested, binds with high affinity and is able to induce the folding of G-rich sequences that form three-dimensional structures known as G-quadruplexes. These include the above mentioned G-rich sequence at the *SOD2* gene promoter, which we demonstrate here is indeed folded as a G-quadruplex under physiological conditions. Taken together these data extend our knowledge on the DNA binding properties of NPM1 and suggest that future efforts should be paid to identify G-quadruplex forming DNA regions that are recognized by NPM1 *in vivo*.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—NPM1-C53 was expressed and purified as previously reported (20).

A gene construct for the expression of NPM1-C70 was obtained through gene synthesis (GeneArt, Regensburg, Germany) and cloned into expression vector pET28a(+) (Novagen, San Diego, CA) between restriction sites NdeI and BamHI. *Escherichia coli* cells, strain BL21(DE3), transformed with the expression vector, were grown to A_{600} 0.5 in LB medium supplemented with kanamycin at 37 °C. At this point 1 mM isopropyl 1-thio-β-D-galactopyranoside was added and cells were further grown at 20 °C for 16 h. Cells were harvested, resuspended in lysis buffer (20 mM Hepes, pH 7.0, 150 mM NaCl, 5 mM β-mercaptoethanol, and 20 mM imidazole) and sonicated. After centrifugation, the supernatant was loaded on a nickel-nitrilotriacetic acid column, pre-equilibrated with lysis buffer. Protein was eluted using a linear gradient of lysis buffer plus imidazole (20 mM to 1.0 M). Fractions containing the protein, as judged from SDS-PAGE, were collected, diluted 10-fold with lysis buffer, and incubated with thrombin (0.5 units/mg of protein) for 1.5 h at 4 °C. After thrombin cleavage the reaction mixture was supplemented with protease inhibitors (Roche Applied Science) and loaded on a nickel-nitrilotriacetic acid column, pre-equilibrated with lysis buffer, to remove the thrombin cleaved N-terminal His tag and uncleaved protein. Protein was recovered from the flow-through, diluted 10-fold with 20 mM Hepes, pH 7.0, 5 mM β-mercaptoethanol, and loaded on a SP-Sepharose column pre-equilibrated with the same buffer. Protein was eluted using a NaCl linear gradient, buffer was exchanged to remove the salt, concentrated up to 40 mg/ml in 20 mM Hepes buffer, pH 7.0, 5 mM β-mercaptoethanol, and stored at −80 °C.

Mutagenesis—NPM1-C70 protein variants were obtained through site-directed mutagenesis and using the QuikChange II kit (Stratagene, La Jolla, CA), following the manufacturer's instructions. Forward oligos used for PCR are the following: K229A, 5′-CATATGCAGGAAAGCTTCgcgAAACAGGAAAAAACGCCGAAAACC-3′; K230A, 5′-CATATGCAGGAAAGCTTCAAAgcgCAGGAAAAAACGCCGAAAACC-3′; K229A/K230A, 5′-CATATGCAGGAAAGCTTCgcggcgCAGGAAAAAACGCCGAAAACCCCGAAAGGCCCG-3′; K233A, 5′-CAGGAAAAAACGCCGgcgACCCCGAAAGGCCCGAGC-3′; K236A, 5′-CCGAAAACCCCGgcgGGCCCGAGCAGCGTG-3′; and K239A, 5′-GAAAGCTTCAAAAAACAGGAAgcgACGCCGAAAACCCCGAAAGGC-3′. NPM1-C70 protein variants were expressed and purified as the wild-type construct.

Oligonucleotides—Oligonucleotides used in this study (Fig. 1*B*) were purchased from PRIMM s.r.l (Milan, Italy) and purified by HPLC. Oligonucleotides used for SPR analysis were biotinylated at the 5′-end. Prior to use, lyophilized oligos were resuspended in the appropriate buffer (20 mM Hepes, pH 7.0, with or without 100 mM NaCl or KCl according to the different experiments), quantified spectrophotometrically and annealed. For annealing, oligos were heated to 95 °C for 15 min and slowly cooled down at RT, overnight.

Surface Plasmon Resonance—The interactions between biotinylated DNA constructs (ligands) with purified proteins NPM1-C70 or NPM1-C53 (analytes) as well as with the porphyrin TmPyP4 (analyte), were all measured using the SPR technique and a Biacore X100 instrument (Biacore, Uppsala, Sweden). Each biotinylated DNA construct was immobilized on a Sensor Chip SA, pre-coated with streptavidin from Biacore AB. The capturing procedure on the biosensor surface was performed according to the manufacturer's instructions and setting the aim for ligand immobilization to 1000 response units. Running buffer was Hepes-buffered saline-EP, which contains 10 mM Hepes, pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% (v/v) Surfactant P20 (Biacore AB). Analytes were dissolved in running buffer, and binding experiments were performed at 25 °C with a flow rate of 30 μl/min. The association phase (k_{on}) was followed for 180 s, whereas the dissociation phase (k_{off}) was followed for 300 s. The complete dissociation of active complex formed was achieved by addition of 10 mM Hepes, 2 M NaCl, 3 mM EDTA, 0.005% (v/v) P20, pH 7.4, for 60 s before each new cycle start. Analytes were tested in a wide range of concentrations to reach at least a 2-fold increase from lower concentration tested. When experimental data met quality criteria, kinetic parameters were estimated according to a 1:1 binding model using Biacore X100 Evaluation Software. Conversely, an affinity steady state model was applied to fit the data. In this latter case two possible situations were exploited: (*a*) single binding site, using the equation $y = R_{max} \times [analyte]/([analyte] + K_D)$; and (*b*) double independent binding site, using the equation $y = R_{max} \times [analyte]/([analyte] + K_D) + R_{max1} \times [analyte]/([analyte] + KD1)$.

Circular Dichroism—All circular dichroism experiments were performed using a Jasco J710 instrument (Jasco Inc., Easton, MD) equipped with a Peltier apparatus for temperature control. Static spectra of G10-loop and c-MYC oligos were collected at 25 °C, using oligos annealed in the appropriate buffer

and concentrated to 20 μM. Spectra were collected using a quartz cell with 1-mm optical path length (Hellma, Plainview, NY) and a scanning speed of 100 nm/min. The reported spectra are the average of five scans. To measure NPM1-C70 induced G-quadruplex formation, oligos dissolved in buffer (without additional salts) but not annealed, and concentrated to 20 μM, were incubated for 1 h with appropriate amounts of NPM1-C70 and spectra collected as above. To monitor NPM-C70 induced unfolding, oligos were annealed in Hepes buffer supplemented with 100 mM NaCl and incubated with NPM1-C70 increasing amounts. The spectral contribution of buffers and proteins was subtracted as appropriate. To measure the effect of the c-MYC and G10-loop oligos on NPM1-C70 structure, CD spectra of the protein (20 μM) in 20 mM Hepes, pH 7.0, alone or incubated with pre-annealed oligos (20 μM) were recorded. The spectral contribution of the oligo alone was then recorded and subtracted as appropriate. Thermal denaturation experiments were performed using a quartz cell with 1-mm optical path length and monitoring the variation of CD signal at 260 nm. Temperature was progressively increased, in 1 °C/min steps, from 25 to 105 °C. The Kaleidagraph software was used for CD spectra analysis and representation.

Urea-induced Protein Denaturation—NPM1-C70 and NPM1-C53 urea-induced denaturations were followed monitoring the change in CD signal at 222 nm, in the presence of increasing amounts of urea. The same experiments were performed in the absence or presence of NaCl 0.25, 0.5, and 1.0 M, respectively. Data were plotted and analyzed using Kaleidagraph software. A simple two-state model was used in all cases to fit the data according to Equation 1,

$$\Delta G_d = m_{D\text{-}N} \times (D - D_{1/2}) \quad \text{(Eq. 1)}$$

where ΔG_d is the free energy of folding at a concentration D of denaturant, $m_{D\text{-}N}$ is the slope of the transition (proportional to the increase in the surface-accessible area on going from the native to the denatured state), and $D_{½}$ is the midpoint of the denaturation transition. An equation that takes into account the pre- and post-transition baselines was used (25).

RESULTS

Surface Plasmon Resonance Analysis of NPM1-DNA Interaction—Full-length NPM1 is able to interact with a G-rich sequence found at the *SOD2* promoter and to function as a cofactor in transcriptional activation (24). This region is predicted to form a hairpin with a five-paired base stem and a stretch of 10 consecutive guanines in the loop (hereby named G10-loop; Fig. 1*B*). Our first aim was to determine whether this DNA binding activity could be mapped to the C-terminal domain of the protein. To this end, a biotinylated version of the G10-loop was immobilized on a streptavidin chip and used as bait in SPR analysis. Two different constructs of the NPM1 C-terminal domain were used as the analytes: NPM1-C70 comprising the last 70 C-terminal residues, and NPM1-C53 comprising the last 53 residues (Fig. 1*A*). The first construct was chosen because previous binding data on truncated proteins mapped the NPM1 nucleic acid binding activity to this segment (3), whereas the NPM1-C53 construct was prepared because it is the minimal folding unit of the C-terminal domain (20) that, as established by NMR data, is made of three helices (Fig. 1*A*) (19).

When testing NPM1-C70 as the analyte we obtained a K_D = 7.2 μM (Fig. 2*A*, Table 1), whereas with NPM1-C53 we obtained a K_D = 169 μM (Fig. 2, *B* and *C*, Table 1). Interestingly, both the association and dissociation rate constants could be determined when NPM1-C70 was the analyte (Fig. 2*A*), whereas they were too rapid to be determined with NPM1-C53 (Fig. 2*B*). This result suggests that the nucleic acid binding site is altered in the shorter protein construct and that, in this case, binding is mainly dictated by electrostatic interactions of the positively charged protein domain (Fig. 1*A*) with the negatively charged G10-loop.

Next, we wanted to clarify the structural properties of the DNA hairpin necessary for binding NPM1. To this end we first tested a so-called T10-loop, where the 10 guanines at the loop of the G10-hairpin are replaced by 10 thymines, whereas all the other bases are conserved (Fig. 1*B*). With NPM1-C70 we obtained a K_D = 307 μM that increased to K_D = 1.17 mM with NPM1-C53 (Table 1). This experiment indicates that a sequence that forms a hairpin structure resembling that of the native G10-loop but with thymines instead of guanines at the loop is poorly recognized, with a 42-fold lower affinity. NPM1-C53 is poorly competent for binding.

To explore the dimensional requirements of the hairpin for high affinity recognition, we next immobilized a G5-loop, which maintains the same hairpin arrangement of the G10-loop but with only five guanines in the loop (Fig. 1*B*). With this oligo, we obtained a K_D = 40 μM with NPM1-C70 and 224 μM with NPM1-C53. These experiments indicate that both the presence of guanines at the hairpin loop and their number contribute to the global affinity, suggesting that the three-dimensional structure of the G10-loop plays an important role.

To further assess the DNA binding properties of NPM1, we next tested an oligo made only of T-bases (38-mer), hereby named poly(T) (Fig. 1*B*), and obtained a K_D = 120 μM with NPM1-C70 and a K_D = 520 μM with NPM1-C53 (Table 1). This suggests that the increased flexibility of the poly(T) linear oligo with respect to the T10-loop can partly compensate for the absence of guanines.

A further experiment was designed to establish the role played by guanines in a DNA sequence that does not natively form hairpin structures. We thought that if the protein preferentially recognizes with high affinity a hairpin loop made of guanines, it might also be able to induce such loop formation in a poly(G) oligo and recognize it with good affinity (Fig. 1*B*). Thus with poly(G) we expected to find a K_D for NPM1-C70 higher than that found with the G10-loop but lower than that found with the poly(T) oligo.

To our surprise, with NPM1-C70 we obtained a K_D = 5.8 μM, comparable with that of the physiological substrate G10-loop; with NPM1-C53 we obtained instead a K_D = 31 μM, the lowest measured so far with this protein construct (Table 1). These results suggest that the recognition of a poly(G) oligo is far more specific than we might have anticipated. This may be rationalized by hypothesizing that this oligo is not linear but has the potential to form a structure that resembles that of the G10-

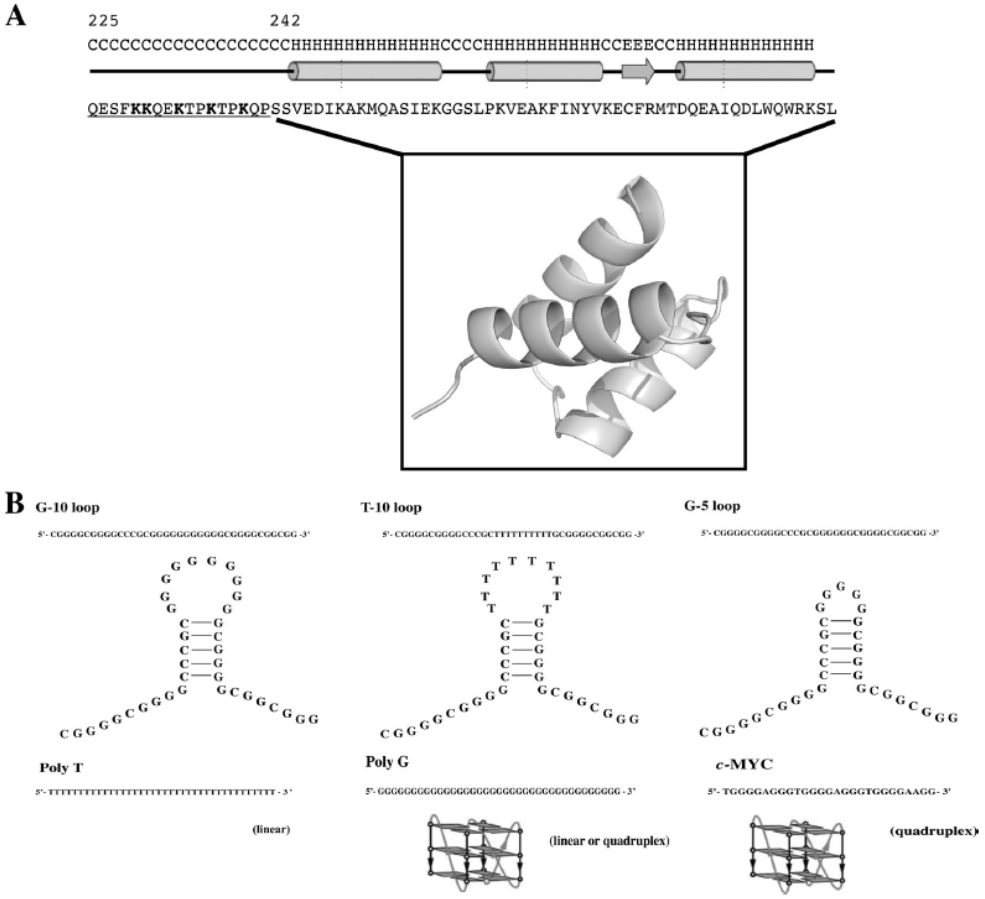

FIGURE 1. **Protein constructs and DNA oligos used in this study.** *A*, sequence of the C-terminal domain of NPM1. The region differentiating NPM1-C70 from NPM1-C53 (whose three-dimensional structure is shown in the *inset*) is *underlined*. Mutated lysines are shown in *bold*. The PSIPRED prediction is also shown (*C* = coil, *H* = α-helix, *E* = β-strand). *B*, oligos used in this study are shown both in sequence and in putative structure.

loop. Indeed, by inspecting the literature, we realized that a poly(G) oligo, if long enough, such as our 38-mer (Fig. 1*B*), has the potential to form three-dimensional structures known as G-quadruplexes (26, 27).

NPM1 C-terminal Domain Binds a Sequence Known to Form a G-quadruplex—G-quadruplexes are formed by sequences displaying at least four stretches of at least three guanines, with no sequence requirements for the intervening loops that are usually one to seven nucleotides long ($G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$) (26, 28). The guanines interact with each other in an arrangement different from the classical B DNA pairing, forming planar tetrads stabilized by the so-called Hoogsteen type H-bonds (27, 28). These structures are greatly stabilized by Na^+ or K^+ ions that intercalate in the rings formed by the four guanines in the tetrad. Repetitive G-rich sequence stretches are highly represented in the human genome and are clustered at gene promoters, suggesting their functional importance (27, 28).

We decided to investigate whether one well characterized G-quadruplex might interact with our NPM1 constructs. We focused our attention on a sequence contained in the NHE_{III} (nuclease hypersensitive element III) region of the c-*Myc* promoter that is known to regulate up to 90% of total c-MYC expression. This is a well characterized example of a parallel G-quadruplex forming region, both *in vitro* and *in vivo* (29). An oligo representing the G-quadruplex forming region of this promoter (hereby c-MYC oligo) was therefore immobilized on a chip for SPR analysis (Fig. 1*B*).

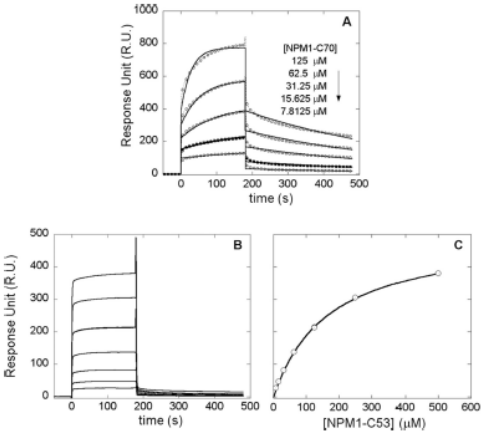

FIGURE 2. **SPR sensorgrams of the interaction between G10-loop and NPM1 constructs.** The oligo, biotinylated at its 5′ end, was immobilized on a SA (streptavidin) sensor chip. *A*, experimental curves (*dashed lines with circles*) represent different concentrations of NPM1-C70 used as the analyte and were fitted according to a single exponential binding model with 1:1 stoichiometry (*continuous lines*). *B*, experimental curves represent different concentrations of NPM1-C53. Both on and off rates were too rapid to be fitted. *C*, Scatchard plot used to determine the dissociation constant with NPM1-C53.

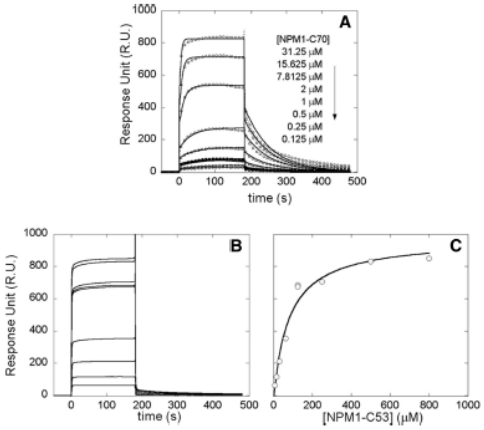

FIGURE 3. **SPR sensorgrams of the interaction between c-MYC and NPM1 constructs.** *A*, experimental curves (*dashed lines with circles*) represent different concentrations of NPM1-C70 used as the analyte and were fitted according to a single exponential binding model with 1:1 stoichiometry (*continuous lines*). *B*, experimental curves represent different concentrations of NPM1-C53. *C*, Scatchard plot referring to curves in *panel B* for dissociation constant determination.

TABLE 1
Dissociation constants and kinetic parameters as determined by SPR analysis

Ligand	Analyte	K_D	k_{on}	k_{off}
		μM	$M^{-1}s^{-1}$	s^{-1}
G10-loop	NPM1-C70	7.2 ± 0.2	$2.8 \pm 0.1 \times 10^2$	$1.98 \pm 0.03 \times 10^{-3}$
	NPM1-C53	169 ± 5		
T10-loop	NPM1-C70	307 ± 7		
	NPM1-C53	1170 ± 40		
G5-loop	NPM1-C70	40 ± 3		
	NPM1-C53	224 ± 6		
Poly(T)	NPM1-C70	120 ± 26		
	NPM1-C53	520 ± 30		
Poly(G)	NPM1-C70	5.8 ± 0.8		
	NPM1-C53	31 ± 4		
c-Myc	NPM1-C70	1.9 ± 0.1	$7.4 \pm 0.3 \times 10^3$	$1.37 \pm 0.03 \times 10^{-2}$
	NPM1-C53	82 ± 15		

Interestingly, we found that NPM1-C70 binds the c-MYC quadruplex with a $K_D = 1.9$ μM (Fig. 3*A* and Table 1), confirming that NPM1-C70 has high affinity for a sequence that adopts a G-quadruplex structure. Similarly to the G10-loop, and contrary to the other oligos, both the association and dissociation rates could be determined, suggesting that c-MYC G-quadruplex recognition is specific. Moreover, the experiment performed using NPM1-C53 as the analyte led to a $K_D = 82$ μM, confirming reduced affinity with this shorter domain (Fig. 3, *B* and *C*, and Table 1).

The G10-loop Forms a G-quadruplex Structure in Vitro—The results reported so far suggest that the NPM1 C-terminal domain recognizes with particularly high affinity sequences known to form G-quadruplex three-dimensional arrangements. Under this light, it is interesting to note that the G10-loop sequence, which is predicted to form a hairpin according to conventional Watson-Crick pairing (Fig. 1B), also matches the above mentioned folding rule for G-quadruplexes and, accordingly, is predicted to form a G-quadruplex by the Quadparser algorithm (26). Therefore, when annealing, this oligo has the potential to form at least two alternative structures. To infer which of these two structures is the most likely to be populated, we first collected the CD spectra of G10-loop and c-MYC, for comparison. It is well known that circular dichroism is diagnostic of G-quadruplex formation (28). In particular, by comparing the spectra of the c-MYC oligo annealed in the absence or presence of 100 mM NaCl or 100 mM KCl, respectively, we observe a red-shift and increase in intensity of the peak at around 260 nm and the formation of a through at 240 nm (Fig. 4*A*). Importantly the same features are observed in the case of the G10-loop spectra (Fig. 4*B*). These variations are both considered hallmarks of parallel G-quadruplex formation (28, 30). A second indication of a G-quadruplex structure for the G10-loop is derived from denaturation experiments. In Fig. 4*C* the thermal melting profiles of G10-loop in the 25–105 °C interval are shown, whereas their corresponding static spectra are reported in Fig. 4*D*. These data indicate that the melting transition is still not complete at 105 °C for the G10-loop in the presence of 100 mM KCl or NaCl, whereas a poorly cooperative transition with a T_m centered at around 65–70 °C is observed when the same experiment is performed in the absence of monovalent cations. The predicted melting temperature for a G10-loop adopting hairpin structure is $T_m = 69.1$ °C in 100 mM monovalent cations, according to the mfold server. Conversely the Quadpredict algorithm predicts higher T_m values, *i.e.* 94.9 and 77.1 °C, in the presence of 100 mM KCl or NaCl, respectively. These higher values are in better agreement with our experimental data, suggesting a G-quadruplex structure in these conditions. Thermal melting profiles were also collected for the c-MYC oligo (Fig.

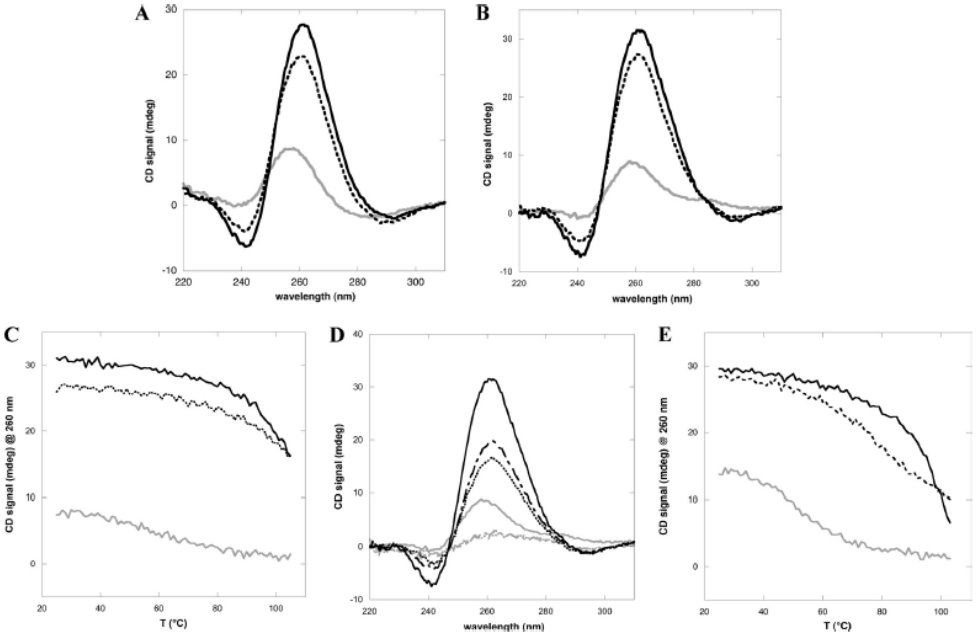

FIGURE 4. **G10-loop adopts a G-quadruplex fold *in vitro*.** *A*, CD spectra of the c-MYC oligo annealed in the absence (*gray continuous line*) or presence of 100 mM NaCl (*dashed line*) or 100 mM KCl (*black continuous line*), respectively. A red shift of the peak at 260 nm and the formation of a through at 240 nm is observed. *B*, CD spectra of the G10-loop oligo. The same features observed with c-MYC oligo are observed. The line format is the same as in *A*. *C*, thermal denaturation profiles of G10-loop in the 25–105 °C interval and in the absence (*gray continuous line*) or presence of 100 mM NaCl (*dashed line*) or 100 mM KCl (*black continuous line*). *D*, CD spectra of the G10-loop in the absence (*gray continuous*, *gray dashed dotted*, and *gray dotted* at 25, 95, and 105 °C, respectively) or presence of 100 mM KCl (*black continuous*, *black dashed dotted*, and *black dotted* at 25, 95, and 105 °C, respectively). Spectra at 95 and 105 °C in the absence of salts are the same, whereas in the presence of 100 mM KCl, they are different, indicating that the melting transition is not yet complete in the latter case. *E*, thermal denaturation profiles of c-MYC oligo in the 25–105 °C interval, collected in the absence (*gray continuous line*) or presence of 100 mM NaCl (*dashed line*) or 100 mM KCl (*black continuous line*).

4*E*), as a control, and found to be similar to those obtained with the G10-loop oligo (Fig. 4*C*).

Finally we analyzed the binding of the porphyrin TmPyP4 to our oligos, by means of SPR. This molecule is known to bind G-quadruplex structures with high affinity and a complex stoichiometry involving at least two binding sites with different affinities (31–33). The interaction of TmPyP4 with poly(T) oligo (Fig. 5*A*) determines a series of sensorgrams that, in the concentration range explored, always reach equilibrium before the end of the contact time between analyte and ligand, allowing both kinetic analysis, with a simple 1:1 model interaction (Fig. 5*A*), and the construction of a Scatchard plot (Fig. 5*B*). The average value for this double determination is $K_D = 345 \pm 15$ nM and is compatible with the presence of a single binding site (Table 2). When analyzing the T10-loop data, again we observed a single binding site with a lower dissociation constant ($K_D = 35 \pm 15$ nM) (Fig. 5, *C* and *D*, and Table 2). Conversely, when analyzing TmPyP4 binding to the c-MYC oligo (Fig. 5, *E* and *F*) a more complex behavior was observed, depending on the TmPyP4 concentration. At low concentrations (*bottom curves* of Fig. 5*E*) the dissociation phase is characterized by a single process corresponding to the off-rate constant of a high affinity binding site. By increasing the TmPyP4 concentration, the off-rate becomes clearly biphasic (*upper curves* of Fig. 5*E*) reflecting the titration of a second binding site with lower affinity. Moreover, not all the traces reach the equilibrium, so that the corresponding Scatchard plot is determined for a subpopulation of the porphyrin concentrations studied. As a result, the equilibrium analysis performed via the Scatchard plot is not compatible with a single binding site (see the *dashed line* in Fig. 5*F*), whereas an excellent agreement is obtained assuming two independent binding sites (*solid line* in Fig. 5*F* and see Table 2 for the corresponding K_D values). Importantly, the experiments performed with the G10-loop indicated that TmPyP4 binding to this oligo follows the same behavior as observed with c-MYC, with the same stoichiometry and similar dissociation constants for the high and low affinity sites, respectively (Fig. 5, *G* and *H*, and Table 2). In conclusion, TmPyP4 binding data obtained with the G10-loop are in agreement with those obtained with c-MYC and with previous work on G-quadruplex forming oligos (31–33), whereas data obtained with the hairpin T10-loop are not. Taken together, our results suggest that, in the presence of

physiological amounts of monovalent cations, the G10-loop folds as a G-quadruplex *in vitro* and might also have this structure when recognized *in vivo* by NPM1 (24).

NPM1-C70 Stimulates the Formation of G-quadruplex Structures—Having established that NPM1-C70 binds with high affinity preformed G-quadruplex structures, we next investigated whether it is able to induce G-quadruplex formation *in vitro*. In Fig. 6*A* we report CD spectra of the not annealed c-MYC oligo titrated with NPM1-C70, in the absence of monovalent ions. By progressively increasing the amount of NPM1-C70 we observe, once again, a red-shift and increase in signal of the 260 nm peak and the progressive formation of a through at 240 nm, which indicates that a G-quadruplex is formed upon protein binding. The same effect, albeit less evident, is obtained when using the G10-loop oligo (Fig. 6*B*). In a mirror experiment, using pre-annealed c-MYC or G10-loop oligos (Fig. 6, *C* and *D*, respectively) we checked whether NPM1-C70 might have the property to destabilize pre-formed G-quadruplex structures. In both cases increasing amounts of protein had no effect on the CD signal of oligos pre-annealed in the presence of monovalent cations. Thus NPM1-C70 is able to induce G-quadruplex formation in unstructured oligos, whereas it does not unwind pre-structured oligos.

Identification of Key Residues in the NPM1-quadruplex Interaction—Data reported so far indicate that the NPM1 region comprised between the longer NPM1-C70 construct and the shorter NPM1-C53 construct (aa 225–241) is necessary for high affinity binding. This region contains five lysine residues that we hypothesized might play a role in the specific recognition played by NPM1-C70 *versus* NPM1-C53 (Fig. 1*A*). To test this hypothesis we mutated each of them into alanine. The K229A/K230A double mutant was also prepared. With these variant proteins we performed a complete set of SPR experiments to

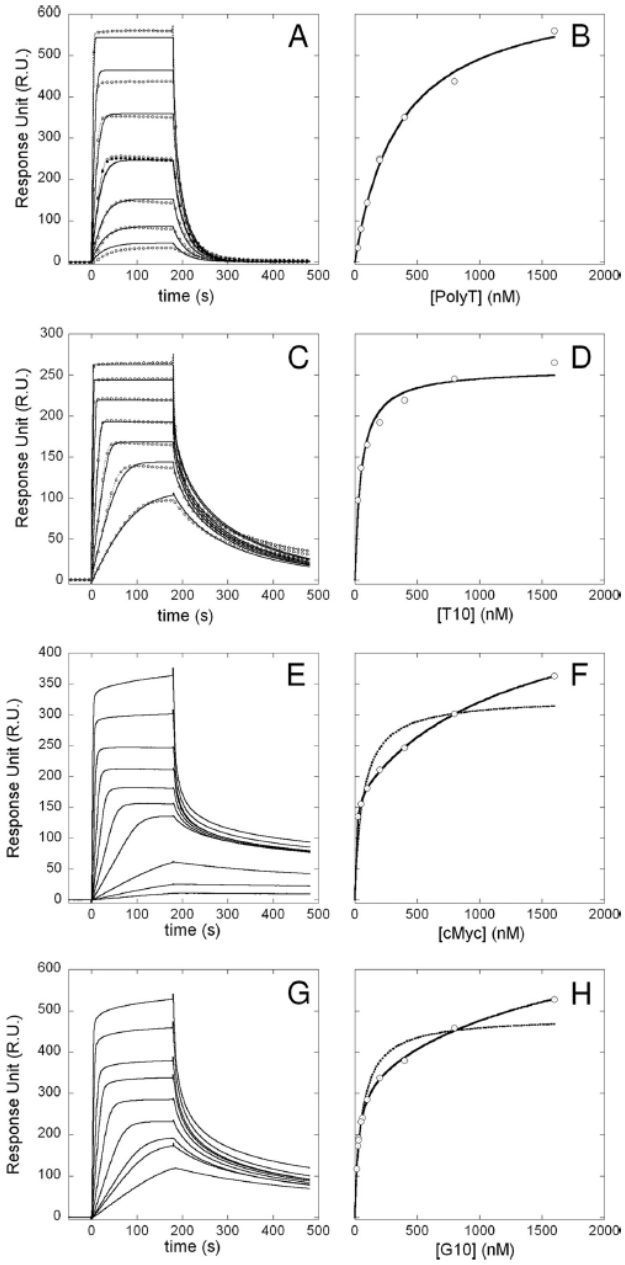

TABLE 2
SPR analysis of TmPyP4 binding to selected DNA oligos

Ligand	K_D	k_{on}	k_{off}
	M	$M^{-1}\ s^{-1}$	s^{-1}
Poly T	$330.9 \pm 0.4 \times 10^{-9}$ (Kinetics) $360 \pm 30 \times 10^{-9}$ (Scatchard) $345 \pm 15 \times 10^{-9}$ (average value)	$1.013 \pm 0.001 \times 10^{8}$	33.52 ± 0.02
T10-loop	$20 \pm 10 \times 10^{-9}$ (Kinetics) $50 \pm 8 \times 10^{-9}$ (Scatchard) $35 \pm 15 \times 10^{-9}$ (average value)	$2.9 \pm 0.8 \times 10^{10}$	$6 \pm 2 \times 10^{2}$
c-Myc	Site 1: $8 \pm 14 \times 10^{-9}$ Site 2: $1.3 \pm 0.2 \times 10^{-6}$		
G10-loop	Site 1: $26 \pm 3 \times 10^{-9}$ Site 2: $1.9 \pm 0.9 \times 10^{-6}$		

measure the binding affinities with the G10-loop and c-MYC oligos. Interestingly, none of the single mutants significantly affected the interaction when the variant proteins were tested against both oligos (Table 3). However, when testing the double mutant K229A/K230A, we observed a marked loss of affinity and obtained a $K_D = 135\ \mu$M for the G10-loop and $K_D = 78\ \mu$M for the c-MYC oligo. These values are very close to those obtained when testing the two oligos with NPM1-C53 (compare Table 3 with Table 1) and suggest that residues Lys^{229} and Lys^{230} cooperate to the specific and high affinity recognition played by NPM1-C70 on both oligos. Besides that, they also suggest that the mode of recognition of the c-MYC and G10-loop is likely to be similar, pointing again to a G-quadruplex structure for the G10-loop oligo.

Folding Studies—Previous structural studies on the NPM1 C-terminal domain (19) as well as our own folding studies (20, 21) were performed on the NPM1-C53 construct rather than on NPM1-C70 (Fig. 1*A*). The reason is that the segment comprising the residues that distinguish NPM1-C70 from NPM1-C53 is predicted to belong to the large natively unstructured domain that separates the chaperone N-terminal domain of the protein from the C-terminal domain (12). In Fig. 1*A* we report the prediction made using the PSIPRED algorithm that shows a high consensus coil prediction for this region (sequence underlined). The same predictions are obtained when using all the algorithms available in the Disprot server (not shown). Because we showed that this region is necessary for high affinity binding, we now wondered if folding predictions were correct and if yes, whether this region folds upon contact with DNA. A first indication comes from CD analysis of the protein in the absence or presence of c-MYC (Fig. 7*A*) or G10-loop (Fig. 7*B*) oligos. In both cases, CD spectra are almost completely superimposible, suggesting that the secondary structure content is invariant, with no extra folding induced upon DNA binding.

To determine whether this segment is actually folded or unfolded, we performed comparative urea denaturation experiments on NPM1-C70 and NPM1-C53 in the presence of increasing amounts of NaCl (Fig. 7, *C* and *D*, respectively). Two critical values are extrapolated by fitting the denaturation curves according to a simple two-state model, *i.e.* the midpoint of denaturation $Urea_{1/2}$ and the $m_{D\text{-}N}$ value. The latter is the slope of the transition and gives an estimate of the change in solvent accessible surface when moving from the denatured to the native state (therefore the higher this value the more compact the native state). The product of these two values represents protein stability (ΔG). In Table 4 it is shown that, in the absence of salt, the ΔG value for NPM1-C70 is considerably higher than the ΔG for NPM1-C53, by roughly 1.5 kcal/mol. In fact, whereas the $m_{D\text{-}N}$ values are similar, the $Urea_{1/2}$ value for NPM1-C70 is 3.8 *versus* 2.6 M for NPM1-C53. If the region differentiating the two constructs were natively unstructured and free to move in the solvent in NPM1-C70, one would expect similar $Urea_{1/2}$ and ΔG values. On the contrary our data suggest that this segment forms a compact structure with the remaining protein segment and thus contributes to overall stability. To further corroborate this finding we performed the same experiments in the presence of increasing salt amounts (Table 3). It is well known that NaCl is highly effective in stabilizing protein structure and promoting compactness (20, 21). Therefore we reasoned that, if the segment were natively unstructured in the absence of salt, it should progressively acquire compactness when increasing the salt concentration and this should be reflected in increased $m_{D\text{-}N}$ values. Our data indicate that this does not occur, with $m_{D\text{-}N}$ values remaining roughly the same, within the experimental error, when increasing NaCl concentration up to 1.0 M. Taken together these experiments indicate that, contrary to predictions, NPM1-C70 adopts a compact structure, without an unstructured N-terminal arm, which constitutes the functional DNA binding domain of the protein.

DISCUSSION

In this work, we have investigated the DNA binding properties of NPM1, building on the foundations already established by others (3, 22–24). Our analysis suggests that, besides NPM1-C53 being able to recognize any oligo tested with low affinity, high affinity recognition is achieved only with NPM1-C70. Moreover, rather than being just a single-stranded binding protein as initially suggested (3, 22), the preferential binding of NPM1 is for oligos with a defined three-dimensional structure. This is the case of the G10-loop, a sequence found at the promoter of the *SOD2* gene, which binds NPM1 and was predicted to form a hairpin based on conventional Watson-Crick pairing (24). However, a hairpin structure is not sufficient to confer high recognition properties, as shown by the results obtained with the T10-loop. The presence and number (see the G5-loop)

FIGURE 5. **SPR sensorgrams of the interaction between the porphyrin TmPyP4 and different oligos (*panels A, C, E,* and *G* for poly(T), T10-loop, c-MYC, and G10-loop, respectively) and their corresponding Scatchard plots (*panels B, D, F,* and *H*).** *A*, poly(T), from bottom to top of the panel [TmPyP4] = 1600, 800, 400, 200, 100, 50, and 25 nM. Experimental data are reported as *dashed lines with circles* (where the concentration was measured twice, also as *dashed line with squares*); best fit assuming a 1:1 interaction is reported as a *solid line*. *B*, poly(T), Scatchard plot, *open circles* are experimental data points and the *solid line* is the fit assuming a single binding site. *C*, T10-loop, from bottom to top of the panel [TmPyP4] = 1600, 800, 400, 200, 100, 50, and 25 nM. Symbols used are the same as in *panel A*. *D*, T10-loop, Scatchard plot, the *continuous line* represents the best fit according to 1:1 interaction model. *E*, c-MYC, from bottom to top of the panel [TmPyP4] = 1600, 800, 400, 200, 100, 50, 25, 12.5, 6.25, and 3.125 nM. Experimental data are reported as *solid lines*. *F*, c-MYC, concentrations used in the Scatchard plot are [TmPyP4] = 1600–25 nM. *Open circles* are experimental data points and the *dashed line* represents the best fit assuming a single binding site, whereas the *solid line* is the best fit assuming two independent binding sites. *G*, G10-loop, from bottom to top of the panel [TmPyP4] = 1600, 800, 400, 200, 100, 62.5, 31.25, 25, and 15.625 nM. Experimental data are reported as *solid lines*. *H*, G10-loop, the concentrations used in the Scatchard plot are [TmPyP4] = 1600–31.25 nM. Symbols used are the same as in *panel F*.

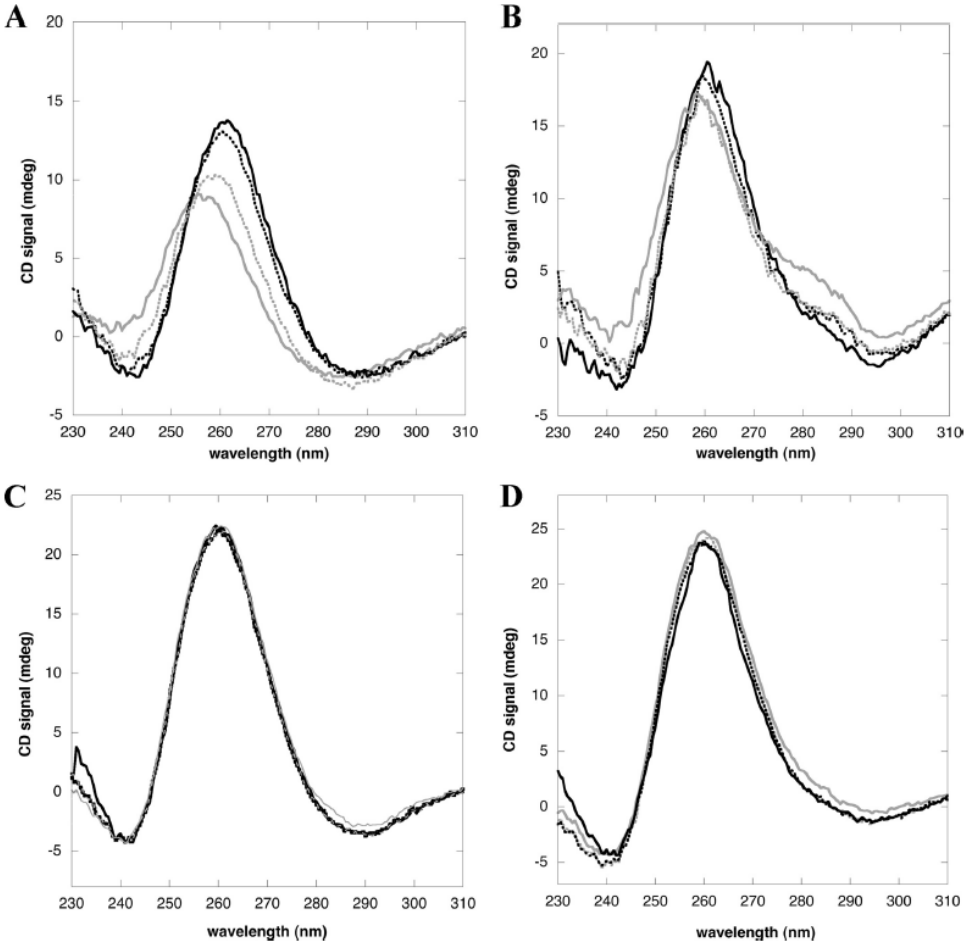

FIGURE 6. **NPM1-C70 is able to promote the formation of G-quadruplex arrangement in unstructured oligos, whereas it does not unfold G-quadruplexes.** *A*, not annealed c-MYC oligo (20 μM) in the absence (*gray solid line*) and presence of increasing amounts of NPM1-C70 (*gray dots*, *black dots*, and *black solid line* for 10, 20, and 30 μM, respectively); the red shift and increase in amplitude at 260 nm and the formation of a through at 240 are indicative of G-quadruplex formation. *B*, not annealed G10-loop oligo (20 μM) in the absence (*gray solid line*) or presence of increasing amounts of NPM1-C70 (*gray dots*, *black dots*, and *black solid line* for 70, 100, and 150 μM, respectively). *C*, spectra of c-MYC oligo (20 μM), pre-annealed in the presence of 100 mM NaCl, and incubated with increasing amounts of NPM1-C70 (same amounts as in *panel A*). *D*, same experiment as in *panel C* performed using the G10-loop (20 μM). Amounts of NPM1-C70 are the same as in *panel B*.

of guanines in the loop, and therefore the three-dimensional structure adopted by this particular sequence, is indeed critical (see Table 1).

Further investigations indicate that the expected hairpin might not be the structure adopted by the G10-loop *in vitro* and possibly also *in vivo*. First, we showed that NPM1-C70 is able to recognize with high affinity G-quadruplex forming oligos, like the poly(G) oligo and, above all, the G-rich sequence found at the $NHEIII_1$ region of the c-*MYC* promoter. Then we presented several lines of evidence that the G10-loop, whose sequence obeys the G-quadruplex folding rule (26), is indeed folded as a parallel G-quadruplex, at least *in vitro*. In fact, we showed by CD spectroscopy that the G10-loop presents all the hallmarks of parallel G-quadruplex formation in the presence of physiological concentrations of K^+ and Na^+ ions. Moreover we found that the thermal unfolding of this oligo in the presence of

TABLE 3
Effect of single alanine mutations on NPM1-C70 binding to G10-loop and c-MYC oligos

	G10-loop			c-MYC		
	K_D	k_{on}	k_{off}	K_D	k_{on}	k_{off}
	μM	$M^{-1}s^{-1}$	s^{-1}	μM	$M^{-1}s^{-1}$	s^{-1}
NPM1-C70	7.2 ± 0.2	2.8 ± 0.1 × 10^2	1.98 ± 0.03 × 10^{-3}	1.9 ± 0.1	7.4 ± 0.3 × 10^3	1.37 ± 0.03 × 10^{-2}
K229A	14.7 ± 0.8	1.6 ± 0.2 × 10^2	2.30 ± 0.04 × 10^{-3}	10.3 ± 0.9	7.3 ± 0.4 × 10^2	7.5 ± 0.3 × 10^{-3}
K230A	10.9 ± 0.3	1.50 ± 0.02 × 10^2	1.60 ± 0.02 × 10^{-3}	15 ± 1	2.8 ± 0.1 × 10^2	4.3 ± 0.1 × 10^{-3}
K233A	7.5 ± 0.5	4.5 ± 0.2 × 10^2	3.4 ± 0.1 × 10^{-3}	17.2 ± 0.6	3.0 ± 0.1 × 10^2	5.2 ± 0.1 × 10^{-3}
K236A	8.1 ± 0.2	3.3 ± 0.1 × 10^2	2.70 ± 0.02 × 10^{-3}	10 ± 1	2.8 ± 0.2 × 10^2	2.7 ± 0.2 × 10^{-2}
K239A	19.8 ± 0.5	1.04 ± 0.02 × 10^2	2.05 ± 0.02 × 10^{-3}	6.8 ± 0.9	8.7 ± 0.7 × 10^2	5.9 ± 0.3 × 10^{-3}
K229A/K230A	135 ± 8			78 ± 7		

monovalent cations is still not complete at temperatures as high as 105 °C, reflecting the enhanced stability conferred to G-quadruplex structures by K^+ and Na^+. Finally we showed that the G10-loop, contrary to the T10-loop, is capable of binding the porphyrin TmPyP4 with the same stoichiometry and dissociation constants comparable with those of the c-MYC oligo, thus behaving as expected for a G-quadruplex forming oligo.

G-quadruplex forming regions are frequent in the human genome and, interestingly, they are particularly abundant at gene promoter regions, with roughly 47% of all gene promoters having putative G-quadruplex forming regions (26, 28). A large number of these regions have been investigated and shown to form fully folded quadruplexes at least *in vitro* (34). In particular, it is interesting to note that these regions are frequently associated to oncogene promoters, including those of c-*Myc*, k-*ras*, *bcl2*, and c-*kit* genes, whereas a markedly reduced frequency at tumor suppressor promoters was observed (34). Moreover telomeres, which are made of the repetitive sequence TTAGGG, were also shown to form G-quadruplexes both *in vitro* and *in vivo*, these structures being detrimental to telomerase function (35). These findings suggested that drugs capable of stabilizing G-quadruplexes might have a therapeutic value in cancer treatment both because of telomerase inhibition or oncogene suppression (27). As a proof of concept, it was shown that the $NHEIII_1$ region at the c-*MYC* promoter suppresses c-*MYC* gene transcription when adopting a G-quadruplex structure and that the porphyrin TmPyP4, which stabilizes the G-quadruplex structure, is able to down-regulate this gene *in vivo* and to inhibit tumor growth in xenograft tumor models (36). The role of G-quadruplexes in gene regulation may, however, differ in other cases and be associated to gene activation, possibly through protein binding (28). This might be the case of the *SOD2* promoter, because we showed a G-quadruplex arrangement for the G10-loop region.

Given their importance in processes such as gene transcription and regulation and telomere elongation control, it is not surprising that a considerable number of proteins have been reported to interact with G-quadruplexes, most of them associated to telomere function (28). These include proteins that only bind G-quadruplexes, proteins that promote G-quadruplex formation, helicases that unwind G-quadruplexes, proteins that destabilize G-quadruplexes, and nucleases that are specific for G-quadruplexes (37). We have shown here that NPM1-C70, besides binding G-quadruplex forming regions, is also capable of promoting G-quadruplex formation *in vitro*, a property that may be relevant for NPM1 function in the cell. Having established that G10-loop folds as a G-quadruplex, and because this particular sequence is recognized by NPM1 *in vivo* and contributes to *SOD2* gene activation (24), it is tempting to speculate that there might be other G-quadruplex forming regions in the genome that are recognized and stabilized by NPM1.

This will be an important subject for future investigations. In fact, we know that mutations of NPM1 associated with acute myeloid leukemia map to the C-terminal G-quadruplex binding domain, destabilize the protein, and drive its aberrant translocation to the cytosol (both wild-type and mutant) (17). Therefore, a significant amount of NPM1 DNA binding activity occurring in the nucleus is likely to be lost in acute myeloid leukemia blasts and this may contribute to driving transformation through the loss of uncharacterized gene regulation activities.

Our studies also suggest that further work is needed to understand the structure and function of the NPM1 C-terminal domain. In fact whereas we showed that NPM1-C70 is the functional domain for G-quadruplex binding, only the structure of NPM1-C53 is available to date (19). This observation led us to focus our attention on the sequence differentiating these two constructs (see Fig. 1*A*). We mutated all five lysines contained in this sequence and found that single alanine substitutions cannot explain the marked loss of function of NPM1-C53 with respect to NPM1-C70. However, when mutating both Lys^{229} and Lys^{230} to alanine, we obtained dissociation constants comparable with those of NPM1-C53 when testing both the G10-loop and c-MYC oligos, suggesting that these two residues cooperate in recognizing some presently unknown structural feature within the G-quadruplex arrangement.

What then is the structure of NPM1-C70 and how are these two lysines positioned with respect to each other? We have previously underlined that all disorder prediction algorithms indicate this protein construct to resemble the structure of NPM1-C53, with an additional natively unfolded tail at the N terminus. However, we have shown here that (i) NPM1-C70 does not acquire additional secondary structure upon DNA binding, (ii) its stability is considerably higher than that of NPM1-C53, and (iii) that structure compactness is not increased by the presence of salts. These findings contrast the predictions and suggest instead that NPM1-C70 is folded as a compact domain, with the region comprised between amino acids 225 and 241 contributing to shape a G-quadruplex-binding site whose structure is presently unknown.

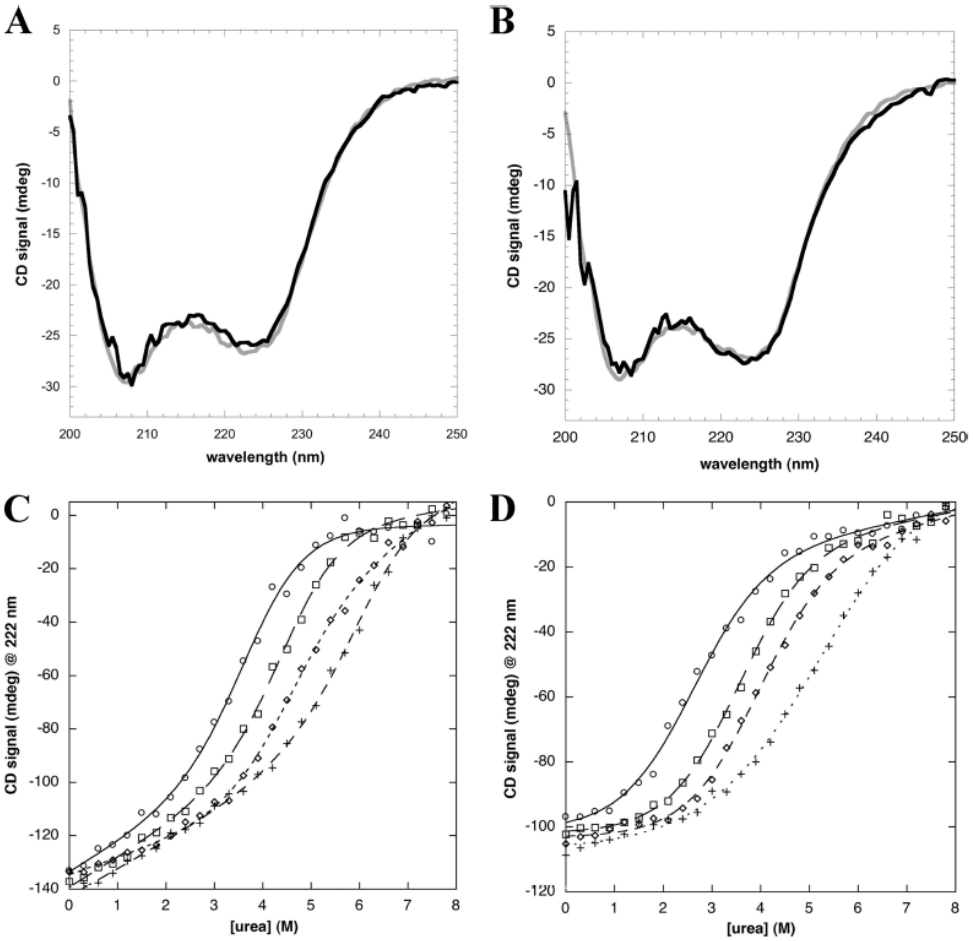

FIGURE 7. **Folding analysis of the region differentiating NPM1-C70 from NPM1-C53 (amino acids 225 to 241).** *A*, the CD spectra of NPM1-C70 (20 μM) in the absence (*gray line*) or presence of c-MYC oligo (20 μM; *black line*) are almost identical. *B*, CD spectra of NPM1-C70 (20 μM) in the absence (*gray line*) or presence of G10-loop oligo (20 μM; *black line*). *C*, urea-induced denaturation of NPM1-C70 in the absence (*open circles*) or presence of 250 mM (*open squares*), 500 mM (*open diamonds*), and 1.0 M (*crosses*) NaCl are reported. Data were fitted according to a two-state model (*lines*) and folding parameters are reported in Table 4. *D*, urea-induced denaturation of NPM1-C53 under the same conditions as in *panel C*. Symbols are also the same.

TABLE 4
Urea-induced denaturation of NPM1-C70 and NPM1-C53 in the presence of increasing salt amounts
$Urea_{1/2}$ and $m_{D\text{-}N}$ were obtained by fitting experimental data. ΔG values are calculated as the product of $Urea_{1/2}$ and $m_{D\text{-}N}$ values.

	$m_{D\text{-}N}$ NPM1-C70	$m_{D\text{-}N}$ NPM1-C53	$Urea_{1/2}$ NPM1-C70	$Urea_{1/2}$ NPM1-C53	ΔG NPM1-C70	ΔG NPM1-C53
No salt	0.93 ± 0.16	0.787 ± 0.07	3.76 ± 0.15	2.55 ± 0.07	3.49	2.02
0.25 M NaCl	0.93 ± 0.12	0.864 ± 0.08	4.51 ± 0.11	3.56 ± 0.09	4.19	3.06
0.50 M NaCl	0.82 ± 0.11	0.843 ± 0.08	4.55 ± 0.14	3.91 ± 0.08	3.77	3.29
1.00 M NaCl	0.79 ± 0.11	0.747 ± 0.09	6.20 ± 0.29	5.06 ± 0.17	4.9	3.79

Therefore structural studies of NPM1-C70 and its complexes with relevant G-quadruplex forming oligos are much in need, also because there is a paucity of structural information concerning the mechanism of G-quadruplex recognition by their interacting proteins (27). Moreover the C-terminal domain of NPM1 is considered as a possible drug target for the treatment

of acute myeloid leukemia with cytoplasmic NPM1 (13, 21). In principle, any molecule that targets this domain could stabilize a native-like state even in the mutated protein and prevent its association with the nuclear export system or translocation. Under this light, the structural analysis of NPM1 binding to nucleic acids would help in the rational design of such molecules.

Acknowledgment—We thank Benedetta Mattei for valuable advice in the design and execution of surface plasmon resonance experiments.

REFERENCES

1. Yung, B. Y., Busch, H., and Chan, P. K. (1985) *Biochim. Biophys. Acta* **826,** 167–173
2. Feuerstein, N., and Mond, J. J. (1987) *J. Biol. Chem.* **262,** 11389–11397
3. Wang, D., Baumann, A., Szebeni, A., and Olson, M. O. (1994) *J. Biol. Chem.* **269,** 30994–30998
4. Herrera, J. E., Savkur, R., and Olson, M. O. (1995) *Nucleic Acids. Res.* **23,** 3974–3979
5. Colombo, E., Marine, J. C., Danovi, D., Falini, B., and Pelicci, P. G. (2002) *Nat. Cell Biol.* **4,** 529–533
6. Korgaonkar, C., Hagen, J., Tompkins, V., Frazier, A. A., Allamargot, C., Quelle, F. W., and Quelle, D. E. (2005) *Mol. Cell. Biol.* **25,** 1258–1271
7. Bertwistle, D., Sugimoto, M., and Sherr, C. J. (2004) *Mol. Cell. Biol.* **24,** 985–996
8. Wang, W., Budhu, A., Forgues, M., and Wang, X. W. (2005) *Nat. Cell Biol.* **7,** 823–830
9. Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K., and Pandolfi, P. P. (2005) *Nature* **437,** 147–153
10. Borer, R. A., Lehner, C. F., Eppenberger, H. M., and Nigg, E. A. (1989) *Cell* **56,** 379–390
11. Szebeni, A., and Olson, M. O. (1999) *Protein Sci.* **8,** 905–912
12. Grisendi, S., Mecucci, C., Falini, B., and Pandolfi, P. P. (2006) *Nat. Rev. Cancer* **6,** 493–505
13. Falini, B., Sportoletti, P., and Martelli, M. P. (2009) *Curr. Opin. Oncol.* **21,** 573–581
14. Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., Bigerna, B., Pacini, R., Pucciarini, A., Liso, A., Vignetti, M., Fazi, P., Meani, N., Pettirossi, V., Saglio, G., Mandelli, F., Lo-Coco, F., Pelicci, P. G., Martelli, M. F., and GIMEMA Acute Leukemia Working Party (2005) *N. Engl. J. Med.* **352,** 254–266
15. Falini, B., Nicoletti, I., Martelli, M. F., and Mecucci, C. (2007) *Blood* **109,** 874–885
16. Garzon, R., Garofalo, M., Martelli, M. P., Briesewitz, R., Wang, L., Fernandez-Cymering, C., Volinia, S., Liu, C. G., Schnittger, S., Haferlach, T., Liso, A., Diverio, D., Mancini, M., Meloni, G., Foa, R., Martelli, M. F., Mecucci, C., Croce, C. M., and Falini, B. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 3945–3950
17. Falini, B., Bolli, N., Liso, A., Martelli, M. P., Mannucci, R., Pileri, S., and Nicoletti, I. (2009) *Leukemia* **23,** 1731–1743
18. Falini, B., Bolli, N., Shan, J., Martelli, M. P., Liso, A., Pucciarini, A., Bigerna, B., Pasqualucci, L., Mannucci, R., Rosati, R., Gorello, P., Diverio, D., Roti, G., Tiacci, E., Cazzaniga, G., Biondi, A., Schnittger, S., Haferlach, T., Hiddemann, W., Martelli, M. F., Gu, W., Mecucci, C., and Nicoletti, I. (2006) *Blood* **107,** 4514–4523
19. Grummitt, C. G., Townsley, F. M., Johnson, C. M., Warren, A. J., and Bycroft, M. (2008) *J. Biol. Chem.* **283,** 23326–23332
20. Scaloni, F., Gianni, S., Federici, L., Falini, B., and Brunori, M. (2009) *FASEB J.* **23,** 2360–2365
21. Scaloni, F., Federici, L., Brunori, M., and Gianni, S. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** 5447–5452
22. Dumbar, T. S., Gentry, G. A., and Olson, M. O. (1989) *Biochemistry* **28,** 9495–9501
23. Hingorani, K., Szebeni, A., and Olson, M. O. (2000) *J. Biol. Chem.* **275,** 24451–24457
24. Xu, Y., Fang, F., Dhar, S. K., St. Clair, W. H., Kasarskis, E. J., and St. Clair, D. (2007) *J. Biol. Chem.* **282,** 15981–15994
25. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry* **27,** 8063–8068
26. Huppert, J. L., and Balasubramanian, S. (2005) *Nucleic Acids Res.* **33,** 2908–2916
27. Neidle, S. (2009) *Curr. Opin. Struct. Biol.* **19,** 239–250
28. Huppert, J. L. (2008) *Chem. Soc. Rev.* **37,** 1375–1384
29. González, V., and Hurley, L. H. (2010) *Annu. Rev. Pharmacol. Toxicol.* **50,** 111–129
30. González, V., Guo, K., Hurley, L., and Sun, D. (2009) *J. Biol. Chem.* **284,** 23622–23635
31. Freyer, M. W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L. H., and Lewis, E. A. (2007) *Biophys. J.* **92,** 2007–2015
32. Wei, C., Jia, G., Yuan, J., Feng, Z., and Li, C. (2006) *Biochemistry* **45,** 6681–6691
33. Wei, C., Wang, J., and Zhang, M. (2010) *Biophys. Chem.* **148,** 51–55
34. Qin, Y., and Hurley, L. H. (2008) *Biochimie* **90,** 1149–1171
35. Lipps, H. J., and Rhodes, D. (2009) *Trends Cell Biol.* **19,** 414–422
36. Grand, C. L., Han, H., Muñoz, R. M., Weitman, S., Von Hoff, D. D., Hurley, L. H., and Bearss, D. J. (2002) *Mol. Cancer Ther.* **1,** 565–573
37. Fry, M. (2007) *Front. Biosci.* **12,** 4336–4351

Structure of Nucleophosmin DNA-binding Domain and Analysis of Its Complex with a G-quadruplex Sequence from the c-*MYC* Promoter*[S]

Received for publication, April 10, 2012, and in revised form, June 13, 2012 Published, JBC Papers in Press, June 15, 2012, DOI 10.1074/jbc.M112.371013

Angelo Gallo[‡1], **Carlo Lo Sterzo**[§1], **Mirko Mori**[‡], **Adele Di Matteo**[§], **Ivano Bertini**[‡], **Lucia Banci**[‡2], **Maurizio Brunori**[§3], **and Luca Federici**[¶]

From the [‡]*Magnetic Resonance Center, Department of Chemistry, University of Florence, 50019 Sesto Fiorentino, Italy,* [§]*Institute of Biology and Molecular Pathology of the National Research Council, Department of Biochemical Sciences "A. Rossi Fanelli," "Sapienza" University of Rome, 00185 Rome, Italy, and* [¶]*Center of Excellence on Aging and Department of Biomedical Sciences, University of Chieti "G. D'Annunzio", 66013 Chieti, Italy*

Background: Nucleophosmin leukemia-associated domain binds G-quadruplex DNA.
Results: NMR structural analysis of the 70-residue nucleophosmin C-terminal domain and its interaction with G-quadruplex DNA from the c-*MYC* promoter was carried out.
Conclusion: The interaction involves helices H1 and H2 of the nucleophosmin terminal three-helix bundle mainly through electrostatic contacts with G-quadruplex phosphates.
Significance: Learning how nucleophosmin interacts with nucleic acids may be crucial in rescuing its function in leukemia.

Nucleophosmin (NPM1) is a nucleocytoplasmic shuttling protein, mainly localized at nucleoli, that plays a key role in several cellular functions, including ribosome maturation and export, centrosome duplication, and response to stress stimuli. More than 50 mutations at the terminal exon of the *NPM1* gene have been identified so far in acute myeloid leukemia; the mutated proteins are aberrantly and stably localized in the cytoplasm due to high destabilization of the NPM1 C-terminal domain and the appearance of a new nuclear export signal. We have shown previously that the 70-residue NPM1 C-terminal domain (NPM1-C70) is able to bind with high affinity a specific region at the c-*MYC* gene promoter characterized by parallel G-quadruplex structure. Here we present the solution structure of the NPM1-C70 domain and NMR analysis of its interaction with a c-*MYC*-derived G-quadruplex. These data were used to calculate an experimentally restrained molecular docking model for the complex. The NPM1-C70 terminal three-helix bundle binds the G-quadruplex DNA at the interface between helices H1 and H2 through electrostatic interactions with the G-quadruplex phosphate backbone. Furthermore, we show that the 17-residue lysine-rich sequence at the N terminus of the three-helix bundle is disordered and, although necessary, does not participate directly in the contact surface in the complex.

Nucleophosmin (also called B23, numatrin, and herein NPM1[4]) is a protein involved in a variety of crucial cellular functions, including ribosome maturation and export, centrosome duplication, and response to stress stimuli (1, 2). NPM1 is a highly mobile protein capable of shuttling between nucleus, nucleoplasm, and cytoplasm, although the bulk of the protein is mainly localized in the nucleoli (3).

The *NPM1* gene, overexpressed in a number of solid tumors, has been proposed as a marker for colon, gastric, ovarian, and prostate carcinomas (1). *NPM1* is also frequently modified in hematopoietic tumors. For instance, in both lymphoid and myeloid disorders, *NPM1* chromosomal translocations lead to the production of several oncogenic fusion proteins (1). Furthermore, *NPM1* is the most frequently mutated gene in acute myeloid leukemia, accounting for 35% of all cases (2, 4); over 50 different mutations, always heterozygous and largely associated to a normal karyotype, were discovered (3, 4). Mutations involve duplication or insertion of small base sequences at the last exon of the gene and lead to a C-terminal domain that has acquired four additional residues compared with wild type and a completely different sequence in the last seven. The nucleolar localization signal is compromised with the loss of one or both of the two critical tryptophan residues, and a third nuclear export signal appears. Moreover, the mutated domain is largely destabilized or totally unfolded (5–7). Both the destabilization of NPM1 C-terminal domain and the presence of an additional nuclear export signal contribute to the aberrant and stable

* This work was supported by Associazione Italiana Ricerca sul Cancro Grants IG2007-4798 (to M. B.) and IG2011-11712 (to L. F.) and by Ministero dell'Istruzione, Università e Ricerca of Italy Grants RBRN07BMCT_007 (to M. B.), RBRN07BMCT (to L. B. and I. B.), and PRIN2009FAKHZT_001 (to L. B. and I. B.).

[S] This article contains a supplemental file showing the coordinates for the lowest energy NPM1-C70·Pu24I structural model.

The atomic coordinates and structure factors (code 2llh) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] Both authors contributed equally to this work.
[2] To whom correspondence may be addressed: Magnetic Resonance Center, University of Florence, Via Sacconi 6, 50019 Sesto Fiorentino, Italy. Tel.: 39-055-4574-263; Fax: 39-055-4574-271; E-mail: banci@cerm.unifi.it.
[3] To whom correspondence may be addressed: Dept. of Biochemical Sciences A. Rossi Fanelli, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy. Tel.: 39-06-4450291; Fax: 39-06-4440062; E-mail: maurizio.brunori@uniroma1.it.

[4] The abbreviations used are: NPM1, nucleophosmin; NPM1-C70, 70-residue NPM1 C-terminal domain; *SOD2*, superoxide dismutase 2; AIRs, ambiguous interaction restraints; NHE III, nuclease hypersensitive element III.

localization of the mutated protein in the cytoplasm, which is the salient feature of this type of leukemia (2).

The NPM1 C-terminal domain is known to bind both duplex and single-stranded DNA as well as RNA with no sequence specificity and with a preference for single-stranded nucleic acids (8). This picture was suggestive of a protein playing mainly a chaperone and transport role for preribosomal RNA particles (1, 9). Recently, however, it was shown that NPM1 binds a specific G-rich sequence at the superoxide dismutase 2 (*SOD2*) gene promoter and participates in the transcriptional activation of this gene (10).

Starting from these premises, we recently showed (11) that (i) NPM1 binds with high affinity DNA sequences that form G-quadruplexes, including the one found at the c-*MYC* oncogene promoter (see later); (ii) NPM1 is able to induce G-quadruplex formation in G-rich unstructured oligos; and (iii) the region of the *SOD2* promoter recognized *in vivo* by NPM1 is indeed folded as a G-quadruplex *in vitro* (11). We also investigated the domain boundaries necessary for DNA binding and demonstrated that a 17-residue segment preceding the C-terminal three-helix bundle (5) is necessary for high affinity recognition (11).

G-quadruplex DNA is gaining increasing attention because it is highly represented (especially in selected regions of the genome, including telomeres and gene promoters) and is involved in a number of regulatory processes (12, 13). Interestingly, G-quadruplex regions are frequently found at oncogene promoters, whereas a reduced frequency at tumor suppressor genes is observed (14). For instance, a well characterized G-quadruplex-forming sequence present at the NHE_{III} region of the c-*MYC* oncogene promoter (15) is recognized by NPM1 C-terminal domain *in vitro* (11) and folds as a G-quadruplex both *in vitro* and *in vivo*, regulating up to 90% of total c-MYC transcription (15–17). The interaction with proteins that stabilize the G-quadruplex fold, such as nucleolin, causes a marked down-regulation of the gene (18, 19), whereas the opposite happens through the interaction with G-quadruplex-unwinding proteins, such as NMR-H2 (20).

Given its selective localization at oncogene promoters and telomeres, G-quadruplex DNA is an attractive target for tumor treatment, and the structures of several G-quadruplex regions alone and in complex with drugs have been reported (21). Conversely, although the list of proteins that bind G-quadruplex DNA is rapidly increasing and the importance of such interactions for a variety of physiological processes is now clear, very little structural information is available concerning the molecular recognition mechanism of complex formation (22). To the best of our knowledge, only two structures are available: (i) thrombin in complex with a synthetic aptamer that folds as a G-quadruplex (23) and (ii) the *Oxytricha nova* telomere-binding protein heterodimer bound to its telomeric sequence (24).

Here we present the high resolution NMR structure of the NPM1 DNA-binding domain and analyze its interaction with the G-quadruplex DNA from the NHE_{III} region of the c-*MYC* promoter. We show (i) that the contact surface involves largely amino acids belonging to helices H1 and H2 of the terminal three-helix bundle and (ii) that a well defined G-quadruplex region is recognized through several electrostatic interactions with the phosphate backbone. The N-terminal lysine-rich region of the NPM1 C-terminal domain, which we show to be unstructured, does not participate directly in the interacting surface although it proved necessary to increase affinity (11). The data presented below unveil the interaction surface between G-quadruplex DNA and NPM1 and may inspire the search for small molecules or aptamers aimed at restoring a native-like fold in NPM1 leukemic mutants.

EXPERIMENTAL PROCEDURES

Oligonucleotides—The oligonucleotides used in this study are Pu27 of sequence 5′-T**GGGGA**G**GGT**G**GGGA**G**GGT**G**G**-**GG**AAGG-3′ and Pu24I of sequence 5′-**TGA**G**GGT**G**G**I**GA**G-**GGTGGGG**AAGG-3′. Pu27 and Pu24I were purchased from Primm (Milan, Italy) and Integrated DNA Technologies, Inc. (Coralville, IA), respectively, and were both HPLC-purified. Lyophilized oligos were dissolved in 20 mM phosphate buffer pH 7.0, 100 mM KCl and annealed. For annealing, oligos were heated to 95 °C for 15 min and then allowed to gently cool down to room temperature overnight. After annealing, the parallel G-quadruplex assembly of both oligos was assessed by inspecting CD spectra collected with a Jasco J-710 spectropolarimeter.

Protein Sample Preparation—A DNA construct for residues 225–294 of human NPM1 was cloned into pET28+(a) vector and transformed into *Escherichia coli* BL21(DE3) cells. For isotope enrichment, cells were grown in a minimal medium with $(^{15}NH_4)_2SO_4$ and [^{13}C]glucose. Protein expression was induced with 1.0 mM isopropyl β-D-thiogalactopyranoside at 20 °C, and cells were harvested after 16 h. NPM1-C70 was purified as reported previously (11). The hexahistidine tag at the N terminus of the protein was thrombin-cleaved and removed by nickel-nitrilotriacetic acid affinity column.

Structure Calculations of the Free Protein—The ^{1}H, ^{13}C, and ^{15}N resonance frequencies of NPM1-C70 were assigned using all classical NMR experiments. NMR experiments used for resonance assignment and structure calculations were performed on ^{13}C,^{15}N-labeled NPM1-C70 sample containing 10% D_2O. NMR spectra were collected at 298 K, processed using standard Bruker software (TOPSPIN 2.1), and analyzed with CARA (25).

Structure calculations were performed with the software package UNIO using as input the amino acid sequence; the chemical shift lists; and three ^{1}H-^{1}H NOE experiments, two-dimensional NOESY, three-dimensional ^{13}C-edited NOESY, and three-dimensional ^{15}N-edited NOESY, recorded at 900 MHz with a mixing time of 100 ms. The standard protocol included in UNIO with seven cycles of peak picking using ATNOS (26), NOE assignment with CANDID (27), and structure calculation with CYANA-2.1 (28) was used. φ and ψ dihedral angles were obtained from the chemical shift analysis using TALOS+ software (29–31). In each ATNOS/CANDID cycle, the angle constraints were combined with the updated NOE upper distance constraints in the input for subsequent CYANA-2.1 structure calculation cycle.

The 20 conformers with the lowest target function values were subjected to restrained energy minimization in explicit water with AMBER 11.0 (32, 33). NOE and torsion angle constraints were used. The quality of the structures was evaluated using the programs PROCHECK, PROCHECK-NMR (34), and

WHAT IF (35). Statistics about the energy-minimized family of conformers are reported in Table 1. The atomic coordinates and structural restraints for NPM1-C70 have been deposited in the Protein Data Bank with accession code 2llh. Resonance assignments are also available at BioMagResBank (accession number 18048).

Structure Calculations of the Complex—To identify intermolecular NOEs in the NPM1-C70·Pu24I complex, a ω_1-^{13}C-edited, ω_2-^{13}C-filtered NOESY experiment was recorded in a two-dimensional plane (^{1}H-^{1}H plane) on ^{13}C,^{15}N-labeled NPM1-C70·unlabeled Pu24I (36). The selected temperature was 290 K, and the mixing time used was 120 ms. ^{1}H, ^{13}C, and ^{15}N backbone resonances of NPM1-C70 in the complex were assigned by performing all the typical experiments for backbone assignment.

To calculate a structural model for the interaction between NPM1-C70 and Pu24I, a data-driven molecular docking was performed using the HADDOCK protocol. HADDOCK comprises a series of Python scripts that run on top of the structure determination programs ARIA and CNS (37–41). The method relies on the definition of ambiguous interaction restraints (AIRs) derived from experimental data.

For our docking calculations, we defined as unambiguous restraints residues resulting from intermolecular NOE cross-peaks on both protein and DNA. Furthermore, we defined as AIRs (i) residues experiencing chemical shift variations above the average variations plus one standard deviation (δ_{HN} = 0.043 ± 0.025 ppm; Fig. 3*B*, *black line*); (ii) residues whose signals broadened their line width (Glu^{245}, Lys^{248}, and Phe^{268}) and/or disappeared during the titration (Phe^{276}); and (iii) DNA atoms interacting with the protein, also identified by chemical shift perturbation, as detected in homonuclear experiments (one-dimensional and two-dimensional). Residues used as AIRs are listed in Table 2.

The HADDOCK docking protocol consists of (i) randomization of orientation and rigid body minimization, (ii) semirigid simulated annealing in torsion angle space, and (iii) final refinement in Cartesian space with explicit solvent. The rigid body docking step was performed five times with 1000 structures generated at each stage, the best 200 of which were refined in the semiflexible stage and subsequently in explicit water. Electrostatic and van der Waals terms were calculated with an 8.5-Å distance cutoff (37).

For the docking procedure, the structures of NPM1-C70 (here determined) and Pu24I (Protein Data Bank code 2A5P) were used as starting points. The coordinates for the lowest energy NPM1-C70·Pu24I structural model are included as supplemental material.

Relaxation Data—Heteronuclear relaxation experiments were performed on ^{15}N-labeled samples of NPM1-C70 at 700 MHz. The ^{15}N backbone longitudinal (R_1) and transverse (R_2) relaxation rates as well as heteronuclear ^{15}N{^{1}H} NOEs were measured using a standard protocol (42, 43).

RESULTS

The Structure of the NPM1-C70 Domain—The short C-terminal domain of NPM1 (residues 242–294; hereafter NPM1-C53) was shown by Grummitt *et al.* (5) to fold as a three-helix bundle. However, NPM1-C53 is poorly competent for DNA binding, and the 17-residue lysine-rich region at its N terminus (residues 225–241) is necessary for high affinity (11). Therefore, we decided to determine by NMR the structure of this longer construct encompassing the last 70 residues of the NPM1 sequence (residues 225–294; hereafter NPM1-C70). Statistics about structure determination are shown in Table 1.

TABLE 1
Statistical analysis of the energy-minimized family of conformers of NPM1-C70

r.m.s., root mean square; r.m.s.d., root mean square deviation; rad, radian; BB, backbone; HA, heavy atoms.

	AMBER 10.0[a] (20 structures)
r.m.s. deviations per meaningful distance constraint (Å)[b]	
Intraresidue (189)	0.0065 ± 0.0063
Sequential (279)	0.0208 ± 0.0062
Medium range (220)[c]	0.0198 ± 0.0031
Long range (40)	0.0106 ± 0.0122
Total (728)	0.0178 ± 0.0015
r.m.s. violations per meaningful dihedral angle constraints (°)[b]	
ϕ (42)	1.20 ± 1.01
ψ (42)	6.17 ± 4.19
Average no. of NOE violations larger than 0.3 Å	0.05 ± 0.22
Average NOE deviation (Å^2)[d]	0.01 ± 0.03
Average angle deviation (rad^2)[d]	0.66 ± 0.10
r.m.s.d. to the mean structure (Å) (BB)[e]	0.80 ± 0.21
r.m.s.d. to the mean structure (Å) (HA)[e]	1.40 ± 0.15
Structural analysis	
Residues in most favorable regions (%)[f,g]	96.5
Residues in allowed regions (%)[f,g]	3.3
Residues in generously allowed regions (%)[f,g]	0.1
Residues in disallowed regions (%)[f,g]	0.0
G-factor[f,g]	0.03
Structure Z-scores[g,h]	
First generation packing quality	1.27
Second generation packing quality	4.51
Ramachandran plot appearance	−1.50
χ_1/χ_2 rotamer normality	−3.04
Backbone conformation	−0.50

[a] AMBER indicates the energy-minimized family of 20 structures.
[b] The number of meaningful constraints for each class is reported in parentheses.
[c] Medium range distance constraints are those between residues (i, i + 2), (i, i + 3), (i, i + 4), and (i, i + 5).
[d] NOE and torsion angle constraints were applied with force constants of 20 kcal mol^{-1} Å^{-2} and 20 kcal mol^{-1} rad^{-2}, respectively.
[e] The r.m.s.d. to the mean structure is reported considering the segment 293–291.
[f] As it results from the Ramachandran plot analysis. For the PROCHECK statistics, an overall G-factor larger than −0.5 is expected for a good quality structure.
[g] The statistical analysis is reported considering the segment 293–291.
[h] Values based on WHAT IF output. A Z-score is defined as the deviation from the average value for this indicator observed in a database of high resolution crystal structures expressed in units of the standard deviation of this database-derived average. Typically, Z-scores below a value of −3 are considered poor, and those below −4 are considered bad.

The structure depicted in Fig. 1*A* comprises a well defined three-helix bundle, similar to the NPM1-C53 construct (5), in terms of length, relative orientation, and hydrophobic interactions between all the paired helices (Fig. 1*A*). On the other hand, the lysine-rich region (residues 225–241) that enhances DNA binding is unstructured as indicated by high values for ^{15}N-^{1}H R_1 (Fig. 1*B*) and low values for ^{15}N-^{1}H R_2 (Fig. 1*C*) and heteronuclear NOEs (Fig. 1*D*) with no secondary structure elements nor any propensity to fractionally take up secondary structures as clearly shown from the chemical shift index analysis (Fig. 1, *B* and *C*). Conversely, the folded part of the NPM1-C70 construct shows ^{15}N-^{1}H R_1 and ^{15}N-^{1}H R_2 values typical of an 8-kDa protein (Fig. 1, *B* and *C*) and positive values for het-

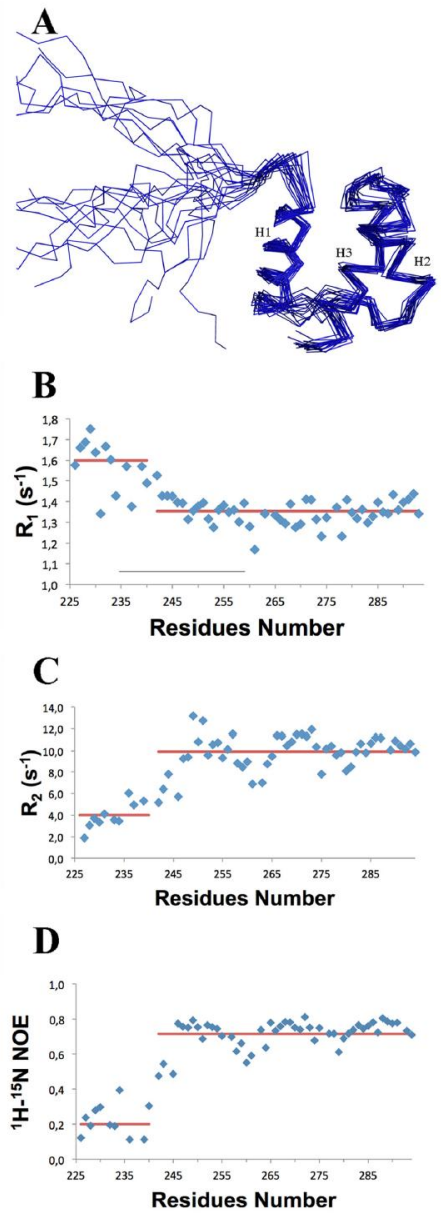

FIGURE 1. **NPM1-C70 quadruplex-binding C-terminal domain encompassing residues 225–294.** *A*, NMR solution structure of NPM1-C70 showing the 20 lowest energy structures. A lysine-rich natively unstructured segment (amino acids 225–242) precedes the terminal three-helix bundle. *B*, ^{15}N-^1H R_1 NOE values. *C*, ^{15}N-^1H R_2 NOE values. Low R_1 and high R_2 values for segment 225–242 are consistent with the N-terminal tail being unstructured. *D*, heteronuclear $^{15}N\{^1H\}$ NOEs are positive but smaller than expected for an 8-kDa protein, indicating fast internal motion in the three-helix bundle. *Red lines* indicate average values for the 225–242 and the 243–294 segments, respectively.

eronuclear $^{15}N\{^1H\}$ NOEs values (Fig. 1*D*), although the latter are lower than those of a rigid protein, thus indicating the occurrence of fast internal motions.

The Complex of NPM1-C70 with c-MYC G-quadruplex DNA—NPM1-C70 binds a DNA oligonucleotide resembling a specific sequence found at the NHE_{III} region of the c-*MYC* promoter (11). This 27-mer region (called Pu27) of sequence 5′-T**GGGGAGGG**T**GGGGAGGG**T**GGGG**AAGG-3′ is known to form a parallel G-quadruplex structure in the presence of K^+ at physiological concentration (100–150 mM) (11, 18). Pu27 contains five runs of three or more consecutive guanines and therefore can, in principle, populate several G-quadruplex structures with different topologies (13). These multiple conformations are indeed observed in the free state as monitored by multiple sets of NMR signals in slow exchange in the one-dimensional ^1H NMR spectra (Fig. 2*A*). Moreover, Pu27 maintains different quadruplex topologies also when bound to NPM1-C70 (Fig. 2*A*). The structural heterogeneity of the Pu27 sample complicates the analysis of its interaction with NPM1-C70. However, recently, Patel and coworkers (44) showed that a c-*MYC*-derived shorter oligonucleotide of 24 residues containing a guanine to inosine substitution in one of the guanine runs populates only one of the possible G-quadruplex conformations (called Pu24I; 5′-T**GAGGG**T**GGI**G**AGGG**T**GGGG**AA-GG-3′). Therefore, we decided to use this oligonucleotide for further studies. The assignment of the ^1H nuclei of Pu24I was provided by courtesy of Anh Tuan Phan and Vitaly Kuryavyi (44). First, by comparing the one-dimensional ^1H NMR spectra of Pu24I in the free state and bound to NPM1-C70, we confirmed that Pu24I displays only a single G-quadruplex topology that is retained after NPM1-C70 binding (Fig. 2*B*). Then, to assess whether Pu24I undergoes major conformational changes upon NPM1-C70 binding, we performed intramolecular NOE experiments for the bound state of Pu24I and compared it with its free state. To characterize the bound state, a ω_1-^{13}C-filtered, ω_2-^{13}C-filtered NOESY experiment was recorded in a two-dimensional plane (^1H-^1H plane), whereas, in the case of the Pu24I free state, a standard two-dimensional NOESY was performed. As shown in Fig. 2*C*, no major variations are visible in the superimposition of the two spectra, indicating that the G-quadruplex structure of Pu24I is maintaining its conformation when bound to NPM1-C70.

To further investigate the interaction of NPM1-C70 with Pu24I, we titrated ^{15}N-labeled NPM1-C70 with increasing amounts of unlabeled Pu24I. We observed both the appearance of new peaks and the disappearance of others, indicating the formation of a new species that exchanges slowly with the free protein on the NMR time scale, *i.e.* $< 10^{-2}\ s^{-1}$ (Fig. 3*A*). The ratio of the intensity of the signals for a given resonance in the two species changes linearly with the amount of Pu24I and reaches a plateau upon addition of stoichiometric amounts of Pu24I, consistent with a 1:1 protein·Pu24I complex. NPM1-

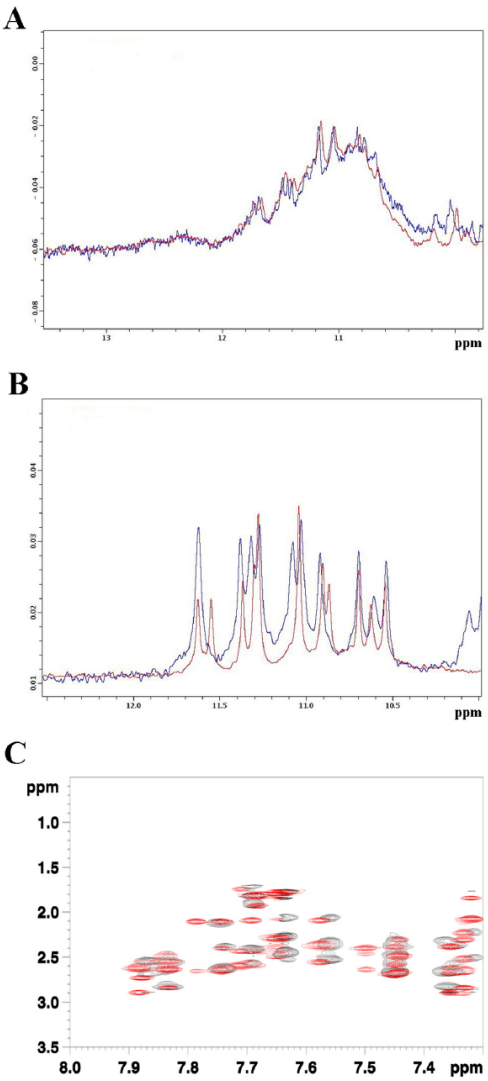

FIGURE 2. **Analysis of Pu27 and Pu24I G-quadruplex conformations.** *A*, superimposition of the one-dimensional ^{1}H NMR spectra of Pu27 in the free state (*blue*) and bound (*red*) to NPM1-C70. Both spectra were acquired at 700 MHz and at 290 K. *B*, superimposition of the one-dimensional ^{1}H NMR spectra of Pu24I in the free state (*blue*) and bound (*red*) to NPM1-C70. Both spectra were acquired at 700 MHz and 290 K. *C*, details of the superimposition between ω_1-^{13}C-filtered, ω_2-^{13}C-filtered NOESY experiments in a two-dimensional plane (^{1}H-^{1}H plane) of Pu24I·NPM1-C70 complex (*black*) and a classical two-dimensional NOESY of Pu24I in its free state (*red*). Both spectra were acquired at 700 MHz and 290 K.

C70 interacts with Pu24I mainly via hydrophilic residues located in helices H1 and H2 and exposed on the same side of the protein (Fig. 3, *B* and *C*). Interestingly, a few residues located on helix H3 also display chemical shift variations (Fig. 3, *B* and *C*). Because the latter are buried and interact with helices H1 and H2, we conclude that a strong coupling is present among the three helices.

In the protein·Pu24I complex, no chemical shift variations were detected for residues belonging to the NPM1-C70 N-terminal unstructured segment, indicating that this tail is not directly involved in the complex (Fig. 3*B*). This was confirmed by analysis of the ^{15}N heteronuclear relaxation data in the complex where the N-terminal tail is still characterized by high values for heteronuclear ^{15}N-^{1}H R_1 (Fig. 3*D*) and low values for heteronuclear ^{15}N-^{1}H R_2 (Fig. 3*E*) and heteronuclear NOEs (Fig. 3*F*), consistent with a natively unstructured state. Conversely, the folded C-terminal region of NPM1-C70 shows ^{15}N-^{1}H R_1 and ^{15}N-^{1}H R_2 values typical of a complex of 16-kDa molecular mass (in agreement with the sum of the molecular weights of NPM1-C70 and Pu24I) and positive values for heteronuclear ^{15}N{^{1}H} NOE values (Fig. 3, *D–F*). Both results suggest that the complex between NPM1-C70 and Pu24I is stable, consistent with the high affinity observed previously for the Pu27 oligonucleotide (11). The relaxation data analysis also pointed out that the heteronuclear ^{15}N{^{1}H} NOE values increase in the folded region of NPM1-C70, indicating an increasing rigidity for the three-helix bundle of NPM1-C70 upon binding of Pu24I (Fig. 3*F*).

Experimentally Restrained Molecular Docking of the NPM1-C70·Pu24I complex—To gain additional information on the structure of the complex, ^{13}C-filtered NOESY experiments were performed. Six intermolecular NOEs were detected between NPM1-C70 and Pu24I in the complex (Table 2). An example of assignment of NOE cross-peaks is shown in Fig. 4*A*. They involve Lys and Asn residues of the protein (whose side chains are free to rotate during the subsequent docking calculation) and sugar backbone protons on Pu24I. These distance restraints, together with the chemical shift perturbations defined as AIRs (also listed in Table 2), were used to calculate a structural model of the NPM1-C70·Pu24I complex within the data-driven HADDOCK docking program (see "Experimental "Procedures" for details about the protocol used). Importantly, HADDOCK calculations identified a single cluster of docking poses. In particular, 184 final complex structures were obtained at the end of the procedure with an root mean square deviation from the lowest energy solution of 1.5 ± 0.9 Å. As an example, the 20 lowest energy docking poses are reported in Fig. 4*B*.

As shown in Fig. 5*A*, the interaction surface involves one side of the three-helix bundle, the interacting residues all being located on helices H1 and H2, with a buried area of 1358.6 ± 92.2 Å^2. Based on the assignment of the intermolecular NOEs (Fig. 4*A* and Table 2), the interaction involves side chains of NPM1-C70 residues Lys^{250}, Lys^{257}, Asn^{270}, and Asn^{274}. On the Pu24I side, intermolecular NOEs indicate the involvement of protons of the backbone (mainly the deoxyribose ring) of nucleotides T1, G11, A12, G13, G15, and G23. From the structure model analysis, a linear stretch of nucleotides (from G11 to G16) located on one side of the G-quadruplex scaffold (Fig. 5*A*) intercalates into a groove formed by helices H1 and H2 (Fig. 5*B*). Interestingly, this stretch contributes to the formation of each of the three stacked guanine tetrads in the G-quadruplex

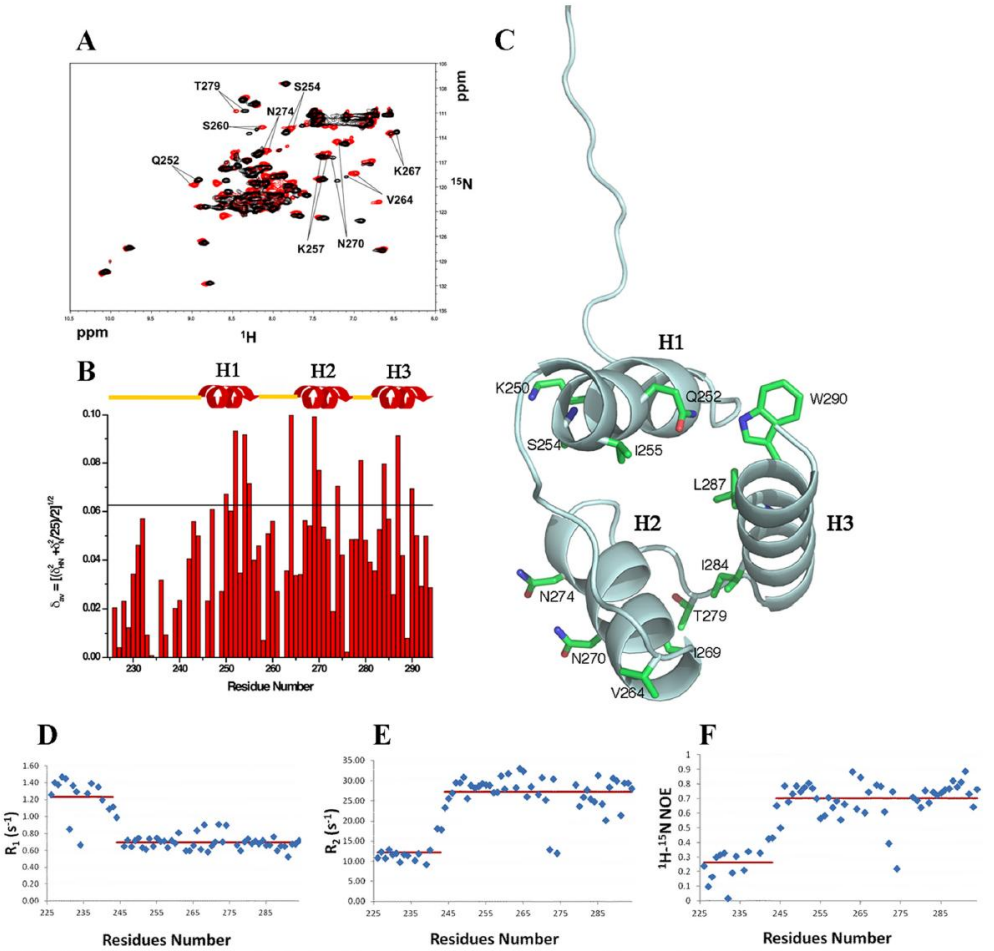

FIGURE 3. **Interaction of NPM1-C70 with the Pu24I G-quadruplex.** *A*, ^{15}N HSQC spectra of the protein before (*black*) and after addition of stoichiometric amounts of unlabeled Pu24I (*red*). Representative chemical shift variations are labeled, indicating relevant residues. *B*, chemical shift variations cluster in the three-helix bundle, whereas they are not found at the N-terminal 225–242 segment. The *horizontal black line* indicates the average chemical shift variation plus one standard deviation upon Pu24I addition. *C*, residues experiencing chemical shift variations higher than the average plus one standard deviation are highlighted on the structure of the protein. Residues belonging to helices H1 and H2 are solvent-exposed. A few hydrophobic residues belonging to helix H3 are also affected, indicating coupling between the helices upon Pu24I binding. Heteronuclear ^{15}N-^{1}H R_1 (*D*) and heteronuclear ^{15}N-^{1}H R_2 values (*E*) for NPM1-C70 in complex with Pu24I indicate that the N-terminal region flanking the three-helix bundle remains unstructured after Pu24I binding. *F*, the increase of heteronuclear ^{15}N{^{1}H} NOE values in the three-helix bundle upon Pu24I binding (see Fig. 1*D* for comparison) suggests increased rigidity.

scaffold. By analyzing the 20 lowest energy docking solutions, it appears that residues Lys^{250}, Lys^{257}, Asn^{270}, and Asn^{274} always participate in the formation of salt bridges or hydrogen bonds with backbone phosphates belonging to the G11–G16 linear stretch. In addition, residues Lys^{267} and Cys^{275} are always found at the interface where they contact phosphate groups in the G11–G16 stretch in all 20 lowest energy docking poses.

DISCUSSION

The propensity of NPM1 to interact with nucleic acids appears to be crucial for several functions that this protein plays both in the nucleoli and in the cytoplasm where a small fraction of the protein is always present due to continuous shuttling back and forth from the nucleus. For instance, a great deal of data show that NPM1 controls both ribosome assembly and

TABLE 2
Intermolecular NOEs and AIRs used for HADDOCK docking calculations
Gua, guanine.

Intermolecular NOEs		
NPM1-C70 atoms	***Pu24I atoms***	***Distance (Å)***
Lys 250-HG3/HG2	G13-H3'	4.01
Lys 250-HE2/HE3	G11-H5'	5.83
Lys 257-HD2/HD3	A12-H1'	5.53
Lys 257-HD3/HD2	T1-H72	4.03
Asn 270-HD22/HD21	G15-H5"	4.52
Asn 274-HD22/HD21	G23-H21	3.12

Ambiguous Interaction Restraints	
Active Residues of NPM1-C70	***Active Residues of Pu24I***
Glu245	Gua5
Lys248	Gua13
Lys250	Gua17
Gln252	Gua20
Ser254	
Ile255	
Lys257	
Val264	
Phe268	
Ile269	
Asn270	
Asn274	
Phe276	
Thr279	
Ile284	
Leu287	
Trp290	

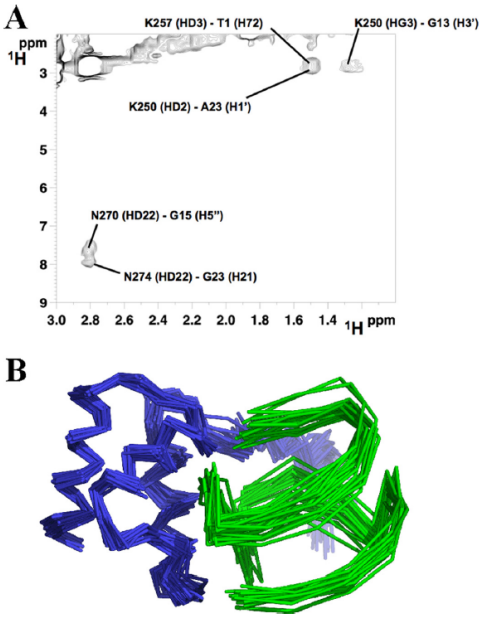

FIGURE 4. **Intermolecular NOEs and HADDOCK calculations.** *A*, an example of NOE cross-peak assignment from ω_1-^{13}C-edited, ω_2-^{13}C-filtered NOESY experiments in a two-dimensional plane (^{1}H-^{1}H plane) of the NPM1-C70·Pu24I complex acquired on a ^{13}C,^{15}N-labeled NPM1-C70·unlabeled Pu24I sample. *B*, the 20 lowest energy complex structures obtained by NMR data-restrained molecular docking calculations using the HADDOCK protocol. NPM1-C70 is represented in *blue*, and Pu24I is represented in *green*.

transport (9). Recently, it was also shown that NPM1 is selectively deposited on the mRNA body during polyadenylation, suggesting a putative role in a variety of post-transcriptional events, including splicing (45). A role of NPM1 in the control of gene transcription has been also suggested either through its association with several transcription factors at gene promoters (9) or through its direct interaction with the G-rich sequence found at the *SOD2* gene promoter (10). We further showed that the latter region is folded as a G-quadruplex at least *in vitro* and that G-quadruplexes are bound by NPM1-C70 with high affinity (11).

Based on these premises, we analyzed the three-dimensional structure of NPM1-C70 bound to G-quadruplexes. Among several oligonucleotides tested, we focused our attention on the interaction of NPM1-C70 with the G-quadruplex at the NHE_{III} region of the c-*MYC* promoter because this is the DNA sequence bound with the highest affinity (11) whose structure is known (40).

The structure of NPM1-C70 alone and bound to the DNA fragment was investigated by a combination of NMR data and docking calculations guided by experimental restraints. As shown in Fig. 5, NPM1-C70 binds the Pu24I G-quadruplex through a specific surface at the interface between helices H1 and H2. Several positively charged and polar residues establish interactions mainly with phosphate groups of a linear stretch of nucleotides that fits the small groove at the H1-H2 interface. Interestingly, this stretch contributes to the formation of the main G-quadruplex scaffold, whereas the interactions with nucleotides belonging to the connecting loops appear marginal. This may explain why NPM1 recognizes with comparable affinity several G-quadruplexes that differ in loop length and distribution (11).

Among the NPM1-C70 residues found at the interface with the Pu24I G-quadruplex, Lys^{257} and Lys^{267} are acetylated *in vivo* by p300 and deacetylated by SIRT1 (46). NPM1 acetylation results in dislocation of the protein from the nucleoli to the nucleoplasm where NPM1 interacts with transcriptionally active RNA polymerase II. Our finding that these two residues are at the interface with Pu24I DNA suggests that loss of nucleolar localization may be due to impaired DNA binding at the nucleoli coupled to acetylation.

Interestingly, we also found from our docking simulations that Cys^{275} is always located at the center of the surface buried by Pu24I. Cys^{275} is targeted and alkylated by the natural antitumoral compound (+)-avrainvillamide (47). Treatment of LNCaP or T-47D cells with (+)-avrainvillamide leads to an increase in cellular p53 concentrations and promotes apoptosis. It is therefore conceivable that these effects may be linked to (+)-avrainvillamide-mediated impairment of NPM1 nucleic acid binding efficiency.

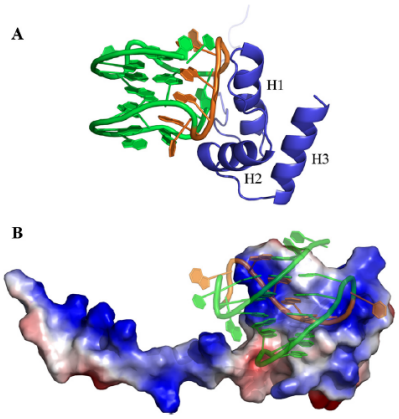

FIGURE 5. **Structural model of the NPM1-C70·Pu24I complex.** *A*, ribbon representation of the lowest energy model showing helices H1 and H2 of NPM1-C70 approaching the G-quadruplex "laterally" and interacting with a specific segment of the backbone (in *orange*). *B*, NPM1-C70 is represented with its electrostatic surface (*blue* for positive and *red* for negative), whereas Pu24I is shown in ribbon representation. The Pu24I structure is shown in *transparency* to highlight the small positively charged groove in between helices H1 and H2 that accommodates a stretch of Pu24I nucleotides (G11–G16; colored in *orange*). The long unstructured tail is also positively charged and may play a role in long range electrostatic interactions with the approaching oligonucleotide.

The N-terminal tail of the NPM1-C70 construct, although necessary for high affinity binding of any DNA tested (11), is natively unstructured both in the isolated domain and in the complex with Pu24I. This finding parallels previous observations in other systems. For instance, the affinity of the cAMP-responsive element-binding protein KID domain for the KIX domain of the cAMP-responsive element-binding protein-binding protein is significantly reduced when an unstructured portion of the domain that does not participate directly to the complex contact surface is deleted (48). Similarly, the interaction of splicing factor 1 with the large subunit of the U2 small nuclear RNA auxiliary factor ($U2AF^{65}$) is affected by flanking unstructured regions that do not physically contact the partner (49). This phenomenon, which is more frequent than previously anticipated, has been termed "flanking fuzziness" (50). Furthermore, it was shown recently that a large fraction of transcription factors are characterized by the presence of unstructured regions that flank the DNA-binding domain at one or both ends and that these regions may impact the affinity for specific or nonspecific DNA sequences (51).

What could the role of this unstructured N-terminal tail of NPM1-C70 be? Although this issue is still under investigation, it is conceivable that the presence of an unstructured segment adjacent to the interacting domain may provide a larger platform for long range electrostatic interactions or even transient physical contacts that facilitate the fine tuning of binding (see Fig. 5*B*) (50). This is supported by the observation that two concomitant Lys to Ala substitutions (Lys^{229}-Lys^{230}) in the unstructured segment of NPM1-C70 result in a dramatic decrease of global affinity (11).

Furthermore, exposed unstructured regions may be modified by post-translational modifications, driving regulatory changes. Several residues in the unstructured segment may be modified, such as Lys^{229} and Lys^{230} that are among the lysines acetylated by p300 and deacetylated by SIRT1 (46). A number of putative phosphorylation sites are also present in the tail, including Ser^{227}, Thr^{234}, Thr^{237}, and Ser^{242}; among these, the phosphorylations of Ser^{227} by PKC and of Thr^{234}-Thr^{237} by the Cdk1·cyclinB complex were experimentally validated (52, 53). Both acetylation and phosphorylation may therefore interfere with NPM1 nucleic acid binding and play a role in NPM1 activity and trafficking throughout the cell cycle.

Leukemia-associated NPM1 mutations cause dramatic destabilization up to total unfolding of the terminal three-helix bundle that is responsible for the aberrant cytosolic translocation of the protein. Scaloni *et al.* (6, 7) showed that a malleable "native-like" structure that accelerates folding is retained in the denatured state of the wild-type NPM1-C53 three-helix bundle and involves helices H2 and H3 (H3 is also the site of leukemic mutations), whereas helix H1 is totally unfolded in the denatured state. A possible strategy to rationally target NPM1 for the treatment of this type of leukemia might be that of developing a drug able to stabilize in the leukemic variant a native-like structure by altering through binding the folding-unfolding equilibrium in favor of the native state. Our folding studies suggested that such a drug should preferentially target helix H1. Because we have shown here that G-quadruplex DNA specifically binds a region in between helices H1 and H2 of NPM1-C70, we may attempt to rationally design aptamers or other smaller molecules that by mimicking the binding properties of G-quadruplexes to NPM1 C-terminal domain might stabilize a native-like state in the leukemic variant.

Acknowledgments—We sincerely thank Prof. Brunangelo Falini (University of Perugia, Italy) for the introduction to nucleophosmin and precious collaboration. The WeNMR project (European FP7 e-Infrastructure grant, Contract Number 261572, supported by the national GRID Initiatives of Belgium, Italy, Germany, the Netherlands (via the Dutch BiG Grid project), Portugal, UK, South Africa, Taiwan, and the Latin America GRID infrastructure via the Gisela project) is acknowledged for the use of web portals and computing and storage facilities.

REFERENCES

1. Grisendi, S., Mecucci, C., Falini, B., and Pandolfi, P. P. (2006) Nucleophosmin and cancer. *Nat. Rev. Cancer* **6**, 493–505
2. Falini, B., Sportoletti, P., and Martelli, M. P. (2009) Acute myeloid leukemia with mutated NPM1: diagnosis, prognosis and therapeutic perspectives. *Curr. Opin. Oncol.* **21**, 573–581
3. Falini, B., Gionfriddo, I., Cecchetti, F., Ballanti, S., Pettirossi, V., and Martelli, M. P. (2011) Acute myeloid leukemia with mutated nucleophosmin (NPM1): any hope for a targeted therapy? *Blood Rev.* **25**, 247–254
4. Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., Bigerna, B., Pacini, R., Pucciarini, A., Liso, A., Vignetti, M., Fazi, P., Meani, N., Pettirossi, V., Saglio, G., Mandelli, F., Lo-Coco, F., Pelicci, P. G., and Martelli, M. F.; GIMEMA Acute Leukemia Working Party (2005) Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N. Engl. J. Med.* **352**, 254–266
5. Grummitt, C. G., Townsley, F. M., Johnson, C. M., Warren, A. J., and Bycroft, M. (2008) Structural consequences of nucleophosmin mutations

in acute myeloid leukemia. *J. Biol. Chem.* **283,** 23326–23332
6. Scaloni, F., Gianni, S., Federici, L., Falini, B., and Brunori, M. (2009) Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. *FASEB J.* **23,** 2360–2365
7. Scaloni, F., Federici, L., Brunori, M., and Gianni, S. (2010) Deciphering the folding transition state structure and denatured state properties of nucleophosmin C-terminal domain. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 5447–5452
8. Hingorani, K., Szebeni, A., and Olson, M. O. (2000) Mapping the functional domains of nucleolar protein B23. *J. Biol. Chem.* **275,** 24451–24457
9. Lindström, M. S. (2011) NPM1/B23: a multifunctional chaperone in ribosome biogenesis and chromatin remodeling. *Biochem. Res. Int.* **2011,** 195209
10. Xu, Y., Fang, F., Dhar, S. K., St Clair, W. H., Kasarskis, E. J., and St Clair, D. K. (2007) The role of a single-stranded nucleotide loop in transcriptional regulation of the human *sod2* gene. *J. Biol. Chem.* **282,** 15981–15994
11. Federici, L., Arcovito, A., Scaglione, G. L., Scaloni, F., Lo Sterzo, C., Di Matteo, A., Falini, B., Giardina, B., and Brunori, M. (2010) Nucleophosmin C-terminal leukemia-associated domain interacts with G-rich quadruplex forming DNA. *J. Biol. Chem.* **285,** 37138–37149
12. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes at the human genome. *Nucleic Acids Res.* **33,** 2908–2916
13. Huppert, J. L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* **37,** 1375–1384
14. Qin, Y., and Hurley, L. H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90,** 1149–1171
15. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 11593–11598
16. Grand, C. L., Han, H., Muñoz, R.M., Weitman, S., Von Hoff, D. D., Hurley, L. H., and Bearss, D. J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth *in vivo*. *Mol. Cancer Ther.* **1,** 565–573
17. Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H., and Brooks, T. A. (2011) Demonstration that drug-targeted downregulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *J. Biol. Chem.* **286,** 41018–41027
18. González, V., Guo, K., Hurley, L., and Sun, D. (2009) Identification and characterization of nucleolin as a *c-myc* G-quadruplex-binding protein. *J. Biol. Chem.* **284,** 23622–23635
19. González, V., and Hurley, L. H. (2010) The cMYC NHEIII (1): function and regulation. *Annu. Rev. Pharmacol. Toxicol.* **50,** 111–129
20. Thakur, R. K., Kumar, P., Halder, K., Verma, A., Kar, A., Parent, J. L., Basundra, R., Kumar, A., and Chowdhury, S. (2009) Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.* **37,** 172–183
21. Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* **19,** 239–250
22. Sissi, C., Gatto, B., and Palumbo, M. (2011) The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie* **93,** 1219–1230
23. Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E., and Tulinsky, A. (1993) The structure of α-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* **268,** 17651–17654
24. Horvath, M. P., and Schultz, S. C. (2001) DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex. *J. Mol. Biol.* **310,** 367–377
25. Keller, R. (2004) *The Computer Aided Resonance Assignment Tutorial*, CANTINA Verlag, Goldau, Switzerland
26. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24,** 171–189
27. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319,** 209–227
28. Güntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278,** 353–378
29. Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J. Biomol. NMR* **4,** 171–180
30. Eghbalnia, H. R., Wang, L., Bahrami, A., Assadi, A., and Markley, J. L. (2005) Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. *J. Biomol. NMR* **32,** 71–81
31. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223
32. Bertini, I., Case, D. A., Ferella, L., Giachetti, A., and Rosato, A. (2011) A grid-enabled web portal for NMR structure refinement with AMBER. *Bioinformatics* **27,** 2384–2390
33. Case, D. A., Darden, T. A., Cheatham, T. E., 3rd, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossvai, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2010) *AMBER 11*, University of California, San Francisco
34. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8,** 477–486
35. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8,** 52–56, 29
36. Zwahlen, C., Legault, P., Vincent, S. J., Greenblatt, J., Konrat, R., and Kay, L. E. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119,** 6711–6721
37. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125,** 1731–1737
38. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921
39. Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001) Automated assignment of ambiguous nuclear Overhauser effects with ARIA. *Methods Enzymol.* **339,** 71–90
40. Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19,** 315–316
41. de Vries, S. J., van Dijk, M., and Bonvin, A. M. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5,** 883–897
42. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33,** 5984–6003
43. Orekhov, V. Yu., Pervushin, K. V., and Arseniev, A. S. (1994) Backbone dynamics of (1–71) bacterioopsin studied by two-dimensional ^{1}H-^{15}N NMR spectroscopy. *Eur. J. Biochem.* **219,** 887–896
44. Phan, A. T., Kuryavyi, V., Gaw, H. Y., and Patel, D. J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* **1,** 167–173
45. Palaniswamy, V., Moraes, K. C., Wilusz, C. J., and Wilusz, J. (2006) Nucleophosmin is selectively deposited on mRNA during polyadenylation. *Nat. Struct. Mol. Biol.* **13,** 429–435
46. Shandilya, J., Swaminathan, V., Gadad, S. S., Choudhari, R., Kodaganur, G. S., and Kundu, T. K. (2009) Acetylated NPM1 localizes in the nucleoplasm and regulates transcriptional activation of genes implicated in oral cancer manifestation. *Mol. Cell. Biol.* **29,** 5115–5127

47. Wulff, J. E., Siegrist, R., and Myers, A. G. (2007) The natural product avrainvillamide binds to the oncoprotein nucleophosmin. *J. Am. Chem. Soc.* **129,** 14444–14451
48. Zor, T., Mayr, B. M., Dyson, H. J., Montminy, M. R., and Wright, P. E. (2002) Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* **277,** 42241–42248
49. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell* **11,** 965–976
50. Tompa, P., and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33,** 2–8
51. Guo, X., Bulyk, M. L., and Hartemink, A. J. (2012) Intrinsic disorder within and flanking the DNA-binding domains of human transcription factors. *Pac. Symp. Biocomput.* **2012,** 104–115
52. Beckmann, R., Buchner, K., Jungblut, P. R., Eckerskorn, C., Weise, C., Hilbert, R., and Hucho, F. (1992) Nuclear substrates of protein kinase C. *Eur. J. Biochem.* **210,** 45–51
53. Okuwaki, M., Tsujimoto, M., and Nagata, K. (2002) The RNA binding activity of a ribosome biogenesis factor, nucleophosmin/B23, is modulated by phosphorylation with a cell cycle-dependent kinase and by association with its subtype. *Mol. Biol. Cell* **13,** 2016–2030