Biochimica et Biophysica Acta 1840 (2014) 2050–2059

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

G-quadruplex DNA recognition by nucleophosmin: New insights from protein dissection

Pasqualina Liana Scognamiglio [a,c,1], Concetta Di Natale [a,c,1], Marilisa Leone [d,b], Mattia Poletto [e], Luigi Vitagliano [d], Gianluca Tell [e], Daniela Marasco [a,b,*]

[a] Department of Pharmacy, University of Naples "Federico II", DFM-Scarl, 80134 Naples, Italy
[b] CIRPEB: Centro Interuniversitario di Ricerca sui Peptidi Bioattivi, University of Naples "Federico II", DFM-Scarl, 80134 Naples, Italy
[c] Center for Advanced Biomaterials for Healthcare@CRIB, Istituto Italiano di Tecnologia (IIT), 80125 Naples, Italy
[d] Institute of Biostructures and Bioimaging, CNR, 80134 Naples, Italy
[e] Department of Medical and Biological Sciences, University of Udine, 33100 Udine, Italy

ARTICLE INFO

Article history:
Received 16 December 2013
Received in revised form 13 February 2014
Accepted 18 February 2014
Available online 24 February 2014

Keywords:
Disordered protein region
Surface Plasmon Resonance
Helical stability

ABSTRACT

Background: Nucleophosmin (NPM1, B23) is a multifunctional protein that is involved in a variety of fundamental biological processes. NPM1/B23 deregulation is implicated in the pathogenesis of several human malignancies. This protein exerts its functions through the interaction with a multiplicity of biological partners. Very recently it is has been shown that NPM1/B23 specifically recognizes DNA G-quadruplexes through its C-terminal region.
Methods: Through a rational dissection approach of protein here we show that the intrinsically unfolded regions of NPM1/B23 significantly contribute to the binding of c-MYC G-quadruplex motif. Interestingly, the analysis of the ability of distinct NPM1/B23 fragments to bind this quadruplex led to the identifications of distinct NPM1/B23-based peptides that individually present a high affinity for this motif.
Results: These results suggest that the tight binding of NPM1/B23 to the G-quadruplex is achieved through the cooperation of both folded and unfolded regions that are individually able to bind it. The dissection of NPM1/B23 also unveils that its H1 helix is intrinsically endowed with an unusual thermal stability.
Conclusions: These findings have implications for the unfolding mechanism of NPM1/B23, for the G-quadruplex affinity of the different NPM1/B23 isoforms and for the design of peptide-based molecules able to interact with this DNA motif.
General observation: This study sheds new light in the molecular mechanism of the complex NPM1/G-quadruplex involved in acute myeloid leukemia (AML) disease.

1. Introduction

Nucleophosmin (NPM1, also known as B23, No38 and numatrin) is an abundant multifunctional protein, initially identified as a phosphoprotein, which is present in high quantities in the granular region of nucleoli [1,2]. NPM1/B23 is, however, capable of shuttling between nucleus and cytoplasm [3]. NPM1/B23 plays a plethora of functions including the regulation of ribosome biogenesis, chromatin remodeling, DNA replication, recombination, transcription, repair and the control of centrosome duplication [4,5]. This protein has been found over-expressed in tumors of different histological origins, including gastric, ovarian, bladder and prostate carcinomas and in various hematological malignancies [6–9]. Notably, NPM1/B23 has been identified as the most frequently mutated gene in acute myeloid leukemia (AML) patients, accounting for approximately 30% of cases [10–14].

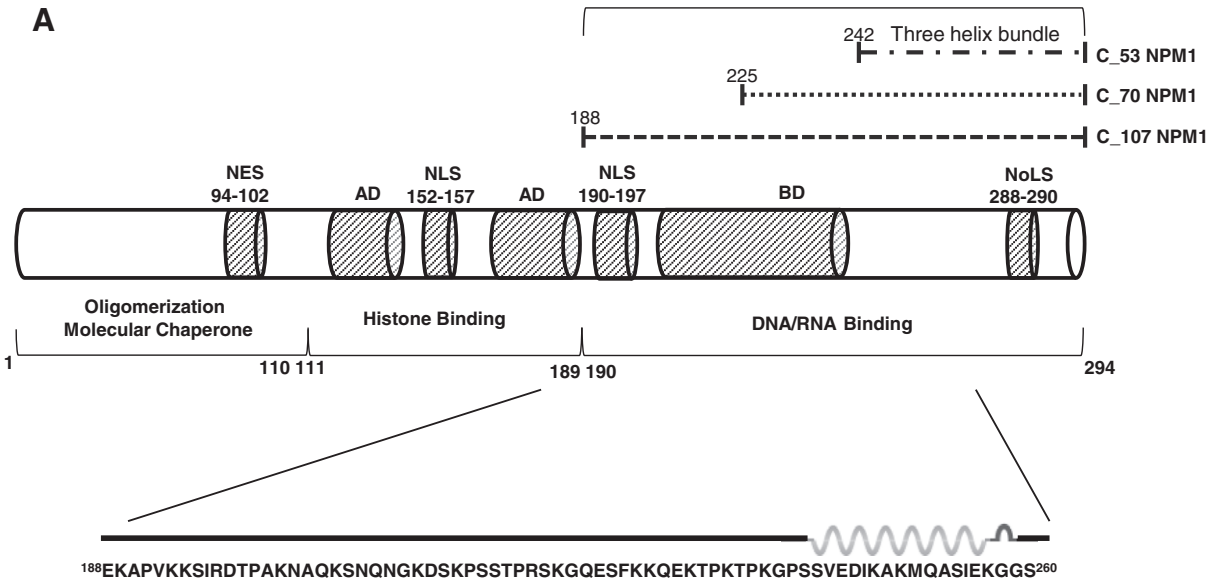
NPM1/B23 belongs to the nucleophosmin/nucleoplasmin family of proteins [15]. Three distinct isoforms of the protein have been reported to be expressed in human cells. B23.1, the longest and the most abundant, is also the best characterized isoform. B23.2 and B23.3 are splicing

Abbreviations: TIS, Triisopropylsilane; TFA, Trifluoroacetic acid; DMF, Dimethylformamide; DCM, Dichloromethane; HBTU, 1-H-Benzotriazolium, 1-[bis(dimethylamino)methylene]-hexafluorophosphate(1-), 3 oxide; HOBt, N-hydroxybenzotriazole; DIEA, Di-isopropylethylamine; Fmoc, Fluorenylmethoxycarbonyl; TCEP, Tris(2-Carboxyethyl) phosphine; HPLC, High Performance Liquid Chromatography; LC–MS, Liquid Chromatography–Mass Spectrometry; EDC, 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride; NHS, N-hydroxysuccinimide; NOESY, Nuclear Overhauser Enhancement Spectroscopy; RMSD, Root Mean Square Deviation; TOCSY, Total Correlation Spectroscopy; C_53 NPM1, NPM1 242–294; C_70 NPM1, NPM1 225–294; C_107 NPM1, NPM1 188–294

* Corresponding author at: Department of Pharmacy, University "Federico II", Via Mezzocannone, 16, 80134 Naples, Italy. Tel.: +39 081 2532043; fax: +39 081 2534574.
E-mail address: daniela.marasco@unina.it (D. Marasco).

[1] These are co-first authors.

http://dx.doi.org/10.1016/j.bbagen.2014.02.017
0304-4165/



B

188-225	EKAPVKKSIRDTPAKNAQKSNQNGKDSKPSSTPRSKGQ
225-243	QESFKKQEKTPKTPKGPSS
242-259 B23.1	SSVEDIKAKMQASIEKGG
242-259 B23.2	SSVEDIKAKMQASIEKAH
259-277 B23.1	GSLPKVEAKFINYVKNCFR

Fig. 1. Schematic representation of NPM1 protein. (A) The modular structure of NPM1, functional, structural domains and primary structure of C_107NPM1 are pointed out and (B) primary sequences of C_107NPM1-based peptides were analyzed in this study.

variants that lack the C-terminal 35 amino-acids and a 29 amino-acid stretch (residues 195–223) in the basic region, respectively [16].

Structural characterizations of B23.1 have shown that the protein is endowed with a modular structure (Fig. 1). The N-terminal domain extends for approximately 100 residues and displays an eight-stranded beta-barrel fold. Five N-terminal domains oligomerize to form a crown-shaped pentamer [17]. Two or more of these pentamers may associate to form decamers or higher oligomeric species [18]. The central portion of NPM1 is characterized by the presence of two acid domains (residues 119–133 and 161–188) and a basic region (residues 198–239). The C-terminal domain (CTD) of B23.1 forms a globular structure consisting of a three helix bundle. The destabilization of this structural unit abolishes the nucleolar localization of the protein [19]. The C-terminal region of the protein has also been the subject of a number of intriguing investigations aimed at unveiling its folding mechanism [20–23]. Despite the structural heterogeneity of the different regions of the protein, recent biophysical studies have highlighted their mutual stabilization upon treatments with temperature or chemical denaturants [21,24].

The analysis of the NPM1/B23 sequence also shows that the protein shuttling abilities between intracellular compartments rely on several signaling sequences: i) a NES (nuclear export) motif with two leucines in the N-terminal domain, ii) NLSs (nuclear localization) between the two acidic stretches and, iii) a NoLS (nucleolar localization) with aromatic-rich residues at the C-terminal domain [20]. NPM1/B23 exerts its functions essentially through interactions with a variety of biological partners. NPM1 interactions with other proteins (e.g. p53, p14arf, Fbw7γ, APE-1) [25–29], ribosomal proteins RPL5, RPS9, RPL23, HIV proteins REV TAT) [30] are generally mediated by the N-terminal domain, whereas the C-terminal moiety regulates the binding of DNA/RNA.

NPM1 was shown to bind both DNA and RNA oligonucleotides, with a preference for single stranded structures with respect to those double-stranded, independently of the sequence [31,32].

Recently, it is has been shown that the C-terminus of NPM1 is able to specifically recognize G-quadruplex DNA motifs [33,34]. This finding is of great interest since these motifs are attractive targets for tumor treatment for their selective localization on promoters of several oncogenes and on telomeric regions [34,35]. In particular, it has been demonstrated that the NPM1 C-terminal domain recognizes in vitro a well-characterized G-quadruplex forming sequence present at the NHEIII region of the c-MYC oncogene promoter [33]. This promoter region adopts a G-quadruplex structure both in vitro and in vivo, regulating up to 90% of the total c-MYC transcriptional rate [36]. NMR studies have shown that this G-quadruplex is recognized by NPM1 primarily by residues belonging to the helices H1 and H2 of the three-helix bundle. Nevertheless, mutagenesis analyses have shown that a fundamental contribution to the G-quadruplex binding is provided by the residues Lys 229 and Lys 230 [34]. Intriguingly, the region embedding these residues is disordered in the NMR structure of the complex. Attempts to explain these puzzling findings were based on the observation that the presence of an unstructured segment adjacent to the interacting domain could provide a larger platform for long-range electrostatic interactions or transient physical contacts [34].

In order to gain further insights into this intricate recognition mechanism, here we evaluated the ability of individual fragments of NPM1 to bind this G-quadruplex. Interestingly, the dissection of NPM1 led to the identification of distinct protein segments that present high affinity for this motif. The structural characterization of these fragments provides new clues about the folding/unfolding process of the triple helix bundle