SAPIENZA
UNIVERSITÀ DI ROMA

PhD course in Biochemistry
XXXI Cycle (2015-2018)

Structural and biochemical characterization of ribosome small subunit-dependent GTPase A (RsgA) from *Pseudomonas aeruginosa*

Docente guida
Prof. Carlo Travaglini-Allocatelli

PhD Student
Serena Rocchio

Tutor
Dr. Adele Di Matteo

PhD coordinator
Prof. Stefano Gianni

Table of contents

1. Introduction

1.1 Bacterial ribosome

Ribosomes are essential macromolecular machines that translate the genetic information into functional proteins. They are particles of more than 2.3 MDa with a diameter of about 20 nm, composed of 65% ribosomal RNA (rRNA) and 35% ribosomal proteins (r-proteins). The bacterial ribosome sediments, during ultracentrifugation, as a 70S particle composed of a small subunit (30S) and a large subunit (50S) (Ramakrishnan et al., 2002). Each subunit is therefore defined by a sedimentation coefficient, which reflect its relative mass, structure and composition differences (Connolly et al., 2008). The small subunit contains 21 ribosomal proteins and a 16S ribosomal RNA (rRNA), whereas the large subunit is made up of 34 proteins and two rRNAs: the 23S and 5S (Shajani et al., 2011) (Figure 1.1).

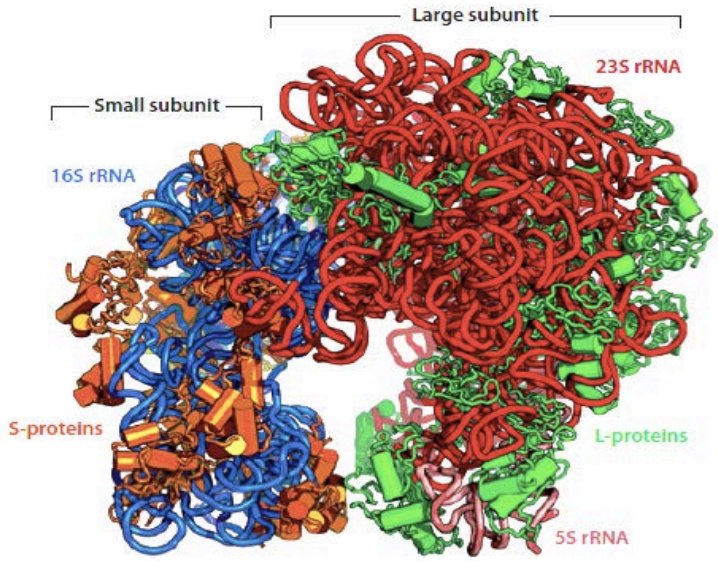

Figure 1.1 *E. coli* 70S ribosome. The small subunit (30S) is shown on the left, with the 16S rRNA in blue and the small subunit proteins (S-proteins) in orange. The large subunit (50S) is shown on the right, with the 23S rRNA in red, the 5S rRNA in pink and the large subunit proteins (L-proteins) in green.
[Figure adapted from Shajani et al., 2011]

The 30S subunit binds mRNA during translation initiation and is mainly responsible for the mRNA decoding function. The 50S subunit contains the peptidyl transferase center and catalyzes peptide-bond formation. The 30S subunit is composed of three domains: the body (5' domain), the platform region (central domain), and the head (3' domain), whereas the 50S subunit principal features are the central protuberance, the L1 arm (L1 stalk) and the L7–L12 region (L7-12 stalk) (Figure 1.2). Ribosome has three functional sites designated as A (aminoacyl), which accepts the incoming aminoacylated tRNA; P (peptidyl), which holds the tRNA with the nascent peptide chain; and E (exit), which holds the deacylated tRNA before it leaves the ribosome (Ramakrishnan et al., 2002).

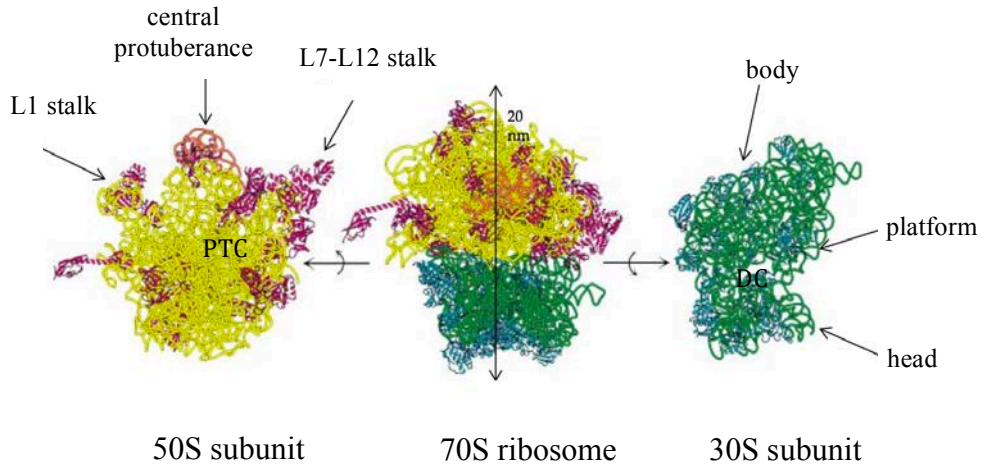

Figure 1.2 Bacerial ribosome. The functional 70S in the middle, the 50S subunit on the left (rRNA in yellow and r-proteins in magenta) and 30S on the right (rRNA in green and r-proteins in cyan). Features in the 50S subunit include the central protuberance, L1 arm (L1 stalk) and L7–L12 region (L7-12 stalk) and the peptidyl transferase center (PTC). In the 30S subunit, these include the head, body and platform as well as the decoding center (DC).

[Figure adapted from Schuwirth et al., 2005]

Structural information on the 70S ribosome (Yusupov et al., 2001; Noeske et al., 2015), the 30S (Wimberly et al., 2000; Schluenzen et al., 2000) and the 50S subunits (Ban et al., 2000) have contributed to shed light on many aspects of ribosome architecture and function (Ramakrishnan et al., 2001).
Together, available structures provide a great deal of information about protein-RNA interactions in each subunit, as well as the details of the interaction of the ribosome with ligands such as initiation and elongation factors, mRNA, tRNA and antibacterial drugs (Noller et al., 2005; Carter et al., 2000; Ramakrishnan, 2001 et al.; Steiz et al., 2003). Despite the acquired knowledge on the ribosome structure and function, ribogenesis (i.e. the processes leading to a functional ribosome in cell) is far to be completely understood even if dozens of proteins involved in ribosome maturation have been identified and many genetic, biochemical and structural data have been accumulated up to now (Goto et al., 2013). To reach a complete picture of how the ribosome get its functional state in vivo is of profound interest both from a biological perspective as well as from a pharmacological point of view, since the ribosome represents an important drug target.

1.1.1 Bacterial ribosome assembly

Ribosome biogenesis is a central cellular program that accounts for a

significant fraction of the energy budget for rapidly growing bacteria, and is an essential process in all living cells. Due to the complexity of this process, understanding how different components of the bacterial ribosome come together and organize themselves remains a daunting challenge.

The in vitro assembly of the small and large subunits from individual components was achieved over 40 years ago (Nierhaus et al., 1991; Nomura et al., 1970; Traub and Nomura, 1969). This pioneering work, led primarily by the Nomura and Nierhaus laboratories, demonstrated that the information for the assembly of these macromolecular complexes resided within the components of the ribosomal subunits themselves. Nevetheless, the non-physiological conditions required for ribosome assembly and the slow kinetics of this process *in vitro* indicated that additional factors are required *in vivo*. Indeed, in bacteria, the cytoplasmic assembly of ribosomes is facilitated by many cofactors, that include ribonucleases, RNA helicases, chaperones, ATPases, GTPases and ribosomal RNA (and r-protein) modification factors (Wilson and Nierhaus 2007). Although the specific role of many of these factors is still unclear, the deletion of genes encoding many of them causes accumulation of precursor rRNAs and immature ribosomal subunits and therefore affects ribosomal assembly (Wilson and Nierhaus 2007; Connolly and Culver 2009). The most notable and enigmatic of these proteins are the GTPases (Brown et al., 2005). Indeed, several GTPases have been correlated with ribogenesis in bacteria like RsgA, Era and YqeH in the 30S maturation and YihA, RbgA, Der and Obg in the 50S maturation (Goto et al., 2013).

Ribogenesis involves different and coordinated events: i) the transcription, processing, and modification of rRNA; ii) the translation and modification of ribosomal proteins; iii) the proper folding of rRNA and ribosomal proteins;

iv) the binding of ribosomal proteins and v) the binding and release of assembly factors. Many of these steps are coupled and occur simultaneously during rRNA transcription through an alternating series of RNA conformational changes and protein-binding events (Karbstein et al., 2007; Williamson et al., 2005; Holmes et al., 2005).

1.1.2 30S subunit maturation

The assembly of the 30S subunit is a multistep process (Figure 1.3) that starts with transcription of the 16S ribosomal RNA (rRNA) and the synthesis of the ribosomal proteins (r-proteins). Folding of the 16S rRNA starts before transcription is completed. This process is coupled with modifications to the RNA and processing of the precursor sequences (Shajani et al., 2011; Connolly et al., 2009). The binding of the 21 r-proteins to the 16 rRNA stabilizes rRNA and suppresses its misfolding (Hosokawa et al., 1966; Traub and Nomura, 1968, 1969; Woodson et al., 2008, 2011).

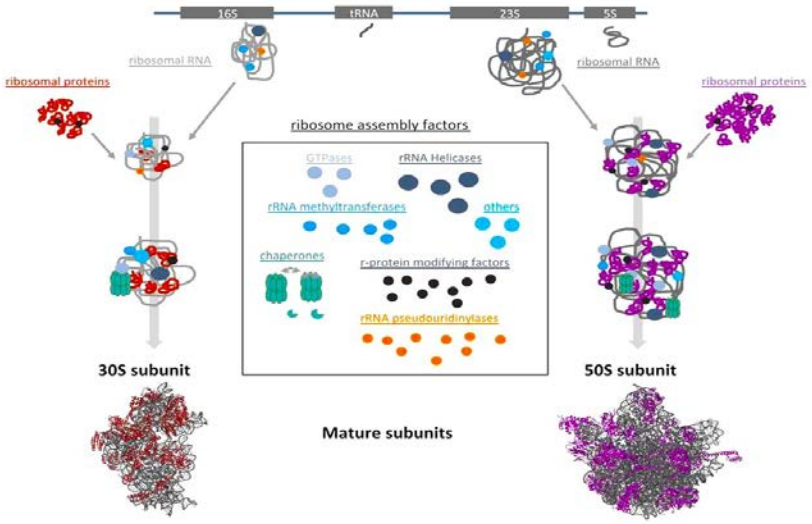

Figure 1.3 Biogenesis pathways of the bacterial ribosomal subunits 30S (left; pdb code 2AVY; RNA shown in gray and r-proteins in red) and 50S (right; PDB code 2AW4; RNA colored gray and r-proteins purple). The biogenesis of these subunits start with the transcription of primary rRNA transcripts, which contain 16S, 23S and 5S rRNA sequences and proceeds through a series of ill-defined steps. Several ribosomal biogenesis factors, shown in the central box, facilitate the ribosomal assembly through a coordinated series of maturation events.

[Figure adapted from Connolly et al., 2009]

Vintage experiments by Nomura (Hosokawa et al., 1966; Traub and Nomura, 1968, 1969) and more recent experiments from the Williamson and Woodson laboratories (Talkington et al. 2005; Adilakshmi et al. 2008) have defined the hierarchy and kinetic pathway of the 21 r-proteins binding to mature 16S rRNA in vitro. They have shown that six primary r-proteins bind the naked 16S rRNA, while secondary r-proteins require one or more primary r-proteins binding. Finally, the tertiary r-proteins bind after a temperature-dependent conformational step that is dependent on the binding of secondary r-proteins (Traub and Nomura, 1968; Shajani et al., 2011) (Figure 1.4).

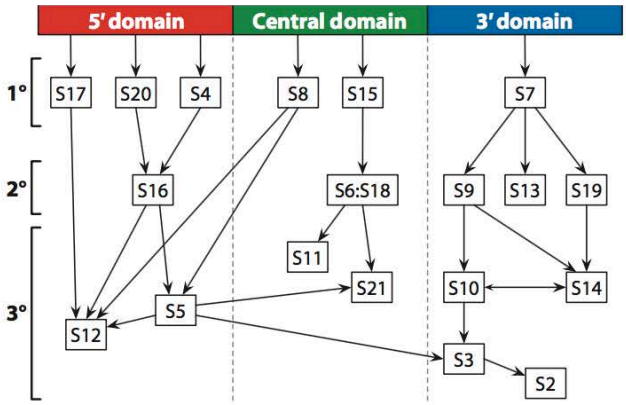

Figure 1.4 The Nomura assembly map. A few proteins, referred to as primary binding proteins (1°), bind directly to the nascent 16S rRNA. The binding of secondary binding proteins (2°) depend on primary binders wheras the tertiary binding proteins (3°) are sequential to secondary binders. The map is divided into 5′ (*red*), central (*green*), and 3′ (*blue*) domains on the basis of binding position relative to the 16S rRNA.
[Figure adapted from Shajani et al., 2011]

This hierarchy in protein binding leads to a cooperative assembly, ensuring each complex forms completely. Cooperativity mostly arises from structural changes in the 16S rRNA induced by the progressive addition of proteins (Culver et al., 2003). The 30S proteins make few base-specific contacts with the rRNA, but recognize the shape of the folded RNA (Brodersen et al., 2002). Co-folding of the RNA and the r-proteins increases the specificity of the assembly (Williamson et al., 2005). Williamson and colleagues, monitoring the binding rates and activation energies of all the 30S ribosomal proteins simultaneously, revealed that the assembly can proceed along multiple alternative pathways, due to the existence of an ensemble of multiple intermediate states from distinct assembly pathways, with no evidence of a single rate-limiting step (Williamson et al., 2005; Talkington et al., 2005; Mulder et al., 2010). They also showed that protein binding to the rRNA drives conformational rearrangements that stabilize the native fold of the 30S subunit (Talkington et al., 2005). This paradigm has been strengthened by Woodson and coworkers who showed multiple early folding nucleation events and induced fit of protein–rRNA complexes (Adilakshmi et al., 2008).

In line with the co-transcriptional nature of ribosome assembly, r-protein binding rates follow the 5’ to 3’ directionality of rRNA transcription, such that r-proteins bind rapidly to the 5’ 16S rRNA domain that forms the 30S body, and more slowly to the 3’ domain that forms the 30S head (Talkington

et al., 2005). Overall the folding of the 5' and central domains is more robust than that of the 3' domain due to redundant and alternative assembly pathways, while assembly of the 3' domain follows a more restrictive pathway that is susceptible to interference and kinetic traps (Xu et al., 2010). Multiple assembly pathways increase the flexibility of the assembly process, while accessory factors and modification enzymes chaperone the late stages of assembly and control the quality of the mature subunits.

During 30S ribosomal subunit biogenesis, assembly factors prevent accumulation of misfolded intermediate states of low free energy that slowly convert into mature 30S subunits. Four protein factors, RsgA (also known as YjeQ), Era, RbfA and RimM, are involved in the late stages of the 30S subunit maturation (Jomaa et al., 2011; Leong et al., 2013; Guo et al., 2013; Jeganathan et al., 2015; Thurolow et al., 2016). Recent work indicates that these factors bind the 30S subunit at or near the decoding center and aid its folding (Lopez-Alonso et al., 2017; Razi et al., 2017; Sharma et al., 2015; Datta et al., 2007); however, the precise mechanisms and the functional interplay among them are very complex and are not yet well understood.

1.2 GTPases and ribosome biogenesis

Guanine nucleotide-binding proteins or G proteins are well-known molecular switches that control several key cellular events (Bourne et al., 2009). Extensive structural and biochemical studies have contributed to our current understanding of their roles in protein synthesis, signaling events leading to cell proliferation and differentiation, endocytosis, protein trafficking, cytoskeletal rearrangement and cell motility (Bourne et al., 2009).

During the genome sequencing revolution in late 1990s it became clear that bacteria harbored several GTPases that had no known function but were

homologous to proteins in eukaryotic organisms. Prior to discover that many of these proteins were implicated in ribosome biogenesis, several groups had shown that many of these uncharacterized GTPases were essential for bacterial growth and had potential links to cell cycle and metabolic pathways (Britton et al., 1998; Morimoto et al., 2002). Although these proteins were initially classified as part of the Ras superfamily of GTPases, nevertheless they are distinct as they have additional domains that are lacking in many of the small monomeric GTPases. Most of these extra domains mediate the interaction of the GTPases with the ribosome through direct binding to rRNA and/or to ribosomal proteins. Mutations affecting the ribosome-associated GTPases, as well as in many Ras superfamily GTPases in eukaryotes, have pleiotropic phenotypes indicating potential connections between the ribosome and the cell cycle, stress, cell growth and nutrient availability.

The GTPases involved in ribosome assembly (referred to as RA-GTPases hereafter) have been studied in multiple bacterial species and share some general features: 1) most of these proteins are essential for growth and, in cases in which null mutations have been made, the mutants show an impaired phenotype including reduced levels of 70S ribosomes in the cell, due to improper assembly of the individual subunits; 2) Many, but not all, RA-GTPases are conserved throughout evolution and homologs can be found in most eukaryotic genomes, including human; 3) The bacterial RA-GTPases interact with ribosome subunits, usually in a GTP-dependent manner. Among the RA-GTPases, RsgA, Era and YqeH are involved in the maturation of the 30S subunit maturation while YihA, RbgA, Der and Obg in the maturation of the 50S subunit (Goto et al., 2013).

1.2.1 General features of GTPases

Monomeric GTPases function as molecular switches, with the GTP-bound form corresponding to the "ON" state and the GDP-bound form to the "OFF" state. Although the bacterial RA-GTPases share many conserved features with traditional GTPases, some unique aspects are highlighted below.

GTPases have conserved motifs recognizable at the sequence and structural levels (Vetter and Wittinghofer, 2001). These include the G1-G2-G3-G4-G5 motifs with G2 and G3 contained in the so-called switch I and switch II regions. These motifs coordinate the binding of guanine nucleotides (G motifs) and the positioning of an Mg^{++} ion and a water molecule for efficient hydrolysis. The highly mobile switch regions, which show different conformations based on the GTP/GDP bound, are often the sites at which effector proteins bind to propagate the downstream GTP mediated signal events (Wittinghofer, Vetter, 2011). The intrinsic rate of GTP hydrolysis is usually low and additional factors are required to speed it up. These include factors such as GTPase activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs). For several RA-GTPases, the ribosome may act as a GAP factor and, in few cases, also a role as a GEF has been proposed, hovewer, the molecular details of these functions have not been fully characterized to date.

The conserved functional 18–20 kDa G domain has a common structure and switch mechanism and consists of a six-stranded β-sheet and five helices on both sides (Vetter and Wittinghofer, 2001). The conserved sequence elements surround the nucleotide-binding site and are designed as G1 (GxxxxGK[S/T]), G2 (T), G3 (DxxG), G4 (N/TKxD) and a weakly conserved G5 sequence motif (often SAK). The loop containing G1, named the P-loop (for phosphate binding) and originally termed the Walker A motif,

interacts with the β- and γ-phosphate groups of GTP. The loops containing G2 and G3, termed switch I and switch II, respectively, make contact with the γ-phosphate and undergo a structured/unstructured shift upon GTP hydrolysis. The aspartate of DxxG in G3 (also called the Walker B motif) makes a water-mediated contact to the Mg^{++} ion, which is required for GTP hydrolysis in most Ras-like and other G proteins. G4 and G5 are instead the major determinants for guanine-base specificity (Vetter and Wittinghofer, 2001). Comparison of available structures of GTPases, in both the di- and triphosphate bound state, confirmed a canonical interaction with nucleotides and led to the conclusion that the switch mechanism is also conserved (Tu et al., 2009, 2011; Foucher et al., 2012). The conformational changes during GTP/GDP transition are confined to switch I and switch II (Milburn et al., 1990; Wittingover and Vetter, 2011). However, in multidomain proteins, these regions are often located in the interdomain interface such that the GTP/GTP transition can induce changes in the relative orientation of G flanking domains (Wittingover and Vetter, 2011).

1.2.2 Circulary permutated GTPases

A subset of RA-GTPases contains a unique circular permutation of the GTP binding domain. These atypical circularly permuted GTPases (cpGTPases) are grouped into distinct subfamilies, represented by the proteins YlqF, YqeH, YjeQ and YawG (Leipe et al., 2002). In the aminoacid sequence of these proteins the occurrence of sequence motifs follows the order G4-G5-G1-G2-G3, instead of the usual order G1-G2-G3-G4-G5 observed in canonical GTPases (Anand et al., 2006). Despite such a variation at the primary sequence level, which should lead to different topological connections between secondary structure elements, the three dimensional fold is well preserved (Shin et al., 2004). Although there are some structural

differences between cpGTPases and traditional GTPases in terms of the GTP binding pocket, the circular permutation does not dramatically alter the way these proteins interact with guanine nucleotides. In these proteins the circular permutation of G region relocates Switch-II to the C-terminus and leaves it unfastened. Since nucleotide-binding and hydrolytic activity require Switch-II to be held rigidly, an additional domain or a super secondary structure able to the proper positioning of Switch-II is required (Anand et al., 2006). As a consequence of the only permutation observed in nature is the creation of a new C-terminus following the DxxG motif in Switch-II. Such feature of cpGTPases confers two advantages: first, the Switch-II is properly positioned and oriented to favour guanine nucleotide binding and hydrolysis. Second, this coupling allows the propagation of conformational changes, mediated by GTP hydrolysis and associated with Switch-II, to the C-terminal domain thus regulating its biochemical functions (Anand et al., 2006).

1.2.3 HAS family of GTPases and mechanism of GTP hydrolysis

All the GTPases involved in ribosome biogenesis are members of the hydrophobic aminoacid substituted (HAS) family of GTPases (Mishra et al., 2005). In classical GTPases, such as Ras, a conserved glutamine residue (Gln61 designated as the Gln^{cat}) is located in the switch II. Hydrolysis of GTP is due to a nucleophilic attack by a water molecule and the role of Gln^{cat} is to stabilize the transition state by orienting the relative positions of the nucleophilic water and the γ-phosphate (Mishra et al., 2005). The importance of the catalytic glutamine in GTP hydrolysis is well documented (Vetter and Wittinghofer, 2001) and its mutation to hydrophobic aminoacids is reported to be oncogenic in Ras. Because of this substitution, the HAS-GTPases are believed to use an alternative mechanism for GTP hydrolysis (Mishra et al., 2005).

Another feature that remains unclear for the RA-GTPases is how the subsequent stabilization of the transition state (TS) is achieved. In the Ras system, RasGAPs stimulate GTP hydrolysis by supplying an additional residue, the "arginine-finger" (Arg^{GAP}), into the active site, which is responsible for ~2,000-fold acceleration of the GTPase reaction by direct electrostatic stabilization of developing negative charges in the transition state (Scheffzek et al., 1997). The Arg^{GAP} is provided either *in cis* by the same molecule or *in trans* by a different molecule and the scenarios are well known (Mishra et al., 2005). Up to now, RA-GTPases seem to be an exception among Ras-related G proteins, as the search for an arginine-finger or an analogous catalytic element has been unsuccessful (Rodnina et al., 2009). It has been proposed that some RA-GTPases use a monovalent cation (M^+ ion) as a structural and catalytic cofactor (Kuhle et al., 2014). These RA-GTPases coordinate a M^+ ion next to the GTP-γ-phosphate in a conserved coordination shell, where it forms a structurally relevant component of the catalytic center. The M^+ ion adds another positive charge to the preorganized active site of GTPases that together with the invariant lysine of the P-loop and the Mg^{++} ion forms a triangle of positively charged moieties around the GTP molecule. Therefore, the M^+ ion would be in a suitable position to neutralize negative charges of the TS in the γ-phosphate as well as the designated leaving group (GDP). This suggests that the M^+ ion might function as the catalytic element that contributes to rapid GTP hydrolysis by providing electrostatic stabilization for the TS, in analogy to the arginine-finger in the Ras-RasGAP system or the M^+ ion in MnmE (Kuhle et al., 2014).

The cation-dependent GTPases can be placed into two distinct groups, according to their behaviour in vitro: those that are stimulated by potassium

ions but not by sodium ions (potassium-selective cation-dependent GTPases) and those that are stimulated by both potassium and sodium ions (sodium-accomodating cation-dependent GTPases) (Ash et al., 2012). Moreover, it has been shown that the intrinsic activities of many bacterial RA-GTPases, YqeH (Anand et al., 2010), Era (Rafay et al., 2012), RbgA (Achila et al., 2012) and Der (Foucher et al., 2012) are enhanced by potassium ions, wheras ObgE activity is stimulated by sodium ions (Gkekas et al., 2017).

1.2.4 Role of GTPases in ribosomal assembly

The next major advancement in the field of RA-GTPases will come from determining how these proteins participate in ribosome biogenesis at molecular level. Few possible mechanisms by which RA-GTPases could function in ribosome assembly have been proposed: i) RA-GTPases could serve to recruit other assembly factors during key points of the assembly process; ii) RA-GTPases may control rRNA structure by regulating the activity of RNA helicases; iii) RA-GTPases could serve as RNA helicases itself and directly unwind or refold rRNA during the assembly process; iv) RA-GTPases could serve to determine whether a particular step in assembly has been properly achieved prior to the occurrency of the following phase (Britton et al., 2009). Under this light, energy from GTP hydrolysis can be used to regulate delivery or removal of proteins to nascent ribosomes as well as to promote a conformational rearrangement within nascent ribosomes. Moreover, GTPases could act as reversible “placeholders”, regulating r-proteins binding to nascent ribosome, as nutrient and enviromental sensors, by perception the cell metabolic state reflected in the GTP/GDP ratio, or as “checkpointers” by preventinting the entry of the 70S into the translational processes if not properly mature (Karbstein et al., 2007). Therefore, RA-GTPases perform essential functions in the assembly of ribosome and a clear

picture of how they work at the molecular level is critical to understand the general process of ribosome biogenesis.

1.3 Ribosome small subunit-dependent GTPase A (RsgA)

Ribosome small subunit-dependent GTPase A (RsgA), also named YjeQ /YloQ/CpgA, is a ribosome assembly factor that intervenes during the late stages of 30S maturation (Daigle et al., 2002; Himeno et al., 2004; Jomaa et al., 2011). RsgA is broadly conserved among bacteria but absent in eukaryotes (Daigle et al., 2002; Leipe et al., 2002). RsgA has low intrinsic GTPase activity that is stimulated by the 30S subunit and the 70S ribosome, but not by the 50S subunit (Himeno et al., 2004). RgsA is found associated with ribosome at very low stoichiometry (1:200) *in vivo* (Daigle and Brown 2004) and *in vitro* binds stably to the 30S subunit in the presence of GDPNP, a nonhydrolyzable GTP analogue, but not GTP or GDP (Daigle and Brown 2004; Himeno et al., 2004). In the presence of GDPNP-RsgA, 70S ribosomes dissociated into their subunits, suggesting an intersubunit localization of the factor (Himeno et al., 2004).

1.3.1 Structural features

RsgA belongs to the TRAFAC (translation factors) class of GTPases and, within this class, to the YlqF/YawG sub-family. YlqF/YawG members share the common characteristic of a circularly permutated GTP binding site in which the canonical G motifs (G1-G2-G3-G4-G5), mediating the nucleotide binding and hydrolysis (Bourne et al., 1991), are circularly permutated and adopt a G4–G5–G1–G2–G3 pattern (Shin et al., 2004; Levdikov et al., 2004;

Nichols et al., 2007). G1 is characterized by the consensus sequence GxxxxGKS/T in which the lysine side chain is responsible for phosphate binding; G2 has only a threonine residue highly conserved and is located in the so called switch I loop; G3, located in the switch II loop, contains the DxxG motif in which the aspartic acid residue is involved in Mg^{++} coordination; G4, characterized by the N/TKxD motif, is responsible for guanine specificity together with the G5 region that is only weakly conserved. In RsgA, the G domain is located within two additional regions: an oligonucleotide/oligosaccharide binding-fold domain (OB-domain) and a zinc-binding domain (Daigle et al., 2002; Nichols et al., 2007; Shin et al., 2004). The OB-fold consists of five antiparallel β-strands defining a β-barrel. The zinc-binding domain is composed of four helices in which the central pair, together with the intervening loop, defines a zinc-binding site. In RsgA structure the switches I and II and are ideally positioned to propagate conformational changes between the GTPase domain and the other two domains (Razi et al., 2017).

1.3.2 RsgA and the 30S subunit

High resolution structural information on the *E. coli* RsgA-30S complex, achived very recently by cryo-EM studies (Razi et al., 2017; Lopez-Alonzo et al., 2017), allowed the unequivocal positioning of RsgA with respect to the 30S subunit (Figure 1.5).

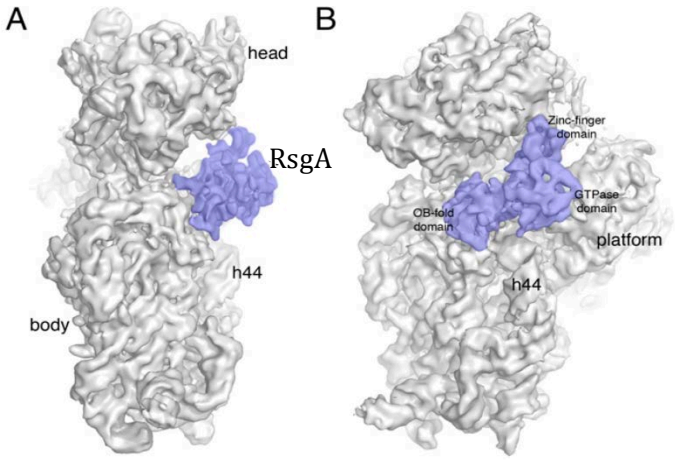

Figure 1.5 Cryo-EM structure of the 30S-RsgA complex. The side view (A) and the front view (B) of the 30S-RsgA complex are shown. Important landmarks of the 30S subunit, in gray, as well as the three domains of RsgA, in purple, are indicated.
[Figure adapted from Razi et al., 2017]

In the GTP-bound form, RsgA binds all three domains of the subunit. The OB-fold contacts the body of the 30S, whereas the zinc-finger domain anchors the protein to both the head and platform domains. Finally, the GTPase domain covers the decoding center almost completely and contacts the platform through a long loop (Figure 1.6). In particular, the OB-fold interacts with the 30S mainly through helix 18 and helix 44, the GTPase domain contacts helix 44, mainly through switch I and switch II, and helix 24 in the platform, and the C-terminal zinc binding domain anchors the protein to the head through helices 29 and 30 and to the platform by contacting helix 45 (Razi et al., 2017). In the GDP form of RsgA the OB-domain is the dominant interface for 30S contact. Accordingly, it has been proposed that the OB-domain acts as an anchoring point to tether RsgA to the 30S subunit (Lopez-Alonso et al., 2017). The interaction of RsgA with the 30S subunit places the GTPase domain in direct contact with the upper part of helix 44

that represents the ribosomal motif undergoing the largest conformational change upon RsgA binding.

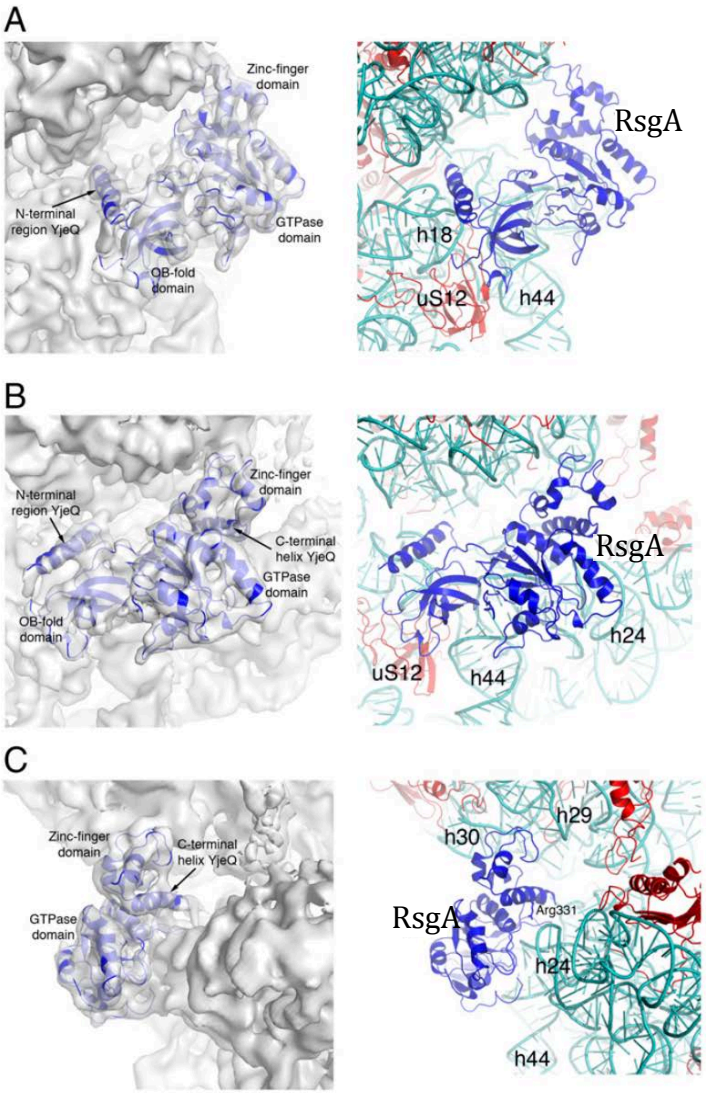

Figure 1.6. The cryo-EM structure and atomic model of RsgA bound to the 30S subunit. The zoom-in view of the density representing RsgA in the cryo-EM map of the 30S-RsgA complex with the atomic model of RsgA superimposed in the cryo-EM map is shown on the left. The equivalent close-up view of the atomic model of the complex with the assembling factor binding to the decoding center of the 30S subunit is shown on the right. The rRNA and r-proteins interacting with RsgA are labeled. The panels shown different views of the 30S-RsgA complex, the side view (A), the front view (B) and the platform view (C).
[Figure adapted from Razi et al., 2017]

As previously suggested, RsgA binds the 30S subunit close to the decoding center, in a position that is incompatible with that of all the three translation initiation factors (IF1, IF2, and IF3) (Carter et al., 2001; Simonetti et al., 2008), as well as A- and P-site tRNAs (Selmen et al., 2006) and the 50S subunit. Recent studies suggest functional interplay between RsgA and other factors involved in the 30S maturation process, like Era, RbfA, and RimM (Inoue et al., 2006; Campbell and Brown, 2008; Goto et al., 2011). Era is of particular interest because of the genetic interaction of this gene with RsgA: overexpression of Era suppresses defects in growth and ribosome maturation of a *rsgA*-null mutant (Campbell and Brown, 2008). Era performs its function in conjunction with RsgA; in particular they both assist the processing of the 3' end of the precursor 17S rRNA. The binding sites of Era and RsgA to the 30S subunit do not overlap, thus simultaneous binding of the two proteins is possible. Moreover, RsgA promotes the release of RbfA from the mature 30S subunit (Goto et al., 2011). RbfA is a small protein that binds the 30S subunit at the junction of the head and body, and its binding alter the position of helices 44 and 45 (Datta et al., 2007). The binding of RsgA to the 30S subunit has a stabilizing effect in the upper region of helix 44. This is the same rRNA motif that appears severely disrupted on RbfA binding (Razi et al., 2017). Therefore, it is likely that binding of RsgA to the 30S subunit

forces helix 44 back into the normal decoding position and triggers the release of RbfA. RsgA also distorts the binding site for IF1, which binds the 30S subunit in the cleft between helix 44 and protein S12 (Carter et al., 2001). In addition, the RsgA binding site partially overlaps with the interaction site of IF2 (Simonetti et al., 2008) and the C-terminal domain of IF3 (Dallas and Noller 2001). Therefore, RsgA might assists ribosome maturation by preventing premature formation of the translation initiation complex (Joomaa et al., 2011). These observations suggest that RsgA might be a general checkpoint protein in the late stage of the 30S subunit biogenesis, not only to release RbfA and other biogenesis factors from the nascent 30S subunit, but also to block the binding of players in translation initiation to the premature 30S subunit. Moreover, RsgA has a role in destabilizing kinetically trapped assembly intermediate as it induces local conformational changes in the 30S structure and disrupts binding of several tertiary r-protein binding, like S2, S3, S12 and S21 (Lopez-Alonso et al., 2017). However, to ultimately describe the existing functional relationships between these factors, additional structural, genetic and biochemical studies of RsgA and other assembly factors are needed.

1.3.3 GTPase activity

Although the GTPase activity of RsgA was characterized early on (Daigle et al., 2002, 2004), the role of this activity in the overall function of RsgA and its regulation remains unclear. The GTPase activity of RsgA is stimulated by over 100-fold in the presence of mature 30S subunits (Daigle et al., 2002, 2004; Himeno et al., 2004). Although the structure does not reveal what triggers this stimulation, it is possible that the specific conformation of helix 44 may stimulate the GTPase activity in RsgA. Recent structural information provides also important clues on the catalytic mechanism of RsgA and on the

30S mediated GTPase activity stimulation (Daigle et al., 2002, 2004; Himeno et al., 2004). Lopez-Alonso and coworkers showed that the catalytic residue in RsgA (performing function that parallel Gln61 in Ras) is an histidine (His248 in *E. coli* sequence) located in the switch 1 loop. The 16S rRNA, in particular the upper region of h44, plays a key role in mantaining His248 in position suitable for the catalitic events; the same role has been proposed for the 23S rRNA in activating EF-Tu (Lopez-Alonzo et al., 2017). Conformational changes in the GTPase domain caused by GTP hydrolysis, ready propagated to neighboring domains, may lead to reorganization of the interface of the complex, causing an overall decrease in the binding affinity of RsgA (Razi et al., 2017) and its release from the 30S subunit. Therefore, the GTPase activity of RsgA may functions as a sensor to facilitate the release of the protein from the 30S subunit once RsgA has performed its functions (Daigle and Brown 2004).

1.3.4 Role of RsgA in 30S subunit maturation

Biochemical and genetic studies over last decade have provided a growing body of evidence implicating a role for RsgA in the late stages of 30S biogenesis (Jomaa et al., 2011; Daigle et al., 2004; Himeno et al., 2004). Involvement of RsgA in ribosome maturation has been firsly suggested by the phenotype arising from *rsgA* deletion. Disruption of the gene for RsgA in *Escherichia coli* affects cell growth, subunit association and processing of the 16S rRNA (Himeno et al., 2004), whereas in *Bacillus subtilis* it has been shown to affects the growth and morphology of the bacterial cells (Campbell et al., 2005; Cladiere et al., 2006). *RsgA* deletion or inactivation of its ribosome small subunit-dependent GTPase activity provides *Escherichia coli* cells with resistance to high salt stress, suggesting a functional connection of the ribosome with the cellular mechanisms of salt tolerance. In addition to an

altered growth rate, *rsgA* deletion in *S. aureus* results in reduced virulence in mouse models (Cambell, 2006), implicating that RsgA is as a valid antibacterial target. Despite the acquired knowledge, the specific roles of RsgA in ribosomal assembly remain elusive. It is known that RsgA promotes the subunits assembly in vivo by changing the kinetics of the assembly process (Shanjani et al., 2011) and facilitating the incorporation of ribosomal proteins, especially those tertiary proteins with very slow binding rates (Talkington et al., 2005), during the late-stage 30S subunit maturation. RsgA plays a role as a checkpoint protein in the mature 30S subunit by testing the ability of the 30S subunit to perform proofreading before the subunit is released to the pool of active ribosomes. Another possible checkpoint role of RsgA is to block the binding of initiation factors to premature 30S subunits and ensuring quality control of the 30S subunit production (Guo et al., 2011). RsgA promotes dissociation of RbfA from the 30S subunit and as such facilitates the docking of the penultimate 16S rRNA helix, h44, on to the body of the 30S subunit (Guo et al., 2011; Jeganathan et al., 2015). Indeed, ribosomes purified from *rsgA* depleted strains are characterized by a distorted decoding center where h44/h45/h24 are not juxtaposed, preventing these particles from associating with the 50S subunit and engaging in translation (Jomaa et al., 2011). Moreover, aminoglycoside antibiotics, such as neomycin, which bind in the decoding center on the interface side of the 30S subunit, inhibit the ribosome dependent GTPase activity of RsgA (Campbell et al., 2005). In contrast chloramphenicol, which binds to the 50S subunit had no effect on its activity (Himeno et al., 2004). However, the role of the GTPase activity in these functions of RsgA remains unclear.

Interestingly, RsgA has been identified as a target for the stringent response nucleotides (p)ppGpp in *S. aureus* (Corrigan et al., 2016) and in *E.coli*

(Zhang et al., 2018). These studies revealed that RsgA, as well as other GTPases involved in ribogenesis, is inhibited by the stringent response nucleotides suggesting a possible mechanism by which the stringent response alarmones can control cell proliferation by interfering with ribosome assembly to inhibit cell growth and promote antimicrobial tolerance.

With the increased importance of ribosome biogenesis as a potential anti-microbial target, the chemical basis of RsgA activity becomes more important (Comartin et al., 2006; Nikolay et al., 2016; Stokes et al., 2014). Additional work and future structures of RsgA alone or in complex with GDP, GTP or transition state analogues as well as in complex with preribosomal particles or 30S-RAs particle, not only would clarify the molecular mechanisms of how this protein assists the maturation of the functional core of the 30S subunit (Razi et al., 2017) but also will be essential to explore the possibility to target RsgA for bacterial infection treatment.

1.4 Targeting ribosomal assembly as novel antibacterial strategy

The widespread and wasteful use of antibiotics in agricolture and clinical applications has strengthened the spread of resistant of and often multi-resistant bacteria via horizontal gene transfer (Davies et al., 2010). In addition to acquired resistance, some bacterial species have an intrinsic or innate resistance to different classes of antibiotics essentially carried out by three mechanisms: restricted uptake and efflux; drug inactivation and changes in targets (Lambert et al., 2002).

Existing antibiotics have limited chemical diversity and few mechanisms of action, making research on novel antibacterial targets a critical factor in

fighting multidrug resistance in bacteria (Poehlsgaard and Douthwaite, 2005). The ribosome and proteins involved in the translational process are among the main antibiotic targets. Crystal structures of naturally produced antibiotics and their semi-synthetic derivatives bound to ribosome have provided unparalleled insight into their mechanisms of action, and have also facilitated the design of more effective compounds for targeting multidrug-resistant bacteria. Many chemically diverse antibiotic compounds target the ribosome at surprisingly few locations, which results in overlap between many of their binding sites. Given the fundamental importance of the rRNA in the translation mechanism, it is not surprising that most ribosome inhibitors target the rRNA-rich surfaces on the 30S and 50S subunits (Figure 1.7). The 30S subunit is targeted by drugs that include tetracycline and aminoglycosides, which hinder the subunit in carrying out its function of deciphering the genetic information encoded in the mRNA (Poehlsgaard and Douthwaite, 2005). On the 50S subunit, most of the antibiotics binding sites cluster at or near the peptidyl-transferase centre (PTC), where peptide-bond formation occurs. The binding sites of macrolides are located near the PTC within the ribosomal exit tunnel, preventing elongation of most nascent chains (Wilson et al., 2014).

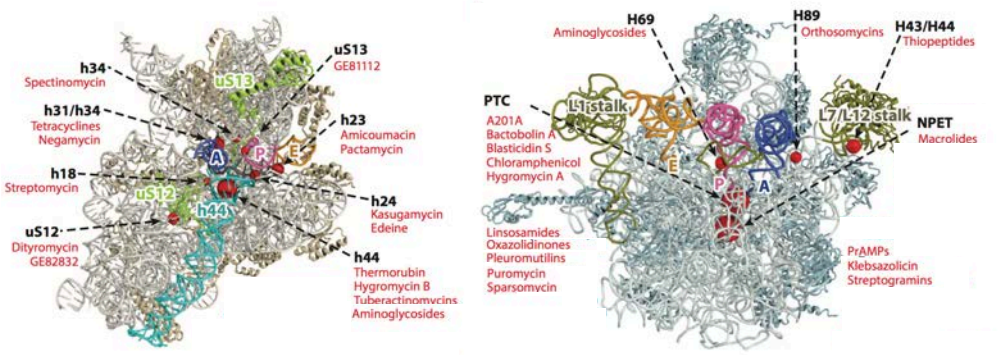

Figure 1.7. Overview of antibiotic binding sites on the 30S and 50S subunits. Major antibiotic binding sites on the 30S (A) and 50S subunits (B) are indicated and shown as red spheres; the names of antibiotics classes bound to each site are listed.
[Figure adapted from Lin et al., 2018]

The plethora of recent structures of antibiotics in complex with the ribosome has highlighted how resistance emerges through mutation or modification of the drug binding sites (Wilson, 2014; Poehlsgaard and Douthwaite, 2005). Indeed, structural studies provided insight about how existing drugs might be improved or novel drugs created. Derivatizing existing drugs to improve interaction at their binding site is an approach that has received considerable attention in the pharmaceutical industry (Wilson, 2014). Moreover, rational approaches based on crystallographic data have been applied to the design of new aminoglycosides and to the development of hybrid drugs (Poehlsgaard and Douthwaite, 2005). These novel drugs target the same sites as the parent compounds but with improved properties (Wilson, 2014).

A considerable challenge still remains to identify and target unexploited sites with novel drugs (Poehlsgaard and Douthwaite, 2005). The prospect of blocking ribosome function by preventing the assembly of subunits has come to light, supported by recent studies of non-ribosomal proteins involved in this process (Comartin and Borwn, 2006). Accumulating evidence indicates that the proteins like Era, Obg, RsgA, YlqF and RimM may be crucial to bacterial ribosome assembly and therefore they may represent novel targets for modern antibacterial drug discovery (Comartin and Borwn, 2006). These assembly factors are small, soluble and amenable to X-ray crystallography, to determine their structure, and cryo-electron microscopy to analyze their interaction with the ribosome. Furthermore, the pleiotropic effects of their inhibition may offer multiple ways to inhibit cell growth through the

impairment of a single target, an attractive feature that might limit the emergence of resistance (Maguire, 2009). Moreover, structures of additional antibiotic-ribosome complexes from diverse species will shed further light on the factors that govern species specificity, which should lead to the development of more selective or broader spectrum antimicrobials.

In conclusion, we find ourselves in a new era of ribosome and antibiotic research. With multi-drug resistance in bacteria being continuous threat to public health, there is enormous interest in rational approaches for the discovery of molecules belonging to new chemical classes and/or displaying novel mechanisms of action that could block protein translation (Comartin and Borwn, 2006).

1.5. Pseudomonas aeruginosa

Among the many bacteria that cause thread to human health *Pseudomonas aeruginosa (P. aeruginosa)* is of particular interest. *P. aeruginosa* is a Gram-negative bacterium that causes infections in individuals suffering from immune deficiency, severe burns and cystic fibrosis (Sousa and Pereira, 2014). Moreover, *P. aeruginosa* is responsible for 10-15 % of the nosocomial infections worldwide (Blanc et al., 1998). These infections are hard to treat due to the natural resistance of the *P. aeruginosa* strains, as well as to the remarkable ability of acquiring further mechanisms of resistance to multiple groups of antimicrobial agents (Stateva and Yordanov, 2009). *P. aeruginosa* is an highly adaptable organism; it can grow on a wide variety of substrates and alter its lifestyle in response to changes in the surrounding environment. During infection, *P. aeruginosa* generate a series of adaptive responses to facilitate its survival and colonization in the hostile host environment,

including the alteration of surface antigens, an increase in antibiotic resistances and the regulation of metabolic pathways (Hogardt and Heesemann, 2010). In addition, *P. aeruginosa* is particularly prone to maintain its survival in the host by promoting biofilm formation (Rybtke et al., 2015).

P. aeruginosa is intrinsically resistant to many structurally unrelated antimicrobial agents (Mesaros et al., 2007) and represents a phenomenon of bacterial resistance, since all known mechanisms of antimicrobial resistance can be seen in it: derepression of chromosomal AmpC b-lactamase (also known as cephalosporinase); production of plasmid or integron-mediated b-lactamases of different molecular classes; diminished outer membrane permeability; overexpression of various efflux pumps with wide substrate specificity; synthesis of aminoglycoside-modifying enzymes; and structural alterations of topoisomerases II and IV determining quinolone resistance (Stateva and Yordanov, 2009). Worryingly, often these mechanisms exist simultaneously conferring combined resistance to many strains (McGowan, 2006). The extensive use of antibiotics to treat *P. aeruginosa*, as well as the emergence of mutator variants, has generated the selective pressure to resistance development (Lambert et al., 2002), leading to serious therapeutic problems as well as the urgency for identification of new potential targets for the development of innovative antibacterial startegies (Stateva and Yordanov, 2009).

2. Aim of the thesis

The increase in antibiotic resistance among pathogenic bacterial strains presents a significant health threat. So far, the main efforts to combat antibiotic resistance are focused on the development of new antibiotics targeting protein biosynthesis. Ribosome, the large molecular machine responsible for this process, and proteins involved in the translational process represent ideal targets of molecules with antibacterial activity.

As discussed in the Introduction, it is now clear that ribosome assembly *in vivo* is an intricate and finely tuned process promoted by the action of several proteins acting as assembly factors, whose precise role is still largely unknown. Small GTPases represent the largest class of ribosome assembly factors in bacteria and are emerging as possible targets to be explored for the development of novel antibacterial strategies. Among them, of particular interest is the Ribosome small subunit-dependent GTPase A (RsgA). The characterization of RsgA from the human pathogenic bacterium *Pseudomonas aeruginosa* (*Pa*RsgA) represents the main focus of this PhD thesis.

RsgA is a late-stage ribosome biogenesis factor involved in the 30S subunit maturation, broadly conserved among bacteria but absent in eukaryotes. Despite the large amount of biochemical, structural and genetic data on RsgA achieved in the last decade, its mechanism of action is still not completely understood.

The main goal of this work is the determination of the *Pa*RsgA structure by X-ray crystallography. To date, no structure is available for RsgA from this opportunistic pathogen. This knowledge will allow investigate the molecular

features for the recognition of GDP and GTP as well as the key determinants for the mechanism of GTP hydrolysis.

Moreover, an accurate kinetic analysis of *Pa*RsgA interaction with GDP and GTP, together with a detailed functional characterization of *Pa*RsgA, provided the determination of substrates affinity and biochemical parameters of GTP hydrolysis.

The results obtained will pave the way for future experiments aimed at the characterization of the binding mechanism underlying ribosome recognition and to get key insight the GTPase activity of *Pa*RsgA in the presence of other assembly factors and/or the ribosomal particle.

3. Materials and methods

3.1 Protein Expression and Purification

*Pa*RsgA synthetic gene (Eurofins Genomics) was cloned into expression vector pET28a(+) between restriction sites NdeI and BamHI. *Escherichia coli* cells, strain BL21(DE3) (Biolabs, Ipswich, MA), transformed with the expression vectors, were grown to A_{600} ~ 0.6 in Luria-Bertani (LB) medium supplemented with kanamycin at 37°C. Expression was induced by addition of 1mM isopropyl-1-thio-ß-D-galactopyranoside (IPTG) and cells were further grown at 20°C for 16 hours. Harvested cells were lysed in a buffer containing 20mM Tris-HCl pH 8.0, 20mM $MgCl_2$, Benzonase nuclease (500U) (Sigma Aldrich), Protease Inhibitor Cocktail Tablet (Roche, Basel, CH) and sonicated. After centrifugation, supernatant was supplemented with 20mM Imidazole and 0.5M NaCl, and loaded on a HisTrap FF (GE Healthcare) equilibrated with the same buffer. Proteins were eluted by a linear gradient of 20mM Tris-HCl pH 8.0, 0.5M NaCl plus imidazole (from 20mM to 1M). Fractions containing *Pa*RsgA, were collected, diluted 10-fold and loaded on an anion exchange column (Q-Sepharose Fast Flow, GE Healthcare) equilibrated with 20mM Tris-HCl pH 8.0. After elution with NaCl, fractions containing the protein were monitored by UV absorption at 280 nm and SDS-PAGE, pooled and concentrated. Sample was buffer exchanged to 20mM Tris-HCl pH 8.0, 200mM NaCl (Storage Buffer) using a PD10 column (GE-Healthcare) and stored at -20 °C. For crystallization experiments protein was further purified by gel filtration on a fast-performance liquid chromatography (FPLC) column (Superdex 75 10/300; GE Healthcare).

3.2 Protein crystallization and structure solution

Purified *Pa*RsgA (concentrated up to 20 mg/ml) was subjected to crystallization procedure using vapour-diffusion technique. Crystallization trials have been performed using the automatic crystallization platform (Phenix-Art Robbins) and different commercial sparse-matrix screens. Promising hits were obtained by using the poly(acrylic acid sodium salt) 5100 as precipitating agent. To improve crystals shape, size and quality various parameters have been screened. High quality crystals were obtained by mixing 2 µl of the protein solution (20 mg/ml) with 3 µl of the reservoir solutions containing 28-30% poly(acrylic acid sodium salt) 5100, 0.1M Hepes pH 7.5, 0.02M $MgCl_2$ and 5% PEG 200. Crystals were cryoprotected increasing the poly(acrylic acid sodium salt) 5100 up to 34 % and flash-frozen. Diffraction data were collected at cryogenic temperature (100 K) at the XRD1 beamline of ELETTRA Synchrotron (Trieste, Italy). Data were indexed and integrated with XDS (Kabsch et al., 2010). The structure was solved by molecular replacement using the structure of RsgA from *Salmonella typhimurium* (pdb code: 2RCN; Nichols et al., 2007) as search model. Molecular replacement as well as refinement procedure was carried out using Phenix (Adams et al., 2010). For the refinement the initial model was subjected to rigid body refinement and, subsequently, to several cycle of refinement and manual visual inspecting and rebuilding using COOT (Emsley et al., 2010). Final model, containing a GDP molecule in the active site and a Zn atom, was refined to an Rwork = 0.248 and an Rfree= 0.286. Data collection and refinement parameters are reported in Table 1. The coordinates and structure factors have been deposited in the Protein Data Bank with accession number 6H4D. Structural superpositions were performed with COOT (Emsley et al., 2010). Figures were prepared with Chimera (Pettersen et al., 2004).

3.3 Preparation of nucleotide-free *Pa*RsgA

*Pa*RsgA protein was purified with the GDP bound to the active site (GDP-*Pa*RsgA). Preparation of nucleotide-free protein was carried out in two-step as reported in Ford et al., 2009. Briefly, GDP was firstly degraded by 1U/mg of alkaline phosphatase (Roche) and replaced with a non hydrolysable GTP-analogue, Gpp(CH_2)p (Sigma-Aldrich), by incubating the protein at 25 °C with 1,5 fold excess of Gpp(CH_2)p for 16 hours. After, 0.002U of snake venum phosphodiesterase (Sigma-Aldrich) per mg of Gpp(CH_2)p-*Pa*RsgA was then added to the solution and incubated one hour at 25°C. Free nucleotide was removed by PD10 column (GE-Healthcare) equilibrated with the Storage Buffer. Enzymes used to obtain the nucleotide-free RsgA were inactivated by 4 steps of rapid freezing-defreezing in liquid nitrogen. The complete removal of GDP was verified by ion-pair reversed-phase HPLC (RP-HPLC) analysis, using a Prevail^{TM} C18 column (GRACE), equilibrated with 100 mM KH_2PO_4/K_2HPO_4 pH 5.8, 10 mM tetrabutylammonium bromide, 8.5% acetonitrile buffer (HPLC buffer) and monitoring the nucleotides absorbance at 254 nm.

3.4 CD and fluorescence spectroscopy

All circular dichroism (CD) experiments were performed in sodium phosphate buffer 20mM pH 7.2, NaCl 150mM using a Jasco J710 instrument (Jasco Inc., Easton, MD, USA) equipped with a Peltier apparatus for temperature control. Static spectra were collected at 20 °C. Proteins concentration was 10 μM. Spectra were collected using a quartz cell with 1-mm optical path length (Hellman, Plainview, NY, USA) and a scanning speed of 100 nm/min. The reported spectra are the average of three scans.

The spectral contribution of buffers was 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 210 nm. Temperature was progressively increased, in 1°C/min steps, from 20 °C to 90 °C.

Thermodynamic stabilities of nucleotide-bound *Pa*RsgA and the nucleotide-free protein (6 μM) were determined at 25 °C by equilibrium urea-induced denaturation experiments, monitoring the change of intrinsic fluorescence emission. Equilibrium denaturations were carried on a Fluoromax-4 spectrofluorometer (Jobin-Yvon) monitoring the Trp fluorescence emission upon addition of urea between 300 and 400 nm at 25 °C in sodium phosphate buffer 20mM pH7.2, NaCl 150mM. Assuming a standard two-state model, the urea-induced denaturation transitions were fitted to the Eq. (1)

(1) $\Delta G_d = m_{D\text{-}N}(D - D_{50})$

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 solvent-accessible surface area from the native to the denatured state) and D_{50} is the midpoint of the denaturation transition. An equation which takes into account the pre- and post-transition baselines was used to fit the observed unfolding transition. The Kaleidagraph software was used for graphical representation.

3.5 Binding kinetic measurements

Binding kinetic experiments were performed on an SX-18 stopped-flow instruments (Applied Photophysics) using the nucleotide-free *Pa*RsgA and fluorescent GDP and GTP analogues (2′ or 3′ mant-GDP and mant-GTP (Life

Technologies)). The excitation wavelength was 355 nm (2.3 nm slit) and the fluorescence emission was measured using a 420 nm cut off filter at 25 °C. Fluorescence intensity of the mant moiety depends on the hydrophobicity of its environment, thus binding of protein to nucleotides leads to an increase of mGDP/mGTP intrinsic fluorescence. Kinetic parameters of *Pa*RsgA-nucleotides association were measured using 1 μM of nucleotides (concentration after mixing) and different concentration of protein (from 0.2 to 5 μM /after mixing) in 50mM Tris-HCl pH 7.5, 50 mM NaCl. The increase in fluorescence intensity of mGDP or mGTP upon *Pa*RsgA binding was monitored over time. Three to five separate curves were averaged for each concentration of protein, and the averaged curves were fitted to a single exponential function, yielding the observed rate constant k_{obs}. The association rate constants (k_{on}), for both GDP and GTP binding, were obtained from the slope of the plot of k_{obs} versus the protein concentration. Dissociation rate constants (k_{off}) were determined by displacement experiments in which nucleotide-free *Pa*RsgA (2 μM), pre-incubated with 5 μM of mGDP/GTP, was rapidly mixed with an excess of non-fluorescent GDP or GTP. Rates of m-nucleotides release were monitored as a decrease in fluorescence intensities. The resulting time traces were fitted on a single exponential, yielding k_{off}. The equilibrium dissociation constant K_D was calculated from the ratio of k_{off} and k_{on}. The GraphPad Prism software was used for graphical representation.

3.6 GTPase activity

The intrinsic GTPase activity of *Pa*RsgA was assessed measuring the GTP-GDP conversion over the time by ion-pair reversed-phase HPLC (RP-HPLC)

analysis.

To define the experimental condition in which perform the quantitative analysis, *Pa*RsgA (5 μM) was incubated with 400 μM GTP (steady state conditions) at 25°C in Tris 20 mM pH 8.0 containing different concentrations (up to 300 mM) of NaCl or KCl, as well as in absence of salt, in presence of 10 mM $MgCl_2$. The GDP production was measured after 120 minutes incubation. For the quantitative determination of catalytic parameters, 5μM of *Pa*Rsga was incubated with different GTP concentrations (50-1200 μM) at 25°C in Tris 20 mM pH 8.0, NaCl 200 mM, in presence of 10 mM $MgCl_2$. At different time intervals the reactions were stopped and the GDP production was monitored. Finally, to investigate the role of the cation over *Pa*Rsga GTPase activity, *Pa*RsgA (5 μM) was incubated with 400 μM GTP (steady state conditions) at 25°C in Tris 20 mM pH 8.0 containing 200 mM of different monovalent cations (NaCl, KCl, NH_4Cl, LiCl and CsCl), in presence of 10 mM $MgCl_2$. The GDP production was measured after 120 minutes incubation.

The reactions were stopped by adding the proteinase K (Sigma) and by heating 10 minutes at 95 °C. The denatured protein was removed by centrifugation and the supernatant applied to the Prevail™ C18 column. The GDP production was monitored by separation of the nucleotides (GTP and GDP) using a Prevail™ C18 column (GRACE, 150 x 4.6 mm) attached to an HLPC AZURA® system (KNAUER). The nucleotides were eluted using a buffer solution containing 100 mM KH_2PO_4/K_2HPO_4 pH 5.8, 10 mM tetrabutylammonium bromide, 8.5% acetonitrile buffer, as mobile phase. Nucleotides were detected by their absorbance at 254 nm and peak areas were converted to concentration using a standard curve derived from known

nucleotides concentrations. For the catalytic parameters estimation, the initial reaction rates (v_o) were obtained as the slope of the linear fit of the initial linear period of the reaction in plot showing the GDP concentration versus time. V_0s were plotted as a function of GTP concentration and fitted with the Michaelis-Menten equation. The GraphPad Prism and the Kaleidagraph software were used for graphical representation.

4. Results

4.1 *Pa*RsgA expression and purification

Rsga from *Pseudomonas aeruginosa* (*Pa*RsgA; Uniprot: Q9HUL3) is a soluble protein of 339 aminoacids (MW = 39241 Da), composed of three domains: a N-terminal Oligonucleotide/Oligosaccharide Binding (OB) domain, a central circularly permutated GTPase (cpGTPase) domain and a C-terminal Zinc-binding domain (Figure 1).

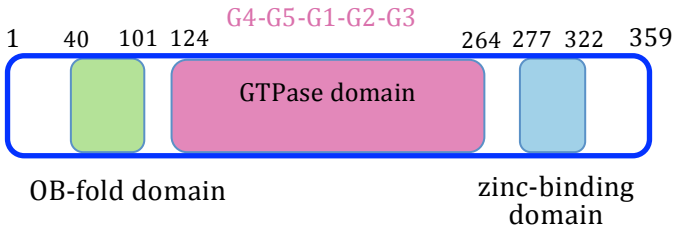

Figure 4.1 Domains organization of *Pa*RsgA. The N-terminal OB-fold domain is shown in green, the central cpGTPase domain in pink and the C-terminal zinc-finger domain in cyan.

*Pa*RsgA was expressed in *E. coli* BL21 (DE3) and purified to homogeneity (Figure 4.2A) by using various chromatographic approaches (affinity, ion-exchange and size-exclusion chromatography) (see Material and Methods). The UV spectrum of the purified protein shows a blunt peak around 280 nm (Figure 4.2B), suggesting the presence of a nucleotide bound to the purified protein.

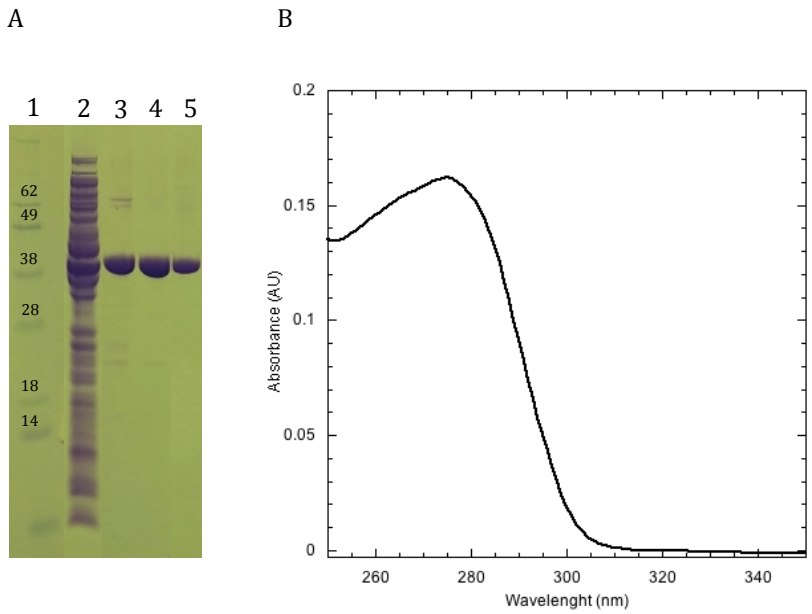

Figure 4.2 SDS PAGE of *Pa*RsgA purification (A). The diffrerent lines are representative of the purification steps, carried out with various cromatography approaches: 1) marker, 2) supernatant, 3) affinity 4) ionic exchange and 5) size exclusion cromatography. (B) UV-spectrum of PaRsgA co-purified with a nucleotide bound.

In order to confirm the presence of a co-purified nucleotide and to unveil its nature, we developed a protocol based on an ion-pair reversed-phase HPLC (RP-HPLC) analysis, using a Prevail^{TM} C18 column and monitoring the absorbance of the sample at 254 nm. Since the polarity of nucleotides increases with the number of phosphate groups, nucleoside triphosphates, unlike nucleoside mono- and diphosphates are weakly retained on reversed-phase chromatography with a conventional mobile phase. Conversely, by using an ion pairing reagent, such as tertrabutylammonium bromide (TBAB), nucleoside diphosphates were retained better than mono-phosphates although more weakly than nucleoside triphosphates. In this case, the separation is based on the formation of ion pair(s) between the positively charged ion-pair

reagent and the negatively charged nucleotide. Since this ion-pair RP-HPLC method enables the efficient separation of guanine nucleotides, it was possible to identify them by referring to the retention times established by the corresponding standards. Representative RP-HPLC chromatogram of purified *Pa*RsgA (Figure 4.3) shows a prominent peak eluting at the retention time corresponding to GDP. The calibration curve of the GDP standard (Figure 4.14) obtained by plotting the peak areas against the concentration of the analyte allows to estimate that *Pa*RsgA is purified with approximately 80% of GDP nucleotide bound.

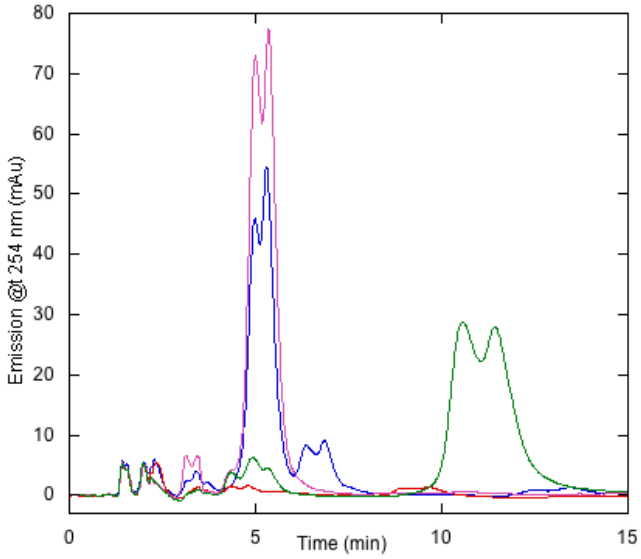

Figure 4.3 Identification of the nucleotide bound to *Pa*RsgA and validation of the nucleotide-free form production through reverse phase cromatograpy. Nucleotide-bound *Pa*RsgA is shown in blue, nucleotide-free *Pa*RsgA in red, GDP standard (50uM) in magenta and GTP standard (50uM) in green.

4.2 *Pa*RsgA structural characterization

4.2.1 *Pa*RsgA crystallization, data collection and structure determination

Purified *Pa*Rsga was concentrated up to 20 mg/ml and subjected to crystallization procedure using the vapour-diffusion technique. Crystallization trials have been performed using commercial sparse-matrix screens and an automatic crystallization platform (Phenix-Art Robbins) at 21°C. Initial promising hits (multiple small crystals or microcrystalline precipitates) were obtained with the INDEX screen (Hampton Research) where the poly(acrylic acid sodium salt) 5100 is present as precipitating agent (Figure 4.4A-B).

To improve crystal quality, we performed different experiments in which various parameters have been screened, such as protein concentration, the crystallization buffer and pH, the presence of additional salts or "additive agents" to slow down the crystallization speed as well as the protein/reservoir solution ratio in the drop. Moreover, different crystallization techniques, such as seeding performed with microcrystals and crystallization under oil (also at 4°C), were attempted in order to improve the diffraction quality. A wide array of *Pa*RsgA crystals were analysed through X-ray diffraction with different synchrotron light source (Elettra, Trieste; Bessy II Berlin; ESRF, Grenoble). The best diffracting crystal (2.9Å resolution) was obtained by mixing 2 µl of the protein solution (20 mg/ml) with 3 µl of the reservoir solutions containing 30% poly(acrylic acid sodium salt) 5100, 0.1 M Hepes pH 7.5, 0.02 M $MgCl_2$ and 5% PEG 200 (Figure 4.4C).

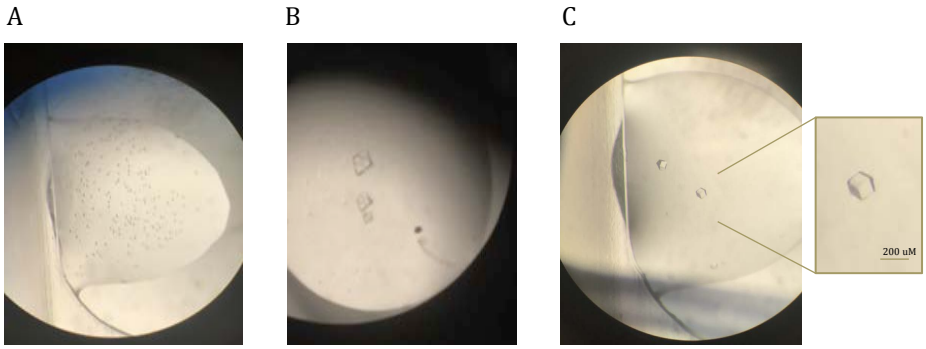

Figure 4.4. GDP-*Pa*RsgA microcrystals (A) and multiple crystals (B). GDP-*Pa*RsgA crystals (C) obtained by the optimized crystallization condition: 28-30% poly(acrilyc acid sodium salt) 5100, 0.1M Hepes pH 7.5, 0.02M $MgCl_2$ and 5% PEG 200 at 20°C.

Diffraction data were collected at 100K (after cryoprotecting the crystal by increasing the poly(acrylic acid sodium salt) 5100 concentration up to 34 %) at the XRD1 beamline of the ELETTRA synchrotron (Trieste, Italy). *Pa*RsgA crystal belongs to the space group $P4_132$, with unit cell parameters of a=b=c= 146.4 Å. One monomer of *Pa*RsgA is present per asymmetric unit. *Pa*RsgA structure (pdb code: 6H4D) was solved by molecular replacement using the structure of RsgA from *Salmonella typhimurium* (pdb code: 2RCN; Nichols et al., 2007) as a search model and refined as reported in the experimental section. The final model, containing a GDP molecule and a Zn atom, was refined to an Rwork = 0.248 and an Rfree= 0.286. Data collection and refinement parameters are reported in Table 1.

TABLE 1

Data collection and refinement statistics for RsgA

Data collection	
Beamline	XRD1
Wavelength (Å)	1.0
Space group	P4132
Cell dimensions (Å)	a=b=c=146.405
Resolution range (Å)	48.8-2.9 (3.08-2.9)[a]
CC(1/2)(%)	99.9 (75.0)
I/sigma I	20.4 (2.5)
Completeness (%)	100 (100)
Reflections	
Total no. observed	214545
Multiplicity	17.2 (18.2)
Wilson plot *B* value	75.8
Refinement statistics	
Resolution range (Å)	48.8-2.9
R_{work}/R_{free}	0.248/0.286
Average B-factor (Å^2) (no. of atoms)	
Protein	89
Waters	78.7
RMSD	
Bond lenght (A)	0.007
Bond angle (°)	1.12
Ramachandran (*n*,%)	
Favoured	98.64
Allowed	3.36
Outliers	0

[a]Values in parentheses refer to the highest resolution shell

Table 1. Data collection and refinement statistics for *Pa*RsgA.

4.2.2 Overall structure

Crystal structure of *Pa*RsgA is depicted in Figure 4.5A and, as reported in the previous section, consists of three domains: a N-terminal Oligonucleotide/ Oligosaccharide Binding (OB) domain, a central circularly permutated domain (cpGTPase), and a C-terminal Zn-binding domain.

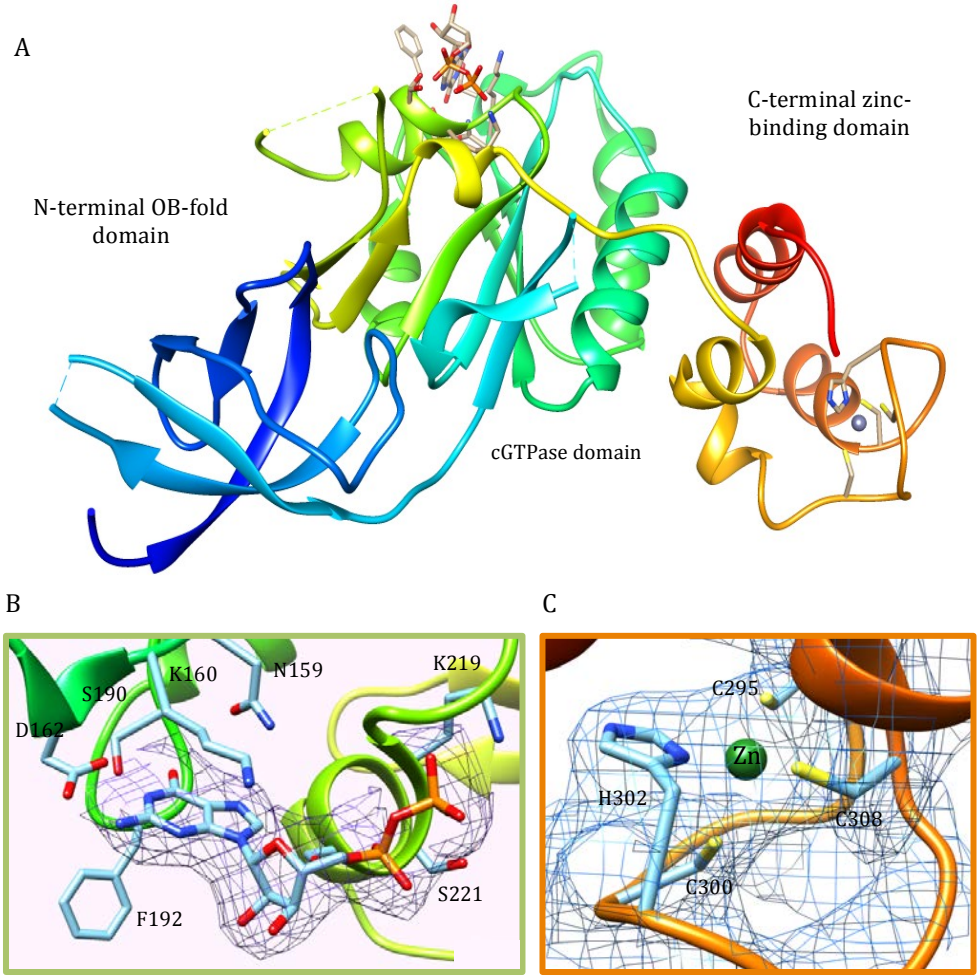

Figure 4.5 Cartoon representation of *Pa*RsgA structure (A) colored in rainbow. Close view of the GDP binding region (B). 2Fobs-Fcalc density map is contoured at 1σ and the GDP and GDP contacting residues are shown in stick representation. Close view of the zinc-binding region (C). 2Fobs-Fcalc density map is contoured at 1σ and the Zn is represented as a green sphere and Zn contacting residues are shown in stick representation.

The quality of the diffraction data is good and electron density is clearly visible for the most of residues with the exception of residues belonging to the N-terminal region (NTE, res: 1-39), residues 90-91, the switch I region

(res: 230-249), and the last four residues. Prediction of intrinsically unstructured regions, performed with the on-line server IUPRED (Dosztányi et al. 2005) on *Pa*RsgA sequence, indicates that the NTE region is natively unstructured. Albeit disordered, this region has been proposed to be involved in ribosome binding and to undergo an disorder-to-order transition upon ribosome binding, as reported in the cryo-EM structure of the *E. coli* 30S-RsgA complex (Lopez-Alonzo et al., 2017). However, a large group of RsgA orthologoues, including *Tm*RsgA (Shin et al., 2004) and *Bs*RsgA (Levdikov et al., 2004), whose 3D structures are available, do not posses this N-terminal region, suggesting a possible difference in the interaction with the ribosome and/or a not yet identified role. The schematic representation of the secondary structure elements arrangement in *Pa*RsgA structure is shown in the topology diagram (Figure 4.6).

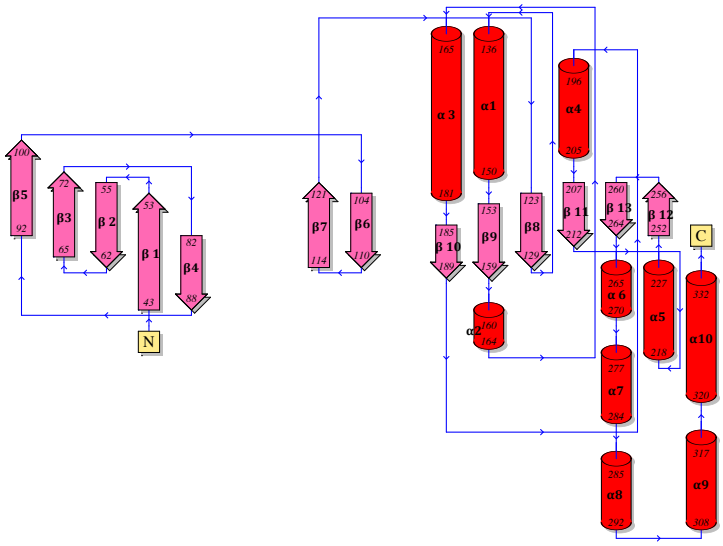

Figure 4.6. Topology diagram of GDP-*Pa*RsgA. β-strands are colored in pink and α-helices in red.

The N-terminal OB-fold domain (residues 40-101) consists of five antiparallel β-strands creating a β-barrel (β1-5β). The OB domain is a compact structural motif frequently used for nucleic acid recognition; despite the well-conserved core structure topology, sequence similarity among OB folds is low. The OB domain is connected to the GTPase domain by a two stranded antiparalleel β-sheet (β6, β7).

The cpGTPase domain spans residues 124-264. The central domain closely resembles the classical small GTPases in terms of 3D structure albeit having a circularly permutated arrangement of the G regions (G4-G5-G1-G2-G3) with respect to the classical ones. It includes the motifs: i) G1 (^{213}GQSGVGKS^{220}) located between β11 and α5; ii) G2 (T^{250}) in the loop connecting α5 and β12; iii) G3 (^{264}DSPG^{267}) following the β13, iv) G4 (^{159}NKAD^{162}) and G5 (^{190}SAK^{192}) located between β9 and α2 and β10 and α4 respectively. The G domain is formed by six stranded β-sheets and 5 α-helices in the arrangement β8, α1, β9, α2, α3, β10, α4, β11, α5, β12 and β13. The loop connecting α5 to β12 (res: 228-251), corresponding to the Switch I in the *Pa*RsgA structure, is disordered. Switch I undergoes important conformational changes upon nucleotides binding, as well as Switch II, that follows the G3 region in *Pa*RsgA. These regions both undertake an order-disorder transition upon GTP hydrolysis, thereby sending "on" or "off" signals to downstream effectors (Witthinghofer and Vetter, 2011). Residues contacting GDP (N159, K160, D162, S190, F192, K219 and S221) are strictly conserved in all protein family members (Figure 4.5B). In the structure of *P*aRsgA, the cpGTPase domain has interfaces with both the OB and the Zn-binding domains, whereas there are no intramolecular contacts between the β-barrel and Zn-binding domains. A small helix (α6) followed by a long loop, located downstream the G3 region, connect the central and the C-terminal domains. This connecting region lies down the entire protein

surface in a position suitable to propagate functional information to all the protein domains. Indeed, in cpGTPases the peculiar arrangement of the catalytic site requires an additional region/domain located downstream to properly position the G3 motif.

The C-terminal domain can be divided into two functional regions: a Zn-binding region (res: 277-319) composed by three helices (α7, α8, α9) plus a loop and a C-terminal helix α10 (res: 320-322). Three conserved cysteine residues (C295, C300, C308) and one histidine (H302) are located in a loop between helices α8 and α9 and are responsible for the tetrahedral coordination of the Zn atom (Figure 4.5C).

4.2.3 Structural comparison with *Pa*RsgA orthologues

To date, 3D structures of RsgA from different bacteria have been reported either in the GDP bound form (*Tm*RsgA: 1U0L; Shin et al., 2004, *St*RsgA: 2RCN; Nichols et al., 2007, and *Aa*RsgA: 2YV5) or in the apo form (*Bs*RsgA: 1T9H; Levdikov et al., 2004). Moreover, recent cryo-EM structure of the *E. coli* 30S-RsgA complex provides information on the GTP-bound RsgA structure (*Ec*RsgA: 5NO2; Lopez-Alonzo, 2017).

Comparison of the present GDP-*Pa*RsgA structure with the structures of orthologous proteins in the GDP-bound form, by using Coot (Emsley et al., 2010), clearly shows that the overall fold is well conserved. Pairwise comparison shows that the rmsd (calculated on the Cα atoms) is 1.3 Å with *St*RsgA, 1.7 Å with *Tm*RsgA and 2.1 Å with *Aa*RsgA (Figure 4.7). Small differences can be found in the region connecting the OB domain to the central domain; this region is composed by a two-stranded β-sheet in *Pa*RsgA and *St*RsgA while it is smaller in *A.aeolicus* and *T. maritima* RsgA structures. Moreover, in the structure of *Aa*RsgA the C-terminus is longer and the last helix is bent with the terminal part folding back (Figure 4.7).

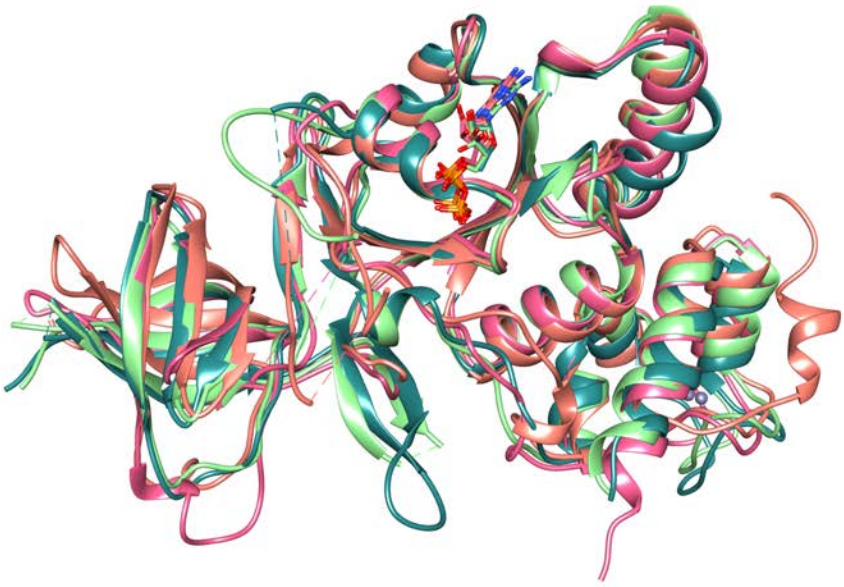

Figure 4.7. Comparison of GDP-PaRsgA structure (pdb code: 6H4D) in dark green with other RsgA structures from different bacterial species: *Salmonella Typhimurium* in light green (pdb code: 2RCN), *Thermotoga maritima* in magenta (pdb code: 1U0L) and *Aquifex aeolicus* in orange (pdb code: 2YV5).

To investigate the conformational changes in *Pa*RsgA structure upon GTP hydrolysis we struggled to obtain either the structure of the protein in complex with a non-hydrolysable-GTP analogue and the structure of the protein without nucleotides; unfortunately, we were not able to obtain crystals from the nucleotide-free form of *Pa*RsgA. We were able to crystallize *Pa*RsgA in complex with a non-hydrolysable-GTP analogue (Gpp(CH_2)p); however, to date, these crystals diffract only at very low resolution. Therefore, in a preliminary analysis, we compared the present GDP-*Pa*RsgA structure with the one from *B. subtilis* solved in the nucleotide-free form (1T9H; Figure 4.8A). The overall fold is preserved in the nucleotide-free *Bs*RsgA structure (rmsd of 2.1 Å). The OB-domain and the Zn-finger domain adopt a similar fold, while the cpGTPase domain

shows some differences due to the presence of GDP in the nucleotide-binding pocket. In particular, in *Bs*RsgA the nucleotide binding site is smaller with respect to the nucleotide-bound form of orthologous proteins whose structures have been determined (see above). This is mainly due to a structuring of the loop connecting β10 to α4 and the terminal end of the helix α5 corresponding to G1 motif (Figure 4.8B). The smaller dimension of the nucleotide binding cleft results in a steric hindrance, which is compatible with the presence of nucleotides, and the relaxation of these regions is required to provide space for nucleotides binding. Moreover, folding of the α5 helix brings the K219 of the G1 motif (corresponding to K177 in *Bs*RsgA) back on the phosphate group of the nucleotide (Figure 4.8B).

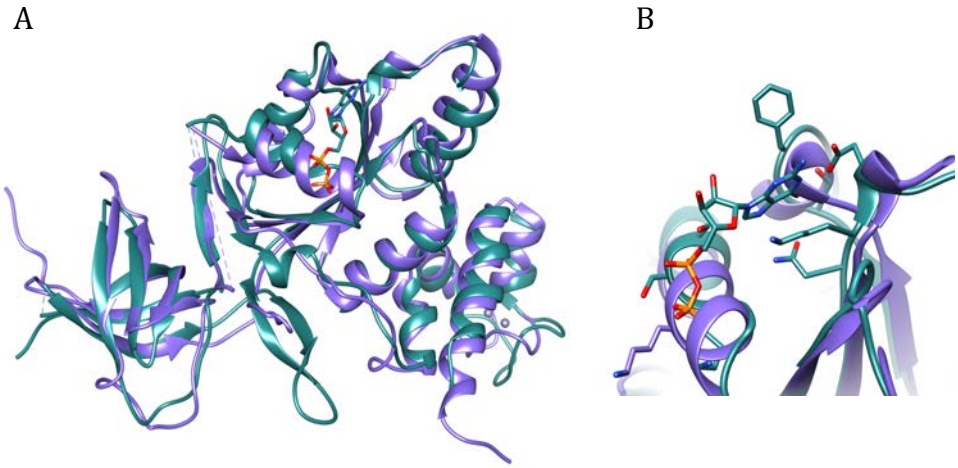

Figure 4.8. Comparison of GDP-*Pa*RsgA structure (pdb code: 6H4D) in dark green with the structure of nucleotide-free RsgA from *Bacillus subtilis* in purple (pdb code: 1T9H). (A) GDP is shown in stick coloured by heteroatoms. Close up view of the GDP binding site (B). Residues involved in nucleotide interaction are shown in stick representation.

Structural comparison performed by superimposing only the GTPase domains of the GDP-bound *Pa*RsgA, the GTP-bound *Ec*RsgA and the nucleotide-free form of *Bs*RsgA, reveals differences in the mutual domains orientation upon nucleotides binding (Figure 4.9). The absence of nucleotide in the *Bs*RsgA structure not only contributes to a local rearrangement of the guanine moiety pocket but also perturbs interdomains orientation, with a greater impact on the OB domain, which is subjected to a large movement respect to the C-terminal one. The two stranded beta sheets located between the OB and G domain could sense structural rearrangement due to the hydrolytic activity and transfer the information to the upstream domain. It is worth mentioning the possibility that the movements of the G flanking regions can be due to the interaction of *Ec*RsgA with the 30S ribosome subunit.

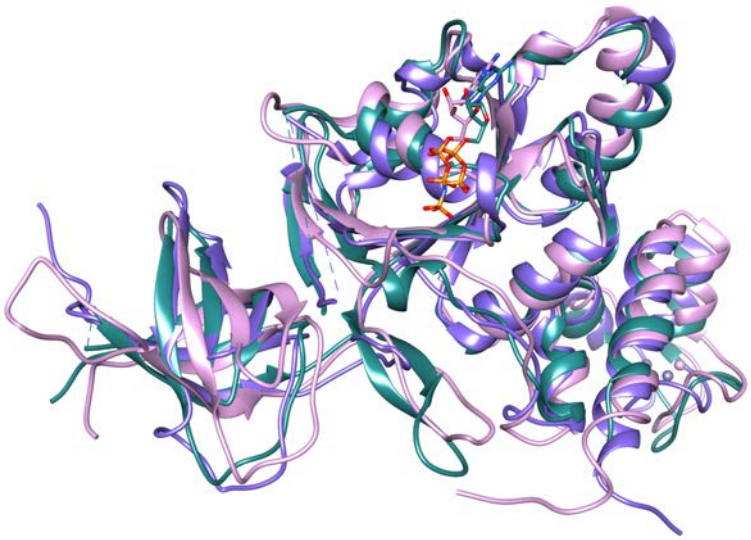

Figure 4.9 Comparison of GDP-*Pa*RsgA structure (pdb code: 6H4D) in dark green, with available structures of RsgA from *Escherichia coli* in pink (pdb code: 2NO5) and *Bacillus subtilis* in purple (pdb code: 1T9H). Nucleotides are shown in stick coloured by heteroatoms.

4.3 Preparation of the nucleotide-free form of *Pa*RsgA

To perform a detailed biochemical characterization of *Pa*RgsA we resorted to purify the nucleotide-free form of the protein and different biochemical procedures have been attempted for this purpose. Two protocols exploited the use of chaotropic agents to partially denature the protein (urea at low concentration) or to weaken electrostatic interactions (ammonium sulphate at high concentration) whereas another one is based on an enzymatic procedure (Ford et al., 2009). Each of these procedures led to the removal of the GDP bound to *Pa*RsgA, however, employing the non-enzymatic protocols we obtained very low yields of the nucleotide-free protein. Therefore, the enzymatic procedure (described in the Materials and methods section) was used to obtain a sufficient amount of the nucleotide-free form of *Pa*RsgA. The success of the procedure, firstly indicated by the presence of a sharp peak at 280 nm in the UV-spectrum (Figure 4.10), was confirmed by ion-pair RP-HPLC analysis as discussed above (see Section 1.*Pa*RsgA expression and purification and Figure 4.3).

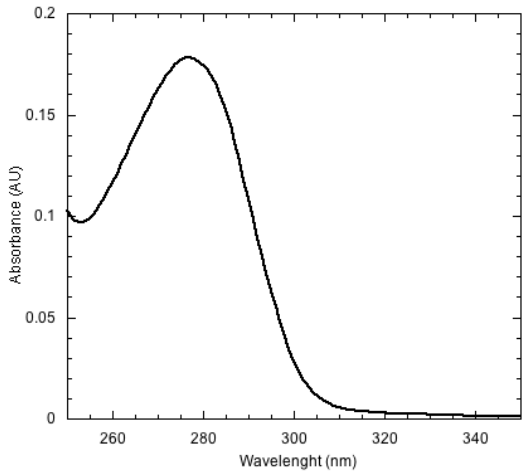

Figure 4.10 UV-spectra of nucleotide-free *Pa*RsgA

4.4. Biophysical characterization of *Pa*RsgA

4.4.1 Far-UV CD spectroscopy

In order to evaluate if the presence of a GDP molecule could alter the secondary structure content of *Pa*RsgA, far-UV CD spectroscopy experiments of both nucleotide-bound and nucleotide-free forms of the protein were carried out. As reported in Figure 11A, the CD spectra of nucleotide-bound and nucleotide-free forms of *Pa*RsgA in the far-UV region (200-250 nm) are very similar and typical of a protein with a mixed content of α-helices and β sheets. This result is in accordance with the known structures of RsgA proteins in the two forms, which do not highlight any conspicuous structural rearrangement upon binding of the nucleotide. Moreover, in order to evaluate the contribution of GDP to the overall thermic stability of *Pa*RsgA, thermal denaturation experiments, from 20 °C to 90 °C, monitoring the variation of the CD signal at 210 nm, have been carried out. Figure 11B shows that the midpoint of the thermal denaturation curve of the nucleotide-free *Pa*RsgA is slightly lower than the value obtained for nucleotide-bound protein. Although a quantitative analysis was not possible given that the thermal denaturation process was not reversible, this experiment suggests that nucleotide binding does not strongly affects the thermal stability of *Pa*RsgA.

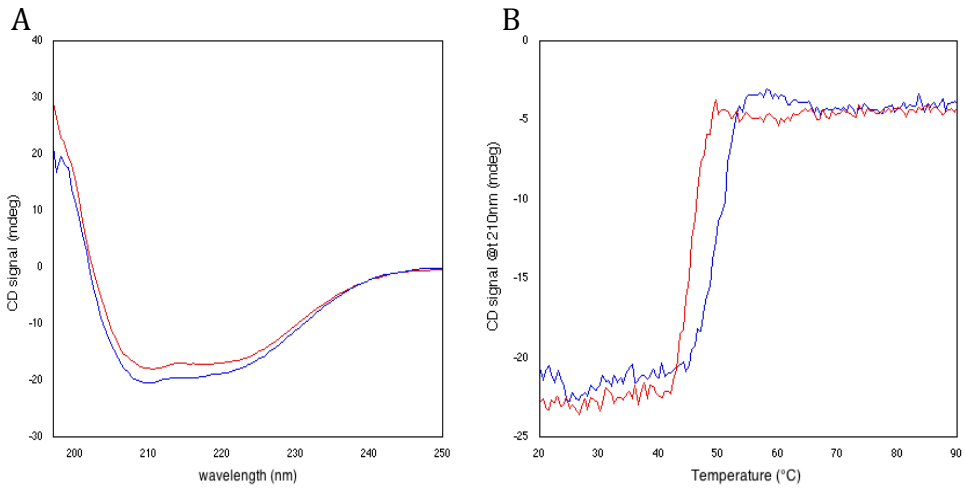

Figure 4.11 Far-UV CD spectra (A) of nucleotide-bound *Pa*RsgA (blue line) and nucleotide-free *Pa*RsgA (red line). Equilibrium thermal denaturation (B) monitored by CD spectroscopy (ellipticity at 210 nm) of nucleotide-bound *Pa*RsgA (blue line) and nucleotide-free protein (red line).

4.4.2 Urea-induced denaturation

Since thermal denaturation of *Pa*RsgA is not reversible, thermodynamic stability parameters of nucleotide-bound and nucleotide-free forms were estimated by equilibrium denaturation experiments using urea as denaturing agent at 25 °C. Fluorescence spectroscopy was used to monitor the increase of tryptophan fluorescence emission at 350 nm after addition of urea (Figure 12). In both cases, the observed unfolding transition follows a simple two-state behaviour, suggesting the absence of stable equilibrium intermediate/s. A quantitative analysis of the observed spectroscopic signals as a function of [urea] allowed us to estimate the thermodynamic stability of the proteins ($\Delta G°$) and to calculate the $m_{D\text{-}N}$ values, which describe the cooperativity of the unfolding process. The calculated parameters determined by applying a two-state model (see Materials and Methods), are $\Delta G° = 4.3 \pm 0.4$ kcal mol^{-1}, $m_{D\text{-}N} = 0.95 \pm 0.04$ kcal mol^{-1} M^{-1} for the nucleotide-bound form and $\Delta G° =$

4.2 ± 0.4 kcal mol^{-1}, $m_{D\text{-}N}$ = 0.92 ± 0.08 kcal mol^{-1} M^{-1} for the nucleotide-free *Pa*RsgA, respectively. The similar $\Delta G°$ values obtained from the denaturation experiments in urea of the two forms quantitatively confirm the observations attained from the CD thermal denaturation experiments described above. It is therefore clear that the presence of the nucleotide provides only a negligible contribution to the total thermodynamic stability of the protein.

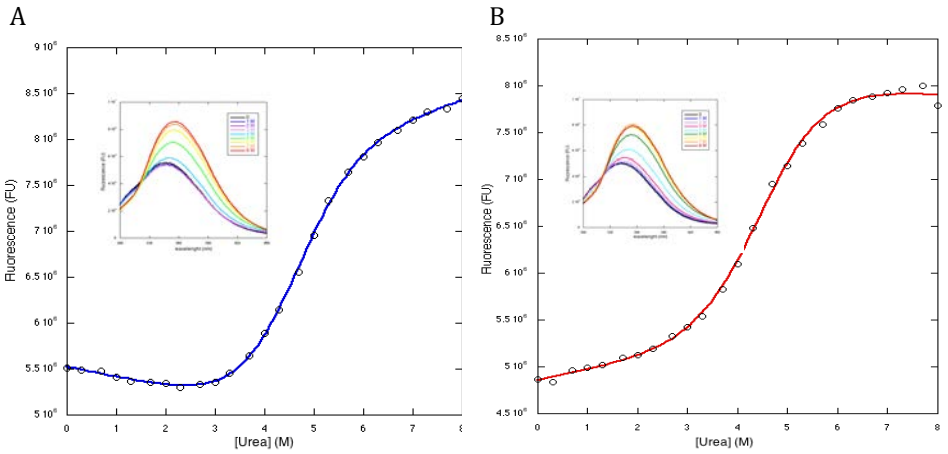

Figure 4.12. Stability of nucleotide-bound *Pa*RsgA (A) and its nucleotide-free form (B) obtained by equilibrium urea-induced denaturation monitoring intrinsic tryptophan fluorescence at 350 nm. Inset: Fluorescence emission spectra at different denaturant concentrations.

4.5 Nucleotide binding kinetics

In order to get information about the kinetic binding mechanism of *Pa*RsgA with GDP or GTP nucleotides and to obtain the relative association and dissociation rate constants (k_{on} and k_{off}), a detailed analysis was performed by time-resolved fluorescence spectroscopy. The kinetics parameters for the binding of nucleotide-free *Pa*RsgA to GDP and GTP were estimated by rapid

stopped-flow binding experiments carried out at 25°C and using fluorescent analogues of the two nucleotides (mGDP, mGTP, see Materials and Methods). The m-derivatized nucleotides provide a good tool to study protein-nucleotide interactions due to the high fluorescence quantum yield of the mant moiety and its sensitivity to the molecular environment.

Representative time courses obtained by mixing a constant amount of derivatized nucleotide against different concentrations of *Pa*RsgA solutions, are shown in Figure 4.13 A and D for both GDP and GTP binding, respectively. At this temperature and under pseudo-first order conditions, the time courses matched single exponentials at all concentrations of *Pa*RsgA. The calculated rate constants were plotted as a function of protein concentration (Figure 4.13 B and E). Values of k_{on} = 9.1 μM^{-1} s^{-1} and 8.1 μM^{-1} s^{-1} for the binding of *Pa*RsgA to GDP and GTP respectively were obtained from the slope of the linear plot of k_{obs} versus *Pa*RsgA concentration. Dissociation rate constants (k_{off}) were evaluated by stopped-flow experiments carried out by mixing *Pa*RsgA-m-nucleotide complex with an excess of the appropriate unlabelled nucleotide. These displacement experiments allowed us to obtain the dissociation rate constant k_{off} for GDP (k_{off} = 0.07 s^{-1}) and GTP (k_{off} = 0.5 s^{-1}) (Figure 4.13 C and F). Equilibrium dissociation constants, calculated from the kinetics parameters, result in a K_D of 0.008 μM for the *Pa*RsgA-GDP complex and 0.06 μM for *Pa*RsgA-GTP complex. Therefore, the higher affinity of *Pa*RsgA towards GDP mostly depends on its slower k_{off} as compared to a faster GTP release.

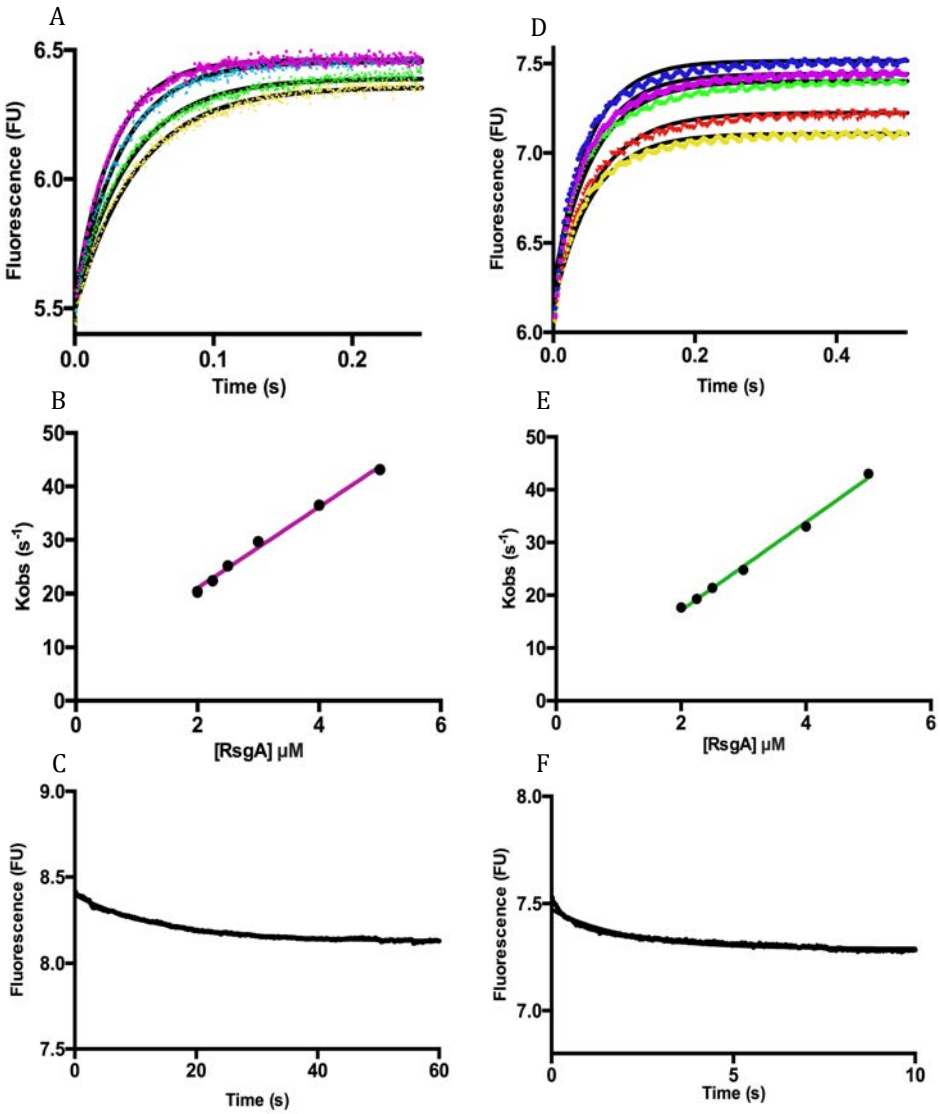

Figure 4.13 GDP and GTP binding and dissociation kinetics of *Pa*RsgA determined via stopped-flow fluorescence analysis. Transients obtained by following mGDP (A) and mGTP (D) (constant concentration of 1 µM/after mixing) fluorescence upon rapid mixing with different concentrations of *Pa*RsgA, shown in rainbow. The second panels (B and E) show the protein concentration dependency of the observed rate constant (k_{obs}). The slope of the linear fit yields k_{on}. The lower panels (C and F) show the direct determination of k_{off} by following the release of mGDP/mGTP from *Pa*RsgA, upon mixing unlabelled nucleotides with a preformed complex of *Pa*RsgA and mGDP/mGTP.

4.6 Enzyme activity

4.6.1 Intrinsic GTPase activity

Previous data showed that RsgA from *E. coli* exhibits weak intrinsic GTPase activity (Daigle et al., 2002) while association of RsgA with the 30S subunit leads to a 160-fold stimulation of its activity (Daigle et al., 2004; Himeno et al., 2004). Moreover, it is well known that the enzyme activity of several RA-GTPases is stimulated in presence of NaCl and/or KCl (Anand et al., 2010; Rafay et al., 2012; Achila et al., 2012; Foucher et al., 2012; Gkekas et al., 2017). In order to get insight into the catalytic activity of *Pa*RsgA, we resorted to investigate its intrinsic GTPase activity measuring the GTP-GDP conversion over time by ion-paring RP-HPLC analysis (see 3.1 Section).

To define the optimal experimental conditions necessary for evaluating the catalytic parameters of *Pa*RsgA GTPase activity, pilot experiments were performed testing different ionic strength conditions. Purified *Pa*RsgA was incubated with 400 μM GTP (steady state conditions) at 25°C in a buffer containing different concentrations (up to 300 mM) of NaCl or KCl, as well as in absence of salt, and the GDP production was measured after 120 minutes incubation. Nucleotides were detected by their absorbance at 254 nm and the corresponding peaks areas were converted to concentrations using calibration curves. In Figure 4.14 is shown the standard curve used to estimate the GDP production in term of concentration for the following analysis.

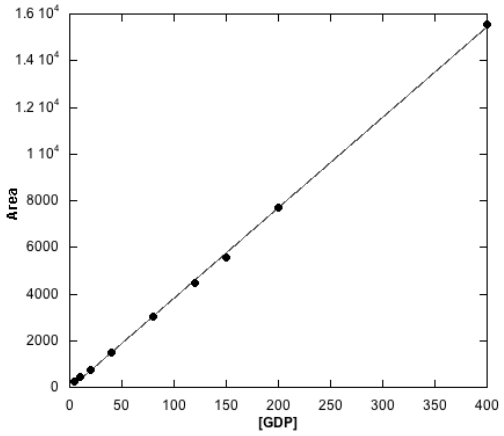

Figure 4.14 Calibration curve of GDP used for quantitative analysis. The plot of analytical signal (the GDP area calculated by ion-paring RP-HPLC analysis) versus GDP concentrations show a linear relationship.

The GDP production rate calculated for the different ionic strength conditions were compared, in term of fold rate, to the condition in absence of salt (Figure 4.15).

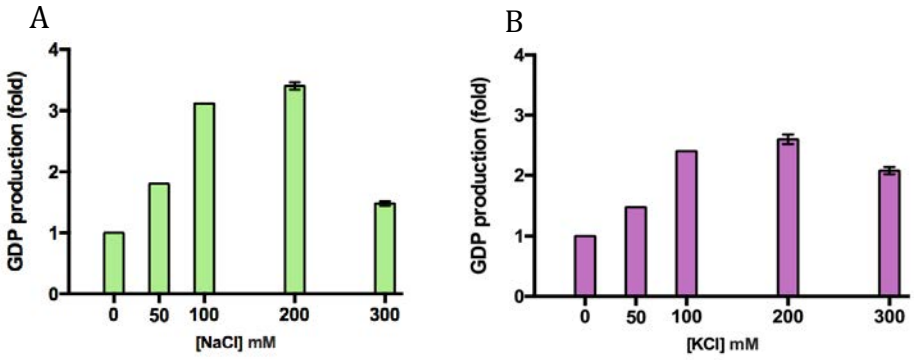

Figure 4.15 Influence of the ionic strength over *Pa*RsgA GTPase activity. The dependency of the GDP production rate over increasing concentrations of NaCl (A) and KCl (B), shown in term of fold stimulation as compared to the reaction carried out in the absence of any salt.

These data shown that the GTPase activity of *Pa*RsgA is dependent to the ionic strength with no selectivity for Na^+ or K^+ ions and that the GDP

production attains a maximum between 100 and 200 mM salt concentration. The maximal stimulation is reached around 200 mM NaCl and KCl, resulting in 3.4 to 2.6 folds rate enhancement, respectively. Therefore, the ionic strength condition of 200 mM NaCl, at whom the maximal stimulation is obtained and afterward referred as internal standard, was used for the following quantitative analysis.

For the quantitative determination of catalytic parameters, purified *Pa*RsgA (5 μM) was incubated with different GTP concentrations (50-1200 μM) at 25°C in presence of 200 mM NaCl and the reactions were stopped at different time intervals. An example of the time course is represented in Figure 4.16, in which the increase of GDP area (corresponding to the GDP production) and the decrease of GTP area over the time are highlighted.

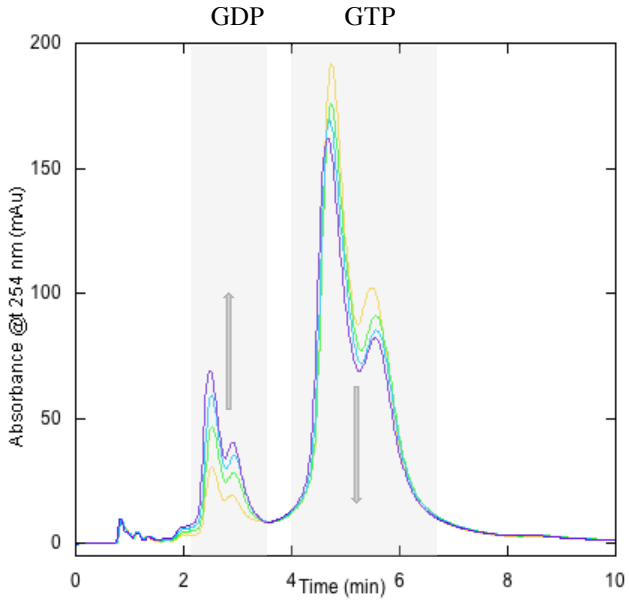

Figure 4.16 Example of the time course in multi-turnover analysis. *Pa*Rsga (5μM) was incubated with 400 μM GTP at 25°C, in presence of 10 mM $MgCl_2$. At different time points (0, 60, 120, 180 min) the reaction was stopped and the GTP hydrolysis was monitored.

Since the reaction is so slow to have a linear trend in the analysed time window, the initial reaction rate (v_o) was obtained as the slope of the linear fit of the initial linear period of the reaction in plot showing the GDP concentration versus time. An example of the initial reaction rate (v_o) evaluation is shown in Figure 4.17.

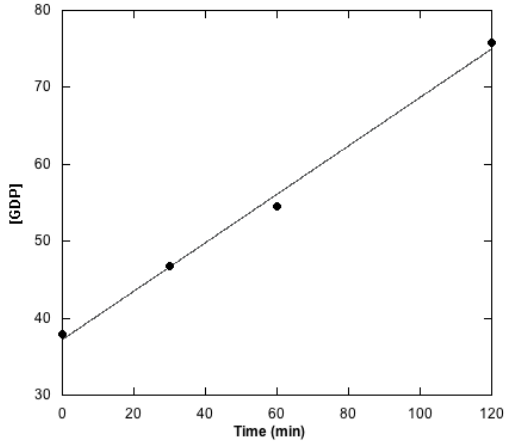

Figure 4.18 *Pa*RsgA GTPase activity. Initial reaction rate (vo), obtained as the slope of GDP concentration versus time plots, are plotted as a function of GTP concentration and fitted to the Michaelis-Menten equation.

The observed velocities at certain concentration were plotted as a function of GTP concentration and fitted with the Michaelis-Menten equation (Figure 4.18). *Pa*RsgA has slow but saturable GTPase activity, as expected for this class of proteins, with K_M = 208 ± 27,7 µM, v_{max} = 0,29 ± 0,011 µM min^{-1} and K_{cat} = 0.058 min^{-1}.

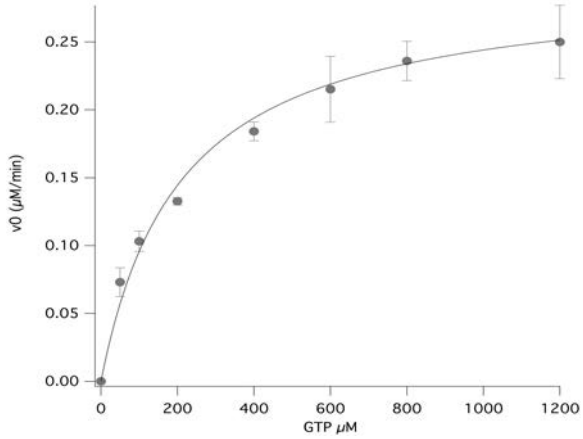

Figure 4.17 Example of initial reaction rate (v_o) calculation. The initial reaction rate (v_o) was obtained as the slope of the linear fit of the initial linear period of the reaction in plot showing the GDP concentration (evaluated at different time intervals by using the standard curve) versus time.

4.6.2 Role of monovalent cations in the GTPase activity of *Pa*RsgA

The presence of a monovalent cation (usually Na^+ or K^+) not only accelerates the intrinsic activity of diverse RA-GTPases but also stabilizes the transition state and thereby is involved in GTP hydrolysis mechanism (Anand et al., 2010; Rafay et al., 2012; Achila et al., 2012; Foucher et al., 2012; Gkekas et al., 2017). This mechanism is in contrast to the well known mechanism employed by classical GTPases like Ras, which utilize an 'Arginine finger' from an interacting GTPase activating protein (GAP) to stabilize the transition state and facilitate GTP hydrolysis. Currently, experimental data about the ability of monovalent cations to influence the GTPase activity of RsgA are not available in the literature.

To test if the nature of the cation is important in modulating the GTPase activity of *Pa*RsgA, its intrinsic activity was investigated in the presence of 200 mM of different monovalent salts such as NaCl, KCl, NH_4Cl, LiCl and

CsCl. The rate of GDP production has been compared, in term of percentage, to the one measured in 200 mM NaCl, used as our internal reference. As reported in Figure 4.19, in the presence of KCl the GDP production is slight lower (around 76%), while in the presence of LiCl it decreases further to approximately 24% and is less than 10% in the presence of NH_4Cl and CsCl.

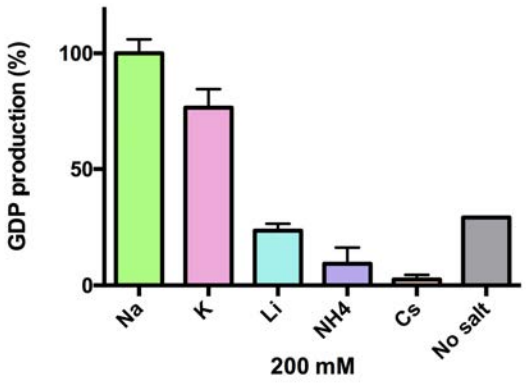

Figure 4.19 Influence of the nature of the cation on *Pa*RsgA GTPase activity. The GDP production rate obtained in presence of different monovalent cations (Na^+, K^+, Li^+, NH_4^+, Cs^+) and any salt (no salt) compared, in terms of percentage, to the condition in which there is the maximal stimulation, namely 200mM NaCl (100% of activity).

These preliminary experiments suggest that *Pa*RsgA catalytic site is suitable for accommodating monovalent cation with an ionic radius in the range of 102 to 138 (ionic radii of Na^+: 102 pm, K^+: 138 pm). Cations smaller or larger in size (Li^+: 76 pm, NH_4^+: 147 pm and Cs^+: 167 pm) have only a slight stimulating effect on *Pa*RsgA GTPase activity.

2.6.3 Possible determinant features for GTP hydrolysis mechanism of *Pa*RsgA

*Pa*RsgA GTPase activity is stimulated by Na^+ and K^+ ions. Those cations

might function as the so far elusive catalytic element in *Pa*RsgA that contributes to rapid GTP hydrolysis by providing electrostatic stabilization for the TS. Therefore, *Pa*RsgA could use the M^+-mediated mechanism, as depicted for other GTPases, like MnmE (Scrima et al., 2006), YqeH (Anand et al., 2010), and RbgA (Ash et al., 2010).

The determinants of M^+-mediated mechanism have been predicted to be two conserved Asn residues, one in the P-loop (G1) and another following it (GxxNxGKSxxxN); and the presence of an insertion termed K-loop, in Switch-I (G2) (Ash et al., 2012). Similarly, in dynamin a conserved Ser present in the P-loop, in place of Asn, stabilizes the M^+ stimulating GTP hydrolysis (Chappie et al., 2010). However, the mere conservation of a P-loop Asn/Ser and a K-loop are not reliable indicators of M^+ stimulated GTP hydrolysis (Rafay, 2012). In the same field, Khule and coworkers have proposed that the key determinants for M^+ ion binding are an Asp in the P-loop and a Gly in the switch I (GxT in G2 motif) (Khule, 2012).

To verify the presence of the structural element predict to trigger the M^+-mediated mechanism, we performed a multiple sequence alignment (MSA), by using the program Clustalω (Sievers et al., 2011), between *Pa*RsgA and different RA-GTPases (*Bs*RbgA, *Bs*YqeH, *Ec*Obg, *Ec*MnmE, *Ec*Era, *Ec*Der, *Ec*EngB). Sequence analysis (Figure 4.20) reveals that Ser215 in G1, like in dynamin, is located one residue downstream in sequence (respect to the first conserved Asn in others RA-GTPases), followed by a Gly, whereas the Asn224 is conserved in *Pa*RsgA sequence. Gly247 in G2, also indicated as important for TS stabilization, is located one residue downstream in the sequence (^{247}**G**THT^{250} respect to the canonical GxT).

Therefore, the putative determinants features of the M^+-mediated mechanism are present in *Pa*RsgA sequence. However, here we can only speculate about the residues involved in GTP hydrolysis mechanism. To gain key insights

into *Pa*RsgA molecular mechanism and to define the key determinants additional structural information are required.

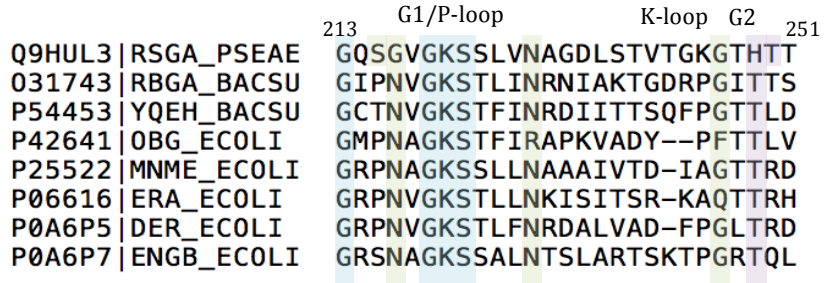

Figure 4.20 Multiple sequence alignment of *Pa*RsgA and different RA-GTPases. Primary sequences of respective GTPases were retrieved in Uniprot and were subjected to multiple sequence alignment (MSA) by using the program clustalω. The region spanning G1 and G2 (residues 213-251 in *Pa*RsgA sequence) is shown. Residues suggested to coordinate the M^+ ion and to stabilize the TS are highlighted in green. Residues conserved in G1 are highlighted in cyan and a conserved Thr from G2 is highlighted in purple.

5. Discussion

Ribosome assembly in vivo is an extremely efficient process owing to the existence of a variety of assembly factors (Woodson et al., 2011). During ribosomal biogenesis, assembly factors prevent accumulation of misfolded intermediate states by facilitating proper rRNA folding and r-protein-rRNA interactions (Shajani et al., 2011) and/or by acting as checkpointers in the context of the mature ribosomal subunits (Woodson et al., 2008). Understanding their specific functions in subunits maturation may help not only to gain a clear picture of how they work, but also to identify key molecular features to target for the development of new antibiotics.

Four protein factors, RsgA, Era, RbfA and RimM, are involved in the late stages of the 30S subunit maturation (Jomaa et al., 2011; Leong et al., 2013; Guo et al., 2013; Jeganathan et al., 2015; Thurolow et al., 2016). These factors bind the small subunit at or near the decoding center and contribute to the 30S proper biogenesis (Lopez-Alonso et al., 2017; Razi et al., 2017; Sharma et al., 2015; Datta et al., 2007); however, the precise mechanisms of action as well as the functional interplay among them are not yet completely understood.

This study focuses on the ribosome small subunit dependent GTPase A from the pathogen *Pseudomonas aeruginosa* (*Pa*RsgA). RsgA is broadly conserved in bacteria and has been shown to be dispensable for growth but critically important for the overall fitness of *E. coli, B. subtilis, and S. aureus.* Moreover, reduced virulence of an *S. aureus rsgA* deletion strain in mouse model has been reported (Tracey et al., 2008). Preliminary experiments performed in collaboration with Dr. Francesco Imperi ("Sapienza" Università di Roma) showed that deletion of the *rsgA* gene in *P. aeruginosa* causes a slight delay of bacterial growth but a significant reduction in pathogenicity in

animal model (*Galleria mellonella*), indicating that although RsgA is not an essential protein in *P. aeruginosa*, it can play a key role during the infectious process. Therefore, *Pa*RsgA represents a possible novel target for antibacterial drug discovery for *P. aeruginosa* infections treatment.

In an attempt to perform a detailed characterization of *Pa*RsgA from a structural and biochemical point of view, *Pa*RsgA was expressed as a recombinant protein in *E. coli* and purified to homogeneity. The protein was submitted to extensive crystallization trials and then crystallised as described in the Results section. The structure was solved with a GDP molecule in the active site at 2.9 Å resolution and the model was deposited in the Protein Data Bank (pdb code: 6H4D). *Pa*RsgA consists of three domains following a N-terminal unstructured region: an OB domain, a central cpGTPase domain and a C-terminal Zn-binding domain (Figure 4.5). The OB domain consists of two three stranded antiparallel β-sheets, packed orthogonally to one another creating a β-barrel; a two-stranded β-sheet connects the OB domain to the central circularly permutated domain, that is made up by a central six-stranded β-sheet surrounded by α helices. The C-terminal region encompasses the zinc-binding domain and a C-terminal extension. Different structures of RsgA orthologous have been solved by X-ray crystallography, either in the GDP bound form (*Tm*RsgA: 1U0L; Shin et al., 2004, *St*RsgA: 2RCN; Nichols et al., 2007, and *Aa*RsgA: 2YV5) and in the apo form (*Bs*RsgA: 1T9H; Levdikov et al., 2004). Recently, cryo-EM structure of the *E. coli* 30S-RsgA complex has provided information on the GTP-bound RsgA structure (*Ec*RsgA: 5NO2; Lopez-Alonzo et al., 2017) as well as on its interaction with the 30S subunit (Razi et al., 2017). Comparison of the present *Pa*RsgA structure with the structures of orthologous proteins in the GDP-bound form indicates that the overall fold is well preserved (Figure 4.7) with most of the differences localizing in the region connecting the OB

domain to the cGTPase domain and the C-terminal region (see Results). The overall conservation among different bacterial species implies a strong link between structure and functional properties. Interestingly, a comparison of the structures of the GDP-bound *Pa*RsgA with the GTP-bound *Ec*RsgA and the nucleotide-free *Bs*RsgA, reveals some differences in the respective domain orientation (Figure 4.9). In particular, the OB domain is subjected to a larger movement with respect to the zinc-binding domains. The absence of nucleotide in the *Bs*RsgA structure not only contributes to a different inter-domains orientation, but also leads to a local rearrangement of the nucleotide-binding pocket. In particular, in *Bs*RsgA the nucleotide-binding cavity is smaller than it is in the nucleotide-bound form of orthologous proteins. Thus a relaxation of the cavity is clearly required to provide space for nucleotide binding. Unfortunately, structures of RsgA available to date (including our *Pa*RsgA structure) do not provide an accurate description of important functional motifs of RsgA. These motifs include the N-terminal unstructured region, which is important for binding to the 30S subunit in *E. coli* (Razi et al., 2017) and the Switch I. Since Switch I is the major determinant of structural rearrangement upon GTP hydrolysis, to investigate these conformational changes we struggled to obtain either the structure of the protein in complex with a non-hydrolysable-GTP analogue ($GppCH_2p$), and the structure of the protein in the nucleotide-free form. However, to date *Pa*RsgA-$ppCH_2p$ crystals diffracted only at low resolution and optimization trials are in progress.

In order to get insight into the function of *Pa*RsgA we have carried out an extensive biochemical and biophysical characterization of this protein. To our knowledge no thermodynamic parameters of RsgA, in both the nucleotide-free and nucleotide-bound forms are currently available.

Since *Pa*RsgA is purified in the nucleotide-bound form, different biochemical

procedures have been optimized to obtain the nucleotide-free form of the protein: an enzymatic procedure and two different protocols that exploited the use of chaotropic agents to partially denature the protein or to weaken electrostatic interactions (Ford et al., 2009). Each of these procedures successfully led to the removal of the GDP bound to *Pa*RsgA. However, the enzymatic procedure, showing the best yields, was used to obtain a sufficient amount of the nucleotide-free form of *Pa*RsgA.

In order to evaluate the contribute of a GDP molecule on the secondary structure content of *Pa*RsgA, far-UV CD spectroscopy experiments of both nucleotide-bound and nucleotide-free forms of the protein were performed. The far-UV spectra do not highlight any conspicuous structural rearrangement upon binding of the nucleotide and therefore the nucleotide bound does not perturb the overall secondary structure of *Pa*RsgA (Figure 4.11). The thermodynamic stability parameters of nucleotide-bound and nucleotide-free forms were estimated by equilibrium denaturation experiments using urea as denaturing agent. The similar $\Delta G°$ values obtained for both *Pa*RsgA forms, quantitatively confirm the observations obtained from the CD thermal denaturation experiments (Figure 4.12). It is therefore clear that the presence of the nucleotide provides only a negligible contribution to the total thermodynamic stability of the protein.

The possibility to obtain large amounts of the nucleotide-free form of the protein, allowed us to perform a detailed kinetic analysis of *Pa*RsgA binding to nucleotides. The kinetics parameters at 25°C for the binding of nucleotide-free RsgA to GDP and GTP were estimated by rapid kinetic binding experiments using a stopped-flow apparatus and fluorescent analogues of the two nucleotides (mGDP, mGTP). Equilibrium dissociation constants (K_D), calculated from the kinetics parameters, are 0.008 μM for the *Pa*RsgA-GDP interaction and 0.06 μM for *Pa*RsgA-GTP interaction. The higher affinity of

*Pa*RsgA binding to GDP, respect to GTP, is mostly dependant on its slower k_{off} as compared to a faster GTP release (Figure 4.13). Nichols and coworkers assessed nucleotide binding of *St*RsgA by isothermal titration calorimetry, resulting in a K_D = 1 μM for GDP and K_D = 8 μM for GTP, values slight higher compared to those here obtained for *Pa*RsgA (Nichols et al., 2007). These differences could be ascribed either to the diverse analytical methodologies and experimental conditions used to evaluate the kinetics parameters or to the use of nucleotide-bound StRsgA in the assay.

RsgA characterized early on exhibit slow intrinsic GTPase activity (Daigle, et al., 2002) enhanced by 160-fold in the presence of mature 30S subunits (Daigle et al., 2004; Himeno et al., 2004). Information available to date do not provide the molecular determinants of this effect even though, very recently, it has been proposed that a specific conformation of helix 44 may be responsible of GTPase stimulation (Razi et al., 2017). Once GTP hydrolysis has occurred, additional conformational changes in switch I and switch II trigger the reorganization of their interactions with the 30S subunit, causing an overall decrease in the binding affinity of RsgA (Daigle et al., 2004; Thurlow et al., 2016) and its release from the 30S subunit. Therefore, the GTPase activity of RsgA functions as a sensor to facilitate the release of the protein factor from the 30S subunit once RsgA has performed its functions (Razi et al., 2017). Apart from the GTP-induced conformational switch, the mechanism of GTP hydrolysis is an unsolved problem in the universal functional cycle of RA-GTPases.

Central to biological processes is the regulation rendered by GTPases. Despite being a critical enzyme family, G-proteins are typically very inefficient catalysts and can have intrinsic turnover numbers as low as 10^{-5} s^{-1} (Ash, 2012). G-proteins must associate with additional factors in vivo that accelerate the conversion of GTP to GDP. Until recently, the GTP

hydrolysis mechanism, exemplified by Ras-family of GTPases, was thought to be universal. The emergence of a novel class of GTPases, like the HAS-GTPases, that carry a hydrophobic residue *in lieu* of the conserved catalytic Gln (Mishra, 2005), suggested that newer mechanisms to hydrolyze GTP must exist. Several RA-GTPases use a monovalent cation (M^+ ion) as a structural and catalytic cofactor that contributes to rapid GTP hydrolysis by providing electrostatic stabilization for the transition state (Kuhle et al., 2014). Moreover, these GTPases does not require a GAP and instead a M^+ (K^+ or/and Na^+) substitutes for the role of an Arginine finger in stabilizing the TS (Scrima et al., 2006; Anand et al., 2010; Ash et al., 2010; Chappie et al., 2010).

To get insight into the molecular mechanism underlying the *Pa*RsgA GTPase activity, the influence of the ionic strength over its hydrolytic activity was investigated and the catalytic parameters were estimated. The GDP production rate reaches maximum values between 100 and 200 mM M^+ ions (KCl and NaCl), according to the relative concentration of these ions in the cell. In presence of 200 mM KCl and NaCl there is a 2.6 to 3.4 folds rate enhancement in the GDP production, respectively. Therefore, there is a dependency of the *Pa*RsgA GTPase activity over the ionic strength but no selectivity for Na^+ or K^+ ions (Figure 4.15), and the maximal stimulation is attained at 200mM NaCl. *Pa*RsgA has slow but saturable GTPase activity, as expected for this class of proteins, with K_M of 208 ± 27,7 μM, v_{max} of 0,29 ± 0,011 μM min^{-1} and K_{cat} of 0.058 min^{-1} (Figure 4.18). The intrinsic *Pa*RsgA activity is then stimulated by the presence of a monovalent cation, where the nature of cation has in important role. The rate of GDP production over the time is stimulated mostly by the presence of Na^+ and K^+, whereas other monovalent cations like Li^+, NH_4^+ and Cs^+ have only a slight effect on the GDP production stimulation (Figure 4.19). This observation suggests that it

is plausible that the nucleotide binding site is suitable for a cation with an ionic radius in the range of 102 pm to 138 pm (respectively, ionic radii of Na^+ and K^+). Smaller (Li, NH_4^+) or larger (Cs^+) cations do not have the same stimulating effect described for Na^+ and K^+. To gain key insights regarding the possibility to use both Na^+ and K^+ to stabilise the TS, structural information on *Pa*RsgA bound to TS analogues are needed. Based on the conservation of the features in the G motifs responsible for the stabilization of TS and for triggering GTP hydrolysis, we checked for the presence of those elements in *Pa*RsgA sequence. The proposed determinants features of M^+-mediated mechanism, the Ser/Asp and Asp (^{215}**S**GVGKSSLV**N**^{224}) in G1 as well as Gly (^{247}**G**THT^{250}) in G2, are conserved in *Pa*RsgA sequence. However, here we can only speculate about the key residues possibly involved in GTP hydrolysis mechanism. To understand the molecular mechanism of GTP hydrolysis of *Pa*RsgA additional structural and biochemical studies are needed; this information ill be pivotal to explore the possibility to target *Pa*RsgA for *P. aeruginosa* infection treatment.

6. Conclusion and future prospects

Here we presented the first structural and functional characterization of Ribosome small subunit dependent GTPase A from *Pseudomonas aeruginosa* (*Pa*RsgA). The crystal structure of *Pa*RsgA has been solved with a GDP molecule in the active site (pdb code: 6H4D). A detailed biochemical and biophysical characterization of both the nucleotide-bound and nucleotide-free form of *Pa*RsgA has been accomplished; the substrates affinities have been estimated as well as the catalytic parameters of its GTPase activity.

Despite it is not yet possible to define the underlying principles of *Pa*RsgA molecular mechanism due to the lack of the completeness of data, the information here reported pose the basis for a wider characterization of *Pa*RsgA. Further structural and biochemical studies are essential to completely decipher its mechanism of action but also to define possible sites to target for the development of novel antimicrobial agents.

To gain key insights into the molecular mechanism of *Pa*RsgA, structural data on GTP-bound form are required. *Pa*RsgA in complex with a non-hydrolysable GTP analogue has been crystallized and improvements of the crystallization conditions are in progress. Structural information on the GTP-bound form of *Pa*RsgA would shed the light on the key determinants responsible for the stabilization of the transition state and for triggering GTP hydrolysis. Moreover, the structure of the GTP-bound form of *Pa*RsgA would provide additional information on the inter-domain interfaces, possible target sites to interfere with the *Pa*RsgA functions. Therefore, a complete structural characterization would be pivotal to explore the possibility to target *Pa*RsgA for *Pseudomonas aeruginosa* infections treatment.

With the increased importance of ribosome biogenesis as a potential anti-microbial target, the chemical basis of RsgA activity and its functional interplay with other assembly factor becomes more important. Therefore, we propose to investigate the functional relationship between *Pa*RsgA and other ribosomal assembly factors involved in 30S subunit maturation. Among them, Ribosome-binding factor A from *P. aeruginosa* (*Pa*RbfA) is the first candidate; this protein has been successfully cloned and purified in our laboratory.

*Pa*RsgA ability to bind *Pa*RbfA will be investigated by surface plasmon resonance (SPR) technique as well as the influence of *Pa*RbfA on *Pa*RsgA GTPase activity. These additional data would help to clarify the molecular mechanisms of how this protein assists the maturation of the functional core of the 30S subunit.

During my PhD I also participated in a project aimed at the structural and functional characterization of Nucleophosmin (NPM), a multifunctional nucleo-cytoplasmic shuttling protein involved in ribosome maturation in eukaryotes, whose role in Acute Myeolid Leukemia and its potential role as a drug target are well documented (Di Matteo et al., 2016).
In particular, I was involved in the characterization of the interaction between NPM N-terminal domain and the tumor suppressor Fbw7γ, that permitted to identify the protein surfaces implicated in recognition and the key aminoacids involved. That analysis was also extended to other NPM partners (HIV-Tat and CENP-W) allowing conclude that NPM uses the same molecular surface as a platform to recognize different protein partners. This study (Di Matteo et al., 2017) contributed to unveil the molecular details of the interaction between NPM and its partners.

In preparation:
"Structural and functional investigation of the Ribosome Small Subunit-dependent GTPase A (RsgA) from *Pseudomonas aeruginosa."*
Rocchio S., Santorelli D., Rinaldo S., Malatesta F., Imperi F., Federici L., Travaglini-Allocatelli, Di Matteo A.

"Folding and aggregration properties of RbfA from *Pseudomonas aeruginosa*." Santorelli D., Rocchio S., Angelucci F., Imperi F., Marasco D., Federici L., Di Matteo A, Travaglini-Allocatelli.

References

Achila, D., Gulati, M., Jain, N. and Britton, R.A. (2012) Biochemical Characterization of Ribosome Assembly GTPase RbgA in Bacillus subtilis. J. Biol. Chem. 287, 8417–8423.

Adams,P.D., Afonine,P.V., Bunko´ czi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, 66, 213–221.

Adilakshmi,T., Bellur,D.L. and Woodson,S.A. (2008) Concurrent nucleation of 16S folding and induced t in 30S ribosome assembly. *Nature*, 455, 1268–1272.

Anand B, Verma SK, Prakash B. 2006. Structural stabilization of GTP-binding domains in circularly permuted GTPases: implications for RNA binding. *Nucleic Acids Res.* 34:2196–205

Anand, B., Surana, P. and Prakash, B. (2010) Deciphering the catalytic machinery in 30S ribosome assembly GTPase YqeH. PLoS ONE 5, e9944.

Ash MR, Maher MJ, Mitchell Guss J, Jormakka M (2012) The cation-dependent G-proteins: in a class of their own. FEBS Lett 586: 2218 – 2224

Ash, M.R., Guilfoyle, A., Clarke, R.J., Guss, J.M., Maher, M.J. and Jormakka, M. (2010) Potassium-activated GTPase reaction in the G Protein-coupled ferrous iron transporter B. J. Biol. Chem. 285, 14594–14602.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A° resolution. *Science* 289:905–20

Blanc, D. S., Petignat, C., Janin, B., Bille, J. & Francioli, P. (1998). Frequency and molecular diversity of Pseudomonas aeruginosa upon admission and during hospitalization: a prospective epidemiologic study. Clin Microbiol Infect 4, 242–247.

Bourne,H.R., Sanders,D.A. and McCormick,F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature*, 349, 117–127.

Britton RA (2009) Role of GTPases in bacterial ribosome assembly. Annu Rev Microbiol 63:155–176.

Britton RA, Powell BS, Dasgupta S, Sun Q, Margolin W, et al. 1998. Cell cycle arrest in Era GTPase mutants: a potential growth rate-regulated checkpoint in *Escherichia coli*. *Mol. Microbiol.* 27:739–50

Campbell TL, Brown ED. 2008. Genetic interaction screens with ordered overexpression and deletion clone sets implicate the *Escherichia coli* GTPase YjeQ in late ribosome biogenesis. *J. Bacteriol.* 190:2537–45

Campbell TL, Daigle DM, Brown ED. 2005. Characterization of the *Bacillus subtilis* GTPase YloQ and its role in ribosome function. *Biochem. J.* 389:843–52

Campbell,T.L., Henderson,J., Heinrichs,D.E. and Brown,E.D. (2006) The yjeQ gene is required for virulence of Staphylococcus aureus. *Infect. Immun.*, 74, 4918–4921.

Carter AP, et al. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature 407:340–348. Carter AP, et al. (2001) Crystal structure of an initiation factor bound to the 30S ribosomal subunit. Science 291:498–501.

Chappie JS, Acharya S, Leonard M, Schmid SL, Dyda F (2010) G domain dimerization controls dynamin's assembly-stimulated GTPase activity. Nature 465: 435 – 440

Cladie`re L, Hamze K, Madec E, Levdikov VM, Wilkinson AJ, Holland IB, Se´ror SJ (2006) The GTPase, CpgA(YloQ), a putative translation factor, is implicated in morphogenesis in Bacillus subtilis. Mol Genet Genomics 275: 409–420

Comartin,D. and Brown,E. (2006) Non-ribosomal factors in ribosome subunit assembly are emerging targets for new antibacterial drugs. *Curr. Opin. Pharmacol.*, 6, 453–458.

Connolly K, Culver G (2009) Deconstructing ribosome construction. Trends Biochem Sci 34:256–263.

Connolly K, Culver G (2009) Deconstructing ribosome construction. Trends Biochem Sci 34:256–263.

Connolly K, Rife JP, Culver G (2008) Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. Mol Microbiol 70:1062–1075.

Corrigan RM, Bellows LE, Wood A, Gründling A. 2016. ppGpp negatively impacts ribosome assembly affecting growth and antimicrobial tol- erance in Gram-positive bacteria. Proc Natl Acad Sci U S A 113: E1710–E1719.

Culver GM. 2003. Assembly of the 30S ribosomal subunit. *Biopolymers* 68:234–49

Daigle DM, Brown ED (2004) Studies of the interaction of Escherichia coli YjeQ with the ribosome in vitro. J Bacteriol 186:1381–1387.

Daigle DM, et al. (2002) YjeQ, an essential, conserved, uncharacterized protein from Escherichia coli, is an unusual GTPase with circularly permuted G-motifs and marked burst kinetics. Biochemistry 41:11109–11117.

Dallas A, Noller HF. (2001). Interaction of translation initiation factor 3 with the 30S ribosomal subunit. Mol Cell. 8(4):855-64.

Datta PP, et al. (2007) Structural aspects of RbfA action during small ribosomal sub- unit assembly. Mol Cell 28:434–445.

Davies, J.; Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 2010, *74*, 417–433.

Dosztányi Z, Csizmok V, Tompa P, Simon I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 21(16):3433-4.

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501.

Foucher AE, Reiser JB, Ebel C, Housset D, Jault JM (2012) Potassium acts as a GTPase-activating element on each nucleotide-binding domain of the essential Bacillus subtilis EngA. PLoS ONE 7: e46795

Foucher AE, Reiser JB, Ebel C, Housset D, Jault JM. (2012) Potassium acts as a GTPase-activating element on each nucleotide-binding domain of the essential Bacillus subtilis EngA. PLoS One. 7(10):e46795.

Gkekas S, Singh RK, Shkumatov AV, Messens J, Fauvart M, Verstraeten N, Michiels J, Versées W. (2017). Structural and biochemical analysis of *Escherichia coli* ObgE, a central regulator of bacterial persistence. J Biol Chem. 292(14):5871-5883.

Goto S, Kato S, Kimura T, Muto A, Himeno H (2011) RsgA releases RbfA from 30S ribosome during a late stage of ribosome biosynthesis. EMBO J 30:104–114.

Goto S, Muto A, Himeno H. (2013). GTPases involved in bacterial ribosome maturation. J Biochem. 153(5):403-14.

Goto,S., Kato,S., Kimura,T., Muto,A. and Himeno,H. (2011) RsgA releases RbfA from 30S ribosome during a late stage of ribosome biosynthesis. *EMBO J.*, 30, 104–114.

Guo Q, et al. (2011) Structural basis for the function of a small GTPase RsgA on the 30S ribosomal subunit maturation revealed by cryoelectron microscopy. Proc Natl Acad Sci USA 108:13100–13105.

Guo Q, et al. (2013) Dissecting the in vivo assembly of the 30S ribosomal subunit reveals the role of RimM and general features of the assembly process. Nucleic Acids Res 41:2609–2620.

Himeno H, et al. (2004) A novel GTPase activated by the small subunit of ribosome. Nucleic Acids Res 32:5303–5309.

Hogardt M, Heesemann J. (2010) Adaptation of Pseudomonas aeruginosa during persistence in the cystic fibrosis lung. Int J Med Microbiol. 300(8):557-62.

Holmes, K. L., and G. M. Culver. 2005. Analysis of conformational changes in 16S rRNA during the course of 30S subunit assembly. J. Mol. Biol. 354:340–357.

Hosokawa K, Fujimura RK, Nomura M (1966) Reconstitution of functionally active ribosomes from inactive subparticles and proteins. Proc Natl Acad Sci USA 55:198–204.

Inoue K, Alsina J, Chen J, Inouye M (2003) Suppression of defective ribosome assembly in a rbfA deletion mutant by overexpression of Era, an essential GTPase in Escherichia coli. Mol Microbiol 48:1005–1016.

Inoue K, Alsina J, Chen JQ, Inouye M. 2003. Suppression of defective ribosome assembly in a *rbfA* deletion mutant by overexpression of Era, an essential GTPase in *Escherichia coli*. *Mol. Microbiol.* 48:1005– 16

Jeganathan A, Razi A, Thurlow B, Ortega J (2015) The C-terminal helix in the YjeQ zinc-finger domain catalyzes the release of RbfA during 30S ribosome subunit as- sembly. RNA 21:1203–1216.

Jeganathan,A., Razi,A., Thurlow,B. and Ortega,J. (2015) The C-terminal helix in the YjeQ zinc-nger domain catalyzes the release of RbfA during 30S ribosome subunit assembly. *J. Mol. Evol.*, doi:10.1261/rna.049171.114.

Jomaa A, et al. (2011) Cryo-electron microscopy structure of the 30S subunit in complex with the YjeQ biogenesis factor. RNA 17:2026–2038.

Jomaa A, et al. (2011) Understanding ribosome assembly: The structure of in vivo assembled immature 30S subunits revealed by cryo-electron microscopy. RNA 17: 697–709.

Jomaa,A., Stewart,G., Mart´ın-Benito,J., Zielke,R., Campbell,T.L., Maddock,J.R., Brown,E.D. and Ortega,J. (2011) Understanding ribosome assembly: the structure of in vivo assembled immature 30S subunits revealed by cryo-electron microscopy. *RNA*, 17, 697–709.

Jomaa,A., Stewart,G., Mears,J.A., Kireeva,I., Brown,E.D. and Ortega,J. (2011) Cryo-electron microscopy structure of the 30S subunit in complex with the YjeQ biogenesis factor. *RNA*, 17, 2026–2038.

Kabash W. 2010. XDS. Akta Crystallogr D Biol Crystallogr. Feb 1; 66(Pt 2): 125–132.

Karbstein K. 2007. Role of GTPases in ribosome assembly. *Biopolymers* 87:1–11

Kuhle B, Ficner R. (2014) A monovalent cation acts as structural and catalytic cofactor in translational GTPases. EMBO J. Nov 3;33(21):2547-63.

Lambert PA. (2002). Mechanisms of antibiotic resistance in Pseudomonas aeruginosa. J R Soc Med 2002;95(Suppl. 41):22–26

Leipe DD, Wolf YI, Koonin EV, Aravind L. 2002. Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* 317:41–72

Leong V, Kent M, Jomaa A, Ortega J (2013) Escherichia coli rimM and yjeQ null strains accumulate immature 30S subunits of similar structure and protein complement. RNA 19:789–802.

Levdikov VM, et al. (2004) The crystal structure of YloQ, a circularly permuted GTPase essential for Bacillus subtilis viability. J Mol Biol 340:767–782.

Lin J, Zhou D, Steitz TA, Polikanov YS, Gagnon MG. (2018). Ribosome-Targeting Antibiotics: Modes of Action, Mechanisms of Resistance, and Implications for Drug Design. Annu. Rev. Biochem. 87:451–78

López-Alonso JP, Kaminishi T, Kikuchi T, Hirata Y, Iturrioz I, Dhimole N, Schedlbauer A, Hase Y, Goto S, Kurita D, Muto A, Zhou S, Naoe C, Mills DJ, Gil-Carton D, Takemoto C, Himeno H, Fucini P, Connell SR. (2017). RsgA couples the maturation state of the 30S ribosomal decoding center to activation of its GTPase pocket. Nucleic Acids Res. 45(11):6945-6959.

Maguire B. 2009. Inhibition of bacterial ribosome assembly: a suitable drug target. *Microbiol.*

Mol. Biol. Rev. 73:22–35

McGowan, J. E., Jr (2006). Resistance in nonfermenting gram- negative bacteria: multidrug resistance to the maximum. Am J Infect Control 34, S29–S37.

Mesaros, N., Nordmann, P., Ple´siat, P., Roussel-Delvallez, M., Van Eldere, J., Glupczynski, Y., Van Laethem, Y., Jacobs, F., Lebesque, P. & other authors (2007). Pseudomonas aeruginosa: resistance and therapeutic options at the turn of the new millenium. Clin Microbiol Infect 13, 560–578.

Milburn MV, Tong L, DeVos AM, Bru¨nger A, Yamaizumi Z, et al. 1990. Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic *ras* proteins. *Science* 247:939–45

Mishra R, Gara SK, Mishra S, Prakash B. 2005. Analysis of GTPases carrying hydrophobic amino acid substitutions in lieu of the catalytic glutamine: implications for GTP hydrolysis. *Proteins* 59:332–38

Morimoto T, Loh PC, Hirai T, Asai K, Kobayashi K, et al. 2002. Six GTP-binding proteins of the Era/Obg family are essential for cell growth in *Bacillus subtilis*. *Microbiology* 148:3539–52

Mulder,A.M., Yoshioka,C., Beck,A.H., Bunner,A.E., Milligan,R.A., Potter,C.S., Carragher,B. and Williamson,J.R. (2010) Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science*, 330, 673–677.

Nichols CE, et al. (2007) Structure of the ribosomal interacting GTPase YjeQ from the enterobacterial species Salmonella typhimurium. Acta Crystallogr Sect F Struct Biol Cryst Commun 63:922–928.

Nierhaus KH. 1991. The assembly of prokaryotic ribosomes. *Biochimie* 73:739–55

Nikolay,R., Schmidt,S., Schlo¨mer,R., Deuerling,E. and Nierhaus,K. (2016) Ribosome assembly as antimicrobial target. *Antibiotics*, 5, 18.

Noeske J, Wasserman MR, Terry DS, Altman RB, Blanchard SC, Cate JH. (2015) High-resolution structure of the Escherichia coli ribosome. Nat Struct Mol Biol. 22(4):336-41.

Noller HF. 2005. RNA structure: reading the ribosome. *Science* 309:1508–14

Nomura M, Erdmann VA. 1970. Reconstitution of 50S ribosomal subunits from dissociated molecular components. *Nature* 228:744–48

Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

Poehlsgaard J, Douthwaite S. (2005). The bacterial ribosome as a target for antibiotics. Nat Rev Microbiol. 3(11):870-81.

Rafay A, Majumdar S, Prakash B. (2012) Exploring potassium-dependent GTP hydrolysis in TEES family GTPases. FEBS Open Bio.Jul 27;2:173-7.

Ramakrishnan V, Moore PB. 2001. Atomic structures at last: the ribosome in 2000. *Curr. Opin. Struct. Biol.* 11:144–54

Ramakrishnan V, Moore PB. Atomic structures at last: the ribosome in 2000. Curr Opin Struct Biol 2001;11:144–154. [PubMed: 11297922]

Razi A, Guarné A, Ortega J. (2017). The cryo-EM structure of YjeQ bound to the 30S subunit suggests a fidelity checkpoint function for this protein in ribosome assembly. Proc Natl Acad Sci U S A. 114(17):E3396-E3403.

Rybtke M, Hultqvist LD, Givskov M, Tolker-Nielsen T. (2015) Pseudomonas aeruginosa Biofilm Infections: Community Structure, Antimicrobial Tolerance and Immune Response. J Mol Biol. 427(23):3628-45

Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, et al. 1997. The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* 277:333– 38

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, et al. 2000. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* 102:615–23

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, et al. 2005. Structures of the bacterial ribosome at 3.5 A° resolution. *Science* 310:827–34

Schuwirth BS, et al. (2005) Structures of the bacterial ribosome at 3.5 A resolution. Science 310:827–834.

Selmer M, et al. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. Science 313:1935–1942.

Shajani Z, Sykes MT, Williamson JR (2011) Assembly of bacterial ribosomes. Annu Rev Biochem 80:501–526.

Sharma MR, et al. (2005) Interaction of Era with the 30S ribosomal subunit: Implica- tions for 30S subunit assembly. Mol Cell 18:319–329.

Shin DH, et al. (2004) Crystal structure of YjeQ from Thermotoga maritima contains a circularly permuted GTPase domain. Proc Natl Acad Sci USA 101:13198–13203.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology. 7:539

Simonetti A, et al. (2008) Structure of the 30S translation initiation complex. Nature 455:416–420.

Sousa AM and Pereira MO. (2014). *Pseudomonas aeruginosa* Diversification during Infection Development in Cystic Fibrosis Lungs—A Review. Pathogens. 3(3): 680–703

SteitzTA,MoorePB.2003.RNA,thefirstmacromolecularcatalyst:Theribosomeisaribozyme.*Trends Biochem. Sci.* 28:411–18

Stokes,J.M., Davis,J.H., Mangat,C.S. and Williamson,J.R. (2014) Discovery of a small molecule that inhibits bacterial ribosome biogenesis. *Elife*, doi:10.7554/eLife.03574.001.

Strateva T and Yordanov D. (2009). Pseudomonas aeruginosa – a phenomenon of bacterial resistance. Journal of Medical Microbiology. 58, 1133–1148

Talkington,M.W.T., Siuzdak,G. and Williamson,J.R. (2005) An assembly landscape for the 30S ribosomal subunit. *Nature*, 438, 628–632.

Thurlow B, et al. (2016) Binding properties of YjeQ (RsgA), RbfA, RimM, and Era to assembly intermediates of the 30S subunit. Nucleic Acids Res 44:9918–9932.

Traub P, Nomura M (1968) Structure and function of E. coli ribosomes, V: Re- constitution of functionally active 30S ribosomal particles from RNA and proteins. Proc Natl Acad Sci USA 59:777–784.

Traub P, Nomura M (1968) Structure and function of Escherichia coli ribosomes, I: Partial fractionation of the functionally active ribosomal proteins and reconstitution of artificial subribosomal particles. J Mol Biol 34:575–593.

Traub P, Nomura M (1969) Studies on the assembly of ribosomes in vitro. Cold Spring Harb Symp Quant Biol 34:63–67.

Tu, C., Zhou, X., Tarasov, S.G., Tropea, J.E., Austin, B.P., Waugh, D.S., Court, D.L., and Ji, X. (2011) The Era GTPase recognizes the GAUCACCUCC sequence and binds helix 45 near the 3^0 end of 16S rRNA. Proc. Natl Acad. Sci. U S A 108, 10156 10161

Tu, C., Zhou, X., Tropea, J.E., Austin, B.P., Waugh, D.S., Court, D.L., and Ji, X. (2009) Structure of ERA in complex with the 3^0 end of 16S rRNA: implications for ribosome biogenesis. Proc. Natl Acad. Sci. U S A 106, 14843 14848

Vetter IR, Wittinghofer A. 2001. The guanine nucleotide-binding switch in three dimensions. *Science* 294:1299–304

Williamson, J. R. 2005. Assembly of the 30S ribosomal subunit. Q. Rev. Biophys. 38:397–403.

Wilson, D.N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev.*

Microbiol. 2014, *12*, 35–48.

Wilson,D.N. and Nierhaus,K.H. (2007) The weird and wonderful world of bacterial ribosome regulation. *Crit. Rev. Biochem. Mol. Biol.*, 42, 187–219.

Wimberly BT, et al. (2000) Structure of the 30S ribosomal subunit. Nature 407: 327–339.

Wittinghofer A, Vetter IR. (2011). Structure-function relationships of the G domain, a canonical switch motif. Annu Rev Biochem. 80:943-71.

Woodson SA (2008) RNA folding and ribosome assembly. Curr Opin Chem Biol 12: 667–673.

Woodson SA (2011) RNA folding pathways and the self-assembly of ribosomes. Acc Chem Res 44:1312–1319.

Xu,Z. and Culver,G.M. (2010) Differential assembly of 16S rRNA domains during 30S subunit formation. *RNA*, 16, 1990–2001.

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, et al. 2001. Crystal structure of the ribosome at 5.5 angstrom resolution. *Science* 292:883–96

Zhang Y, Zborníková E, Rejman D, Gerdes . 2018) Novel (p)ppGpp Binding and Metabolizing Proteins of *Escherichia coli*. MBio. 6;9(2). pii: e02188-17.

Appendix

Review

Molecules that target nucleophosmin for cancer treatment: an update

Adele Di Matteo[1], Mimma Franceschini[2,3], Sara Chiarella[2,3], Serena Rocchio[4], Carlo Travaglini-Allocatelli[4] and Luca Federici[2,3]

[1] Istituto di Biologia e Patologia Molecolari, Consiglio Nazionale delle Ricerche, Rome, Italy

[2] Dipartimento di Scienze Mediche, Orali e Biotecnologiche, Università di Chieti "G. d'Annunzio", Chieti, Italy

[3] Ce.S.I.-MeT Centro Scienze dell'Invecchiamento-Medicina Traslazionale, Università di Chieti "G. d'Annunzio", Chieti, Italy

[4] Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università di Roma "Sapienza", Rome, Italy

Correspondence to: *Luca Federici*, ***email***: *lfederici@unich.it*

Adele Di Matteo, ***email***: *adele.dimatteo@uniroma1.it*

Keywords: *nucleophosmin, B23, acute myeloid leukemia, solid tumours, targeted therapy*

Received: *January 05, 2016* ***Accepted***: *March 28, 2016* ***Published***: *April 05, 2016*

ABSTRACT

Nucleophosmin is a highly and ubiquitously expressed protein, mainly localized in nucleoli but able to shuttle between nucleus and cytoplasm. Nucleophosmin plays crucial roles in ribosome maturation and export, centrosome duplication, cell cycle progression, histone assembly and response to a variety of stress stimuli. Much interest in this protein has arisen in the past ten years, since the discovery of heterozygous mutations in the terminal exon of the *NPM1* gene, which are the most frequent genetic alteration in acute myeloid leukemia. Nucleophosmin is also frequently overexpressed in solid tumours and, in many cases, its overexpression correlates with mitotic index and metastatization. Therefore it is considered as a promising target for the treatment of both haematologic and solid malignancies. NPM1 targeting molecules may suppress different functions of the protein, interfere with its subcellular localization, with its oligomerization properties or drive its degradation. In the recent years, several such molecules have been described and here we review what is currently known about them, their interaction with nucleophosmin and the mechanistic basis of their toxicity. Collectively, these molecules exemplify a number of different strategies that can be adopted to target nucleophosmin and we summarize them at the end of the review.

INTRODUCTION

Nucleophosmin (also known as NPM1, B23, No38, numatrin) is a phosphoprotein, mainly localized at nucleoli [1]. The *NPM1* gene maps to chromosome 5q35 and is expressed in three isoforms through alternative splicing (Figure 1A). Isoform NPM1.1 (P06748-1) (294 residues) is the most abundant one and displays nucleolar localization. Isoform NPM1.2 (P06748-2) lacks an in-frame exon (exon 8) resulting in a shorter protein with respect to NPM1.1 in which an internal segment (residues 195-223) is lacking. The third isoform, NPM1.3 (formerly known as B23.2; P06748-3), uses an alternative exon at the 3' end, which is responsible for a shorter protein construct lacking the last 35 aminoacids with respect to NPM1.1 [2]. This isoform is expressed to low levels and has nucleoplasmic localization. The most abundant NPM1.1 isoform, which will be called NPM1 from now on, is expressed in all tissues. All studies we report here are focused on this isoform.

NPM1 is one of the most abundant proteins in the granular region of nucleoli; it plays a crucial role in maintaining nucleolar structure and is therefore considered one of the "hub" proteins of the nucleolus [3, 4]. NPM1 is involved in many and different cellular functions which have been extensively reviewed recently [5, 6, 7]. Among them NPM1 plays a role in: i) rRNA expression and maturation [8, 9], ii) ribosome assembly and export [10, 11], iii) centrosome duplication [12, 13], iv) DNA replication, recombination, transcription, and repair [1,

5, 6, 14, 15], v) molecular chaperoning for histones and other proteins [5, 16, 17, 18]. Many of these functions are fulfilled through the interaction with different protein partners and indeed NPM1 has been reported to interact with a plethora of proteins [reviewed in 6]. Of particular importance here is the involvement of NPM1 in the p14ARF-HDM2-p53 signaling axis. NPM1 has been reported to interact with all these three proteins in different cellular contexts [19, 20, 21]. In particular NPM1 appears to be the major cellular interactor of p14ARF [22] and to be responsible for p14ARF nucleolar localization in unstressed cells. p14ARF mutants failing to interact with NPM1 are highly unstable and display low anti-proliferative activity [23]. Moreover, overexpressed NPM1 directly interacts with c-MYC and controls c-MYC induced hyperproliferation and transformation activities [24], while being also essential for c-MYC nucleolar localization and c-MYC mediated rRNA transcription [25].

Various aspects of the NPM1 structure, trafficking and post-translational modifications are central to its pleiotropic behavior. NPM1 displays a modular organization in distinct domains each endowed with specific activities (see below). Furthermore, NPM1 mainly resides in nucleoli but can shuttle between nucleoli, nucleoplasm, and cytoplasm thanks to its different localization signals. NPM1 cellular localization during the various phases of the cell cycle or in response to stress signals, as well as the interactions established by NPM1, are all tightly regulated through post-translational modifications. Indeed NPM1 phosphorylation, acetylation, sumoylation, ADP ribosylation and poly-ubiquitination

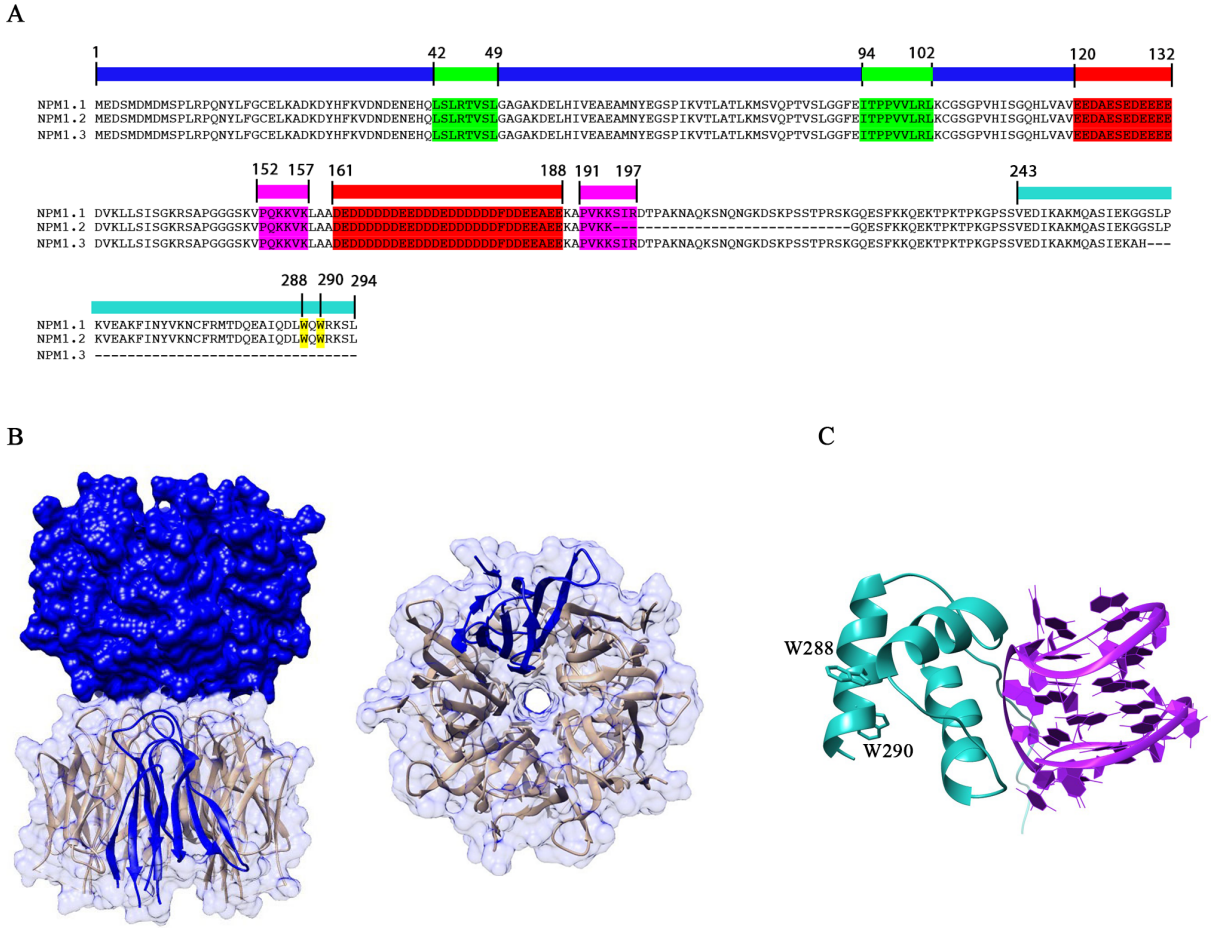

Figure 1: Domain organization and structure of NPM1. A. Primary structures of NPM1.1, NPM1.2 and NPM1.3 are shown. The blue bar marks the N-terminal core domain. Nuclear export signals (NES) in the N-terminal domain are highlighted in green. Acidic-rich regions in the central domain are highlighted in red, while the bipartite nuclear localization signal (NLS) is highlighted in magenta. The cyan bar marks the C-terminal nucleic acid binding domain. Trp288 and Trp290, which form the nucleolar localization signal (NoLS) are highlighted in yellow. Notably NPM1.2 lacks part of the NLS while NPM1.3 lacks most of the C-terminal domain and its NoLS. **B.** Crystal structure of the N-terminal core domain [32]. Five monomers associate to form a pentameric assembly and two pentamers interact in a head-to-head fashion to generate a decamer. **C.** The structure of the C-terminal domain of NPM1 (in cyan) is shown in complex with a G-quadruplex sequence from the *c-MYC* promoter (in magenta) [43]. Trp288 and Trp290 side chains are also highlighted.

have all been reported [reviewed in 5].

NPM1 STRUCTURE

NPM1 shows a modular organization in which three distinct regions can be envisaged: i) the N-terminal region, often referred to as the "core" domain, is mainly responsible for the chaperone activities and for the interaction with protein partners [26, 27, 28]. This region contains two nuclear export signals (NES) (Figure 1A); ii) a central region, predicted to be natively unstructured, which contains a bipartite nuclear localization signal (NLS) (Figure 1A); iii) the C-terminal region of the protein is important for the interactions with nucleic acids [27, 29], contains the nucleolar localization signal (NoLS) (Figure 1A), and is the region where AML-associated mutations occurr [30].

The NPM1 N-terminal core domain shows high similarity to proteins belonging to the nucleoplasmin family [31]. The three-dimensional structures of this domain from human [32] and mouse [28] NPM1 as well as from the homologous Drosophila dNLP [33] and Xenopus NO38 [34] proteins have been determined. They all consist of eight antiparallel β-strands forming a β-barrel with jelly-roll topology. Five monomers tightly associate to form a crown-shaped pentamer (Figure 1B) and two pentamers interact in a head-to-head fashion to form a decamer, with few contacts between the two pentamers (Figure 1B). Whether this decameric assembly is physiologically relevant is presently unknown. NPM1 functions and cellular localization have been reported to be correlated with its oligomerization state [5, 35, 36]. In all determined structures, the pentameric assembly is stabilized by extensive hydrophobic and hydrogen-bonding interactions, ensuring for this domain a high thermal and chemical stability, typical of molecular chaperones [37]. Experimental evidences suggest that ionic strength, divalent ions and interactions with protein partners also contribute to stabilize the pentameric organization of NPM1 [28, 38]. However, a key and opposed role appears to be played by phosphorylation of serine and threonine residues by several different kinases [28]. The sequential phosphorylation of such residues, first those solvent exposed and then those buried at the monomer-monomer interface, greatly destabilizes the oligomeric state of NPM1 promoting monomerization [28]. Monomers are in turn intrinsically unstable and, following pentamer dissociation, completely unfold [28].

The N-terminal core domain is followed by a poorly characterized central region, predicted to be natively unstructured. This region contains two long acidic stretches, spanning residues 120-132 and 161-188 respectively. They are composed of several consecutive glutammate or aspartate residues (Figure 1A) and thought to be important, in cooperation with the N-terminal domain, for the histone chaperoning activity played by NPM1 [31]. Before and after the second acidic stretch, a bipartite nuclear localization signal (NLS) is present (residues 152-157 and 191-197). The terminal part of the region is instead markedly basic and may cooperate with the C-terminal domain in shaping its binding properties. Interestingly, a protein segment comprising the second acidic stretch, the terminal basic region and part of the C-terminal domain (residues 140-259) has been associated to a ribonuclease activity on the internal transcribed segment 2 (ITS2) of 47S ribosomal pre-mRNA, essential for ribosome maturation [10].

The C-terminal domain of NPM1 consists of a three-helix bundle (Figure 1C) [39], stabilized by a set of strictly conserved aromatic residues (Phe268, Phe276, Trp288, and Trp290) [7]. This domain contains the NPM1 nucleolar localization signal (NoLS), encompassing the two tryptophan residues Trp288 and Trp290. This aromatic-rich NoLS is very atypical and, to our knowledge, present only in NPM1. Mutations of one or both Trp288 and Trp290 residues cause unfolding of the three-helix bundle and loss of NPM1 nucleolar localization [39, 40]. Such mutations are typical of AML patients, as it will be detailed later.

The DNA and RNA binding activity of NPM1 is exerted by its C-terminal domain. This was initially shown to have a preference for single stranded DNA and RNA over duplex DNA, however no sequence requirements for nucleic acid binding were uncovered [27, 41]. More recently, NPM1 was shown to specifically target a G-rich sequence at the *SOD2* gene promoter [42] which adopts a G-quadruplex structure under physiological conditions [29]. The structure of NPM1 C-terminal domain in complex with a G-quadruplex oligonucleotide derived from the *c-MYC* promoter was also investigated [43]. It was shown that the three-helix bundle engages the G-quadruplex phosphate scaffold with a positively charged groove located between helices H1 and H2 (Figure 1C). Structural studies also revealed that the terminal part of the central domain, which is unstructured and markedly positively charged, is also necessary for high affinity binding, through both long range electrostatic effects and transient interactions with the G-quadruplex [44, 45]. NPM1 loses its nucleolar localization following lysine acetylation played by p300 [46] and, consistently with structural studies, both lysine residues located in the three-helix bundle at the G-quadruplex interface (Lys250, Lys257 and Lys267) [43] and lysine residues located in the flanking unstructured tail (Lys229 and Lys230) are acetylated by p300 [46].

NPM1 AND CANCER

NPM1 is over-expressed in a variety of solid tumours of different histological origin including prostate [47], liver [48], thyroid [49], colon [50], gastric [51], pancreas [52], glioma and glioblastoma [53, 54],

astrocytoma [55] and others. NPM1 overexpression often correlates with mitotic index and metastatization and it has been proposed as an adverse prognostic marker in a number of such malignancies [5, 56]. Even though amplification of the *NPM1* locus has never been shown, the *NPM1* gene is a target of the oncogenic c-MYC, which stimulates NPM1 expression by direct binding to the *NPM1* promoter [57].

The specific contribution of overexpressed NPM1 to cancer development is not fully understood but it may arise from multiple factors. First, NPM1 overexpression is correlated to increased ribosome biogenesis and protein synthesis and both these functions are amplified in tumour cells [58]. Furthermore, NPM1 plays a role in stimulating DNA repair following oncogene activation and reduces apoptotic or senescence response [5, 6, 59]. Accordingly, a model has been recently proposed [5] whereas, when oncogene activation arises in a normal cell as a first genetic event, overexpression of NPM1 may contribute to reinforce the DNA damage response thus keeping DNA damage and the consequent genomic instability to a level that the cell can sustain. This would in turn allow cells to select for the cooperative mutations necessary for transformation.

NPM1 is also heavily implicated in haematological malignancies, being its gene both the target of different chromosomal translocations or of frequent mutations. In 30% of anaplastic large cell lymphoma (ALCL) patients a t(2;5) translocation fuses the 5' end of NPM1 gene with the 3' portion of the ALK (anaplastic limphoma kinase) gene. This leads to the expression, in the cytoplasm of cancer cells, of a chimeric protein consisting of the NPM1 N-terminal oligomerization domain fused to the ALK tyrosine kinase domain [60, 61]. This chimera is a major driver in ALCL tumourigenesis [62] and the role of the NPM1 moiety is thought to be that of facilitating, through its oligomerization, the dimerization and thus the constitutive activation of the ALK tyrosine kinase domain.

A second rare event was found in acute promyelocytic leukemia (APL) patients. Here, as a consequence of a t(5;17)(q35;q31) translocation, the NPM1 N-terminal domain is fused to the DNA-binding domain of retinoic acid receptor α (RARα). Also in this case, and similarly to the most common PML-RARα chimera, NPM1 facilitates dimerization of the RARα moiety thus interfering with the RARα transcriptional activity [63]. The resulting arrest of myeloid differentiation may be reversed by treatment with all-trans retinoic acid (ATRA) [64].

A third chromosomal translocation at t(3;5)(q25;q35) has been identified in a small subgroup (less than 1%) of acute myeloid leukemia (AML) patients. This event generates a protein chimera comprising the first 175 aminoacids of NPM1 (the N-terminal domain plus a portion of the central unstructured region) and the entire coding sequence of the protein MLF1 (myelodisplasia/myeloid leukemia factor 1) [65]. The role of this chimera in tumourigenesis has not been fully elucidated.

Beside the specific role played by the NPM1 moiety in the different chimeras, in all cases haploinsufficiency for the *NPM1* WT gene is generated and partial dislocation of the NPM1 WT protein in the cytosol is observed, as a consequence of heteroligomerization with protein chimeras. Interestingly, NPM1 haploinsufficiency is also observed in myelodisplastic syndromes with 5q deletion [66], suggesting that it may confer *per se* a proliferative advantage in the myeloid lineage.

In 2005 the *NPM1* gene was identified as the most frequently mutated one in AML, accounting for around 60% of patients with normal kariotype and 35% of total cases [30]. Mutations map to the last exon of the gene and are always heterozygous [67, 68]. More than 30 different mutations have been identified but the consequences at the protein level are similar in all cases: due to duplication or insertion of short nucleotide stretches at exon XII of the NPM1 gene, the reading frame is altered leading to a protein that has acquired four additional residues at the C-terminus and has a different sequence in the last seven residues. Both triptophans 288 and 290, or only Trp288 in some unfrequent mutants, which constitute the nucleolar localization signal (NoLS), are replaced and the whole C-terminal domain of the protein is totally unfolded or largely destabilized [39, 69, 70, 71]. Furthermore, the newly generated sequence forms a novel NES which reinforces the two already present at the N-terminal region of the protein. Disruption of the NoLS and appearance of a new NES account for the aberrant cytoplasmic localization of mutated NPM1 [40], which is the hallmark of this kind of leukemia (hence mutated NPM1 is also termed NPM1c+, from cytoplasmic positive). Furthermore, since NPM1c+ oligomerizes with the wild-type protein produced by the normal allele, through its unaltered N-terminal domain, the majority of wild-type NPM1 is also translocated in the cytosol and only a small fraction still resides in the nucleoli of leukemic blasts. A wealth of different data [72, 73] suggests that NPM1 mutations act as a founder genetic lesion in this kind of leukemia and therefore AML with NPM1 mutation has been included as a new provisional entity in the WHO 2008 classification of myeloid neoplasms [74]. AML with mutated *NPM1* may be further stratified into two different categories: those patients where concomitant *FLT3-ITD* (FMS-like tyrosine kinase internal tandem duplication) is absent, usually respond to standard induction therapy and have favourable prognosis; when *FLT3-ITD* sums up to *NPM1* mutations (around 30% of cases) the prognosis is much worse. In all cases relapse is frequent and *NPM1* mutations are typically present at relapse [75].

The exact mechanism through which NPM1c+ exerts its transforming activity is not yet fully understood, but all evidences point to the hypothesis that "placing a critical regulator at the wrong place in the wrong time"

may be the driving force [76]. In particular NPM1c+ may confer a proliferative advantage to blasts through the gain of several and unwanted "new" cytosolic functions. A number of them have been elucidated. For instance NPM1c+ interacts, through its unchanged N-terminal domain, with the tumour suppressors p14ARF and Fbw7γ, the major E3-ubiquin ligase targeting c-MYC, which are both translocated in the cytosol and there proteasomally degraded [77, 78]. Therefore, an important tumour suppressor pathway is hampered while, at the same time, an oncogene product like c-MYC is stabilized. Also, NPM1c+ binds and inhibits caspases 6 and 8, thus directly impacting on the execution of apoptosis [79]. Furthermore, NPM1c+ binds and inhibits the PTEN deubiquitinating enzyme HAUSP, resulting in PTEN cytoplasmic polyubiquitinilation and degradation [80]. Thus a third important tumour suppressor is also deregulated by NPM1c+ [80]. These cytosolic activities promote leukemic blasts survival and may also counteract the nucleolar stress caused by NPM1 delocalization. Additionally, the unfolded C-terminal domain of NPM1c+ has been recently shown to contain amyloidogenic regions that may also contribute to the gain of cytosolic functions played by the protein [81].

Targeting NPM1 for cancer treatment

A comprehensive description of the pleiotropic functions played by NPM1 and its role in cancer has been the subject of a number of excellent reviews to

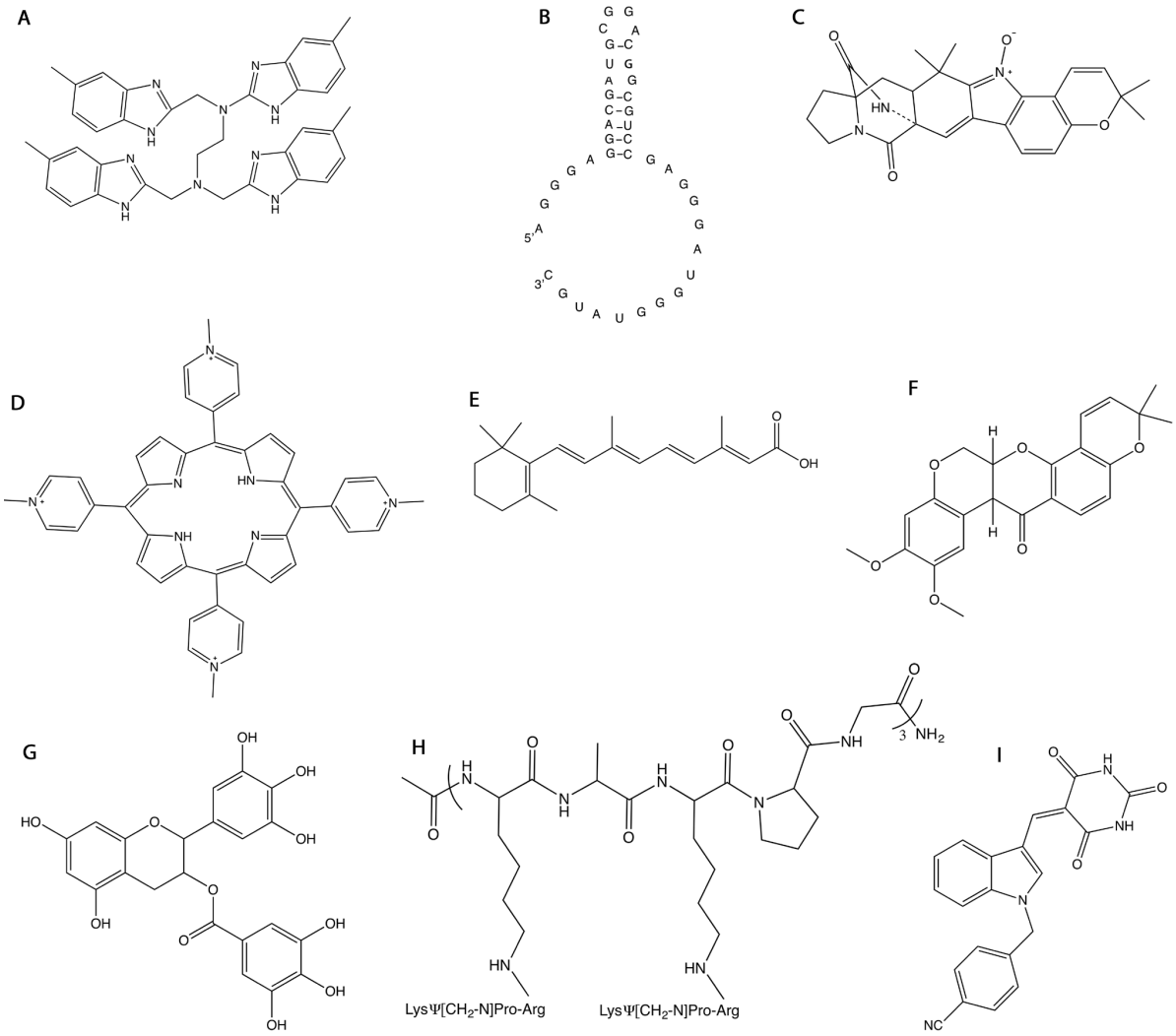

Figure 2: Structures of molecules that target NPM1. A. NSC348884. **B.** 1A1(1-40) RNA aptamer. **C.** (+)-avrainvillamide. **D.** TmPyP4. **E.** All trans-retinoic acid (ATRA). **F.** deguelin. **G.** (-)-epigallocatechin-3-gallate (ECGT). **H.** NucAnt 6L. **I**.YTR107.

which we refer the reader [1, 5, 6, 7, 71, 72]. A common conclusion in all these analyses is that NPM1 should be considered as a target for the treatment of several tumours, noticeably haematological malignancies where the *NPM1* gene is mutated or found at the junction of chromosomal translocations, but also solid tumours where the gene is overexpressed.

Interestingly, over the course of the last 10 years, several molecules that target NPM1 have been indeed discovered and their effect and therapeutic potential has been investigated to various extent. Therefore we thought that it might have been timely and appropriate to review here what is currently known about these molecules, their interaction with NPM1 and the mechanistic basis for their toxicity.

NSC348884

NPM1 oligomerizes through interactions mediated by its N-terminal domain. Loss of oligomerization has been shown to promote unfolding of the domain [28], suggesting a consequent impairment of its functions. The hypothesis of targeting the oligomerization properties of NPM1 to interfere with its functions, including its antiapoptotic activities, led to the identification of NSC348884, the first small molecule inhibitor reported to specifically interact with NPM1 [82]. A pharmacophore hypothesis was devised from the analysis of the hydrophobic interface between monomers in the pentameric ring and used to screen "in silico" a large library of compounds. NSC348884 ((di-[((6-methyl-1H -benzo[d]imidazol-2-yl)methyl)((5-methyl-3-oxo-3H -indol-2-yl)methyl)]) aminoethane) (Figure 2A and Figure 3) was the best hit and used for subsequent functional studies. NSC348884 was initially shown to promote monomerization of NPM1 in LNCaP (androgen-sensitive prostate adenocarcinoma) and HCT116 (colorectal carcinoma) cell lines, in which NPM1 is wild-type and highly expressed. Cell viability assays, conducted on LNCaP and Granta (mantle cell lymphoma) cells showed cellular toxicity with IC_{50} varying in the 1.5-4 micromolar range (Table 1). Subsequent experiments demonstrated a remarkable synergic effect of NSC348884 with doxorubicin. The combined used of sub-cytotoxic doses of both drugs led to a complete loss of cell viability [82]. These promising data led to the investigation of the mechanistic basis of cellular toxicity. First, it was shown that NSC348884 counteracts the anti-apoptotic activity of over-expressed NPM1 and promotes apoptosis in LNCaP and Granta cells in a dose-dependent fashion, as seen both by morphologic analysis and annexin V staining. It is well known that NPM1 knockdown results in increased levels of p53 and of its phosphorylation at the Ser15 site [83]. Similarly, NSC348884 treatment was shown to exert both effects and also to elevate levels of p21, a key transcriptional target of p53. NPM1 is known to interact with the tumour suppressor p14ARF and to sequester it in nucleoli. When moving to the nucleoplasm, p14ARF interacts with HDM2, the E3-ubiquitin ligase for p53, resulting in elevated p53 levels [84]. NPM1 has also been shown to directly interact with p53 and prevent its phosphorylation at Ser15 [83]. Both of these NPM1 antiapoptotic activities may be compromised by treatment with NSC348884 thus explaining its apoptotic effect in cells overexpressing NPM1. Similar results, i.e. NPM1 monomerization and cell growth inhibition, were also recently shown in the hepatic carcinoma HepG2 cells treated with NSC348884 [85].

The effect of NSC348884 in AML was analysed by comparing cells expressing NPM1c+ (OCI-AML3) and cells expressing only wild type NPM1 (HL-60 and OCI-AML2) [86]. Interestingly, NSC348883 was found to be more effective in disrupting NPM1 oligomerization in OCI-AML3 cells compared to HL-60 and OCI-AML2 cells. This is somewhat unexpected since mutations target the C-terminal domain of the protein while NSC348884 targets the N-terminal domain. Furthermore, NSC348883 markedly induced higher levels of apoptosis in OCI-AML3 cells than in the NPM1 wt cells. A synergic effect was found in the combined treatment of OCI-AML3 cells or primary AML cells expressing NPM1c+ with NSC348884 and all-*trans*-retinoic acid (ATRA), but not in cells expressing wild-type NPM1 only. However, NSC348883 treatment alone or co-treatment with ATRA was significantly less effective in inducing apoptosis in primary AML cells co-expressing NPM1c+ and FLT3-ITD, consistent with the poorer prognosis of patients carrying both alterations with respect to patient with NPM1c+ only [86].

Overall, these data suggest that molecules targeting NPM1 oligomerization may be effective against both solid malignancies overexpressing NPM1 and in AML with NPM1c+ expression, especially when used in combination with established chemotherapy. This therapeutic strategy may be also attractive for tackling haematological disorders caused by chromosomal translocations in which the N-terminal region of NPM1 is fused to other proteins, i.e. NPM-ALK, NPM-RARα and NPM-MLF1. As already discussed above, in such chimeric oncoproteins the role of the NPM1 moiety is probably that of favouring, through its oligomerization, the dimerization of the fusion partners, which are then constitutively activated. However, NSC348884 has not been yet tested in such tumours.

REV-NLS

Different studies have suggested that NPM1 behaves as a nucleolar "hub", favouring the localization and accumulation of its protein partners in the nucleolus thanks to reciprocal interactions [87]. Although extensive biochemical and structural characterization of these interactions is still lacking and many observations

available in the literature are conflicting, various studies have identified NPM1 and in particular its N-terminal region as one of the receptors responsible for the nucleolar localization of many proteins (Figure 3) [28].

This topic was initiated by studies on the interaction between NPM1 and the Rev protein from the human immunodeficiency virus-1 (HIV-1). Rev is a protein essential for virus replication and was initially shown to interact with NPM1 [88, 89]. Later on, the Rev sequence recognized by NPM1 was identified and shown to coincide with the highly basic sequence necessary for Rev nuclear/nucleolar localization (i.e. its NoLS) [90]. To analyse the effect played in cancer cells by interfering with NPM1 protein-protein interactions, different Rev peptides were administered to Ras-3T3 cells [91]. In particular, the Rev37-47 peptide (ARRNRRRRWRE), which binds *in vitro* NPM1 with submicromolar affinity [90] was shown to be active with GI_{50}=95.1 μM (Table 1). Other Rev-derived peptides with reduced or no affinity for NPM1 had virtually any effect. Rev37-47 also efficiently inhibited colony formation in soft agar, suggesting that the peptide could revert the transformed phenotype of Ras-3T3 cells to a normal phenotype. Further experiments on nude mice inoculated subcutaneously with Ras-3T3 cells confirmed the efficacy of Rev37-47 to consistently reduce tumour growth. Finally Rev37-47 was shown to synergize with doxorubicin in reducing tumour growth in the xenografts. Similarly to what already shown with drugs that influence the oligomeric state of NPM1, treatment with Rev37-47 results in p53 increased levels and transcriptional activity. Consistently, the cytotoxic effect of this Rev peptide was abrogated by siRNA directed against p53 [91].

Overall, though very preliminary, these data suggest that a molecule targeting the NPM1 surface that interacts

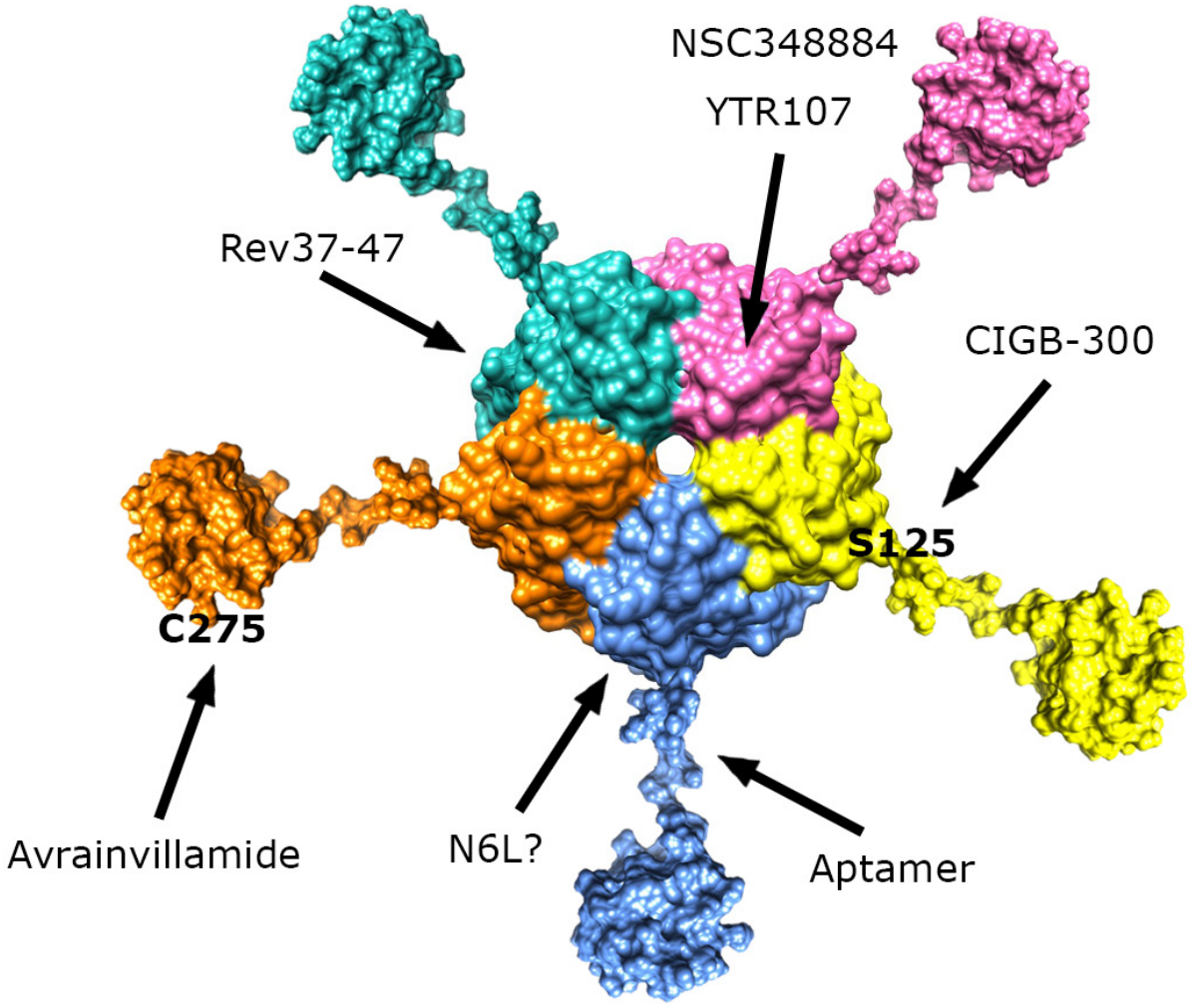

Figure 3: Schematic representation of full-length NPM1 structure showing the pentamer formed by the N-terminal domains, followed by the central unstructured regions and culminating with the folded C-terminal domains. The sites recognized by NPM1-interacting molecules are indicated, when known.

with HIV Rev, plays a toxic effect on cancer cells. Since NPM1 interacts with many proteins, the protein-protein interaction surface targeted by Rev37-47 may be common to other NPM1 partners (Figure 3). Therefore a single molecule may compromise many of the interactions established by NPM1 at once.

RNA APTAMERS

Aptamers are small synthetic RNA or single stranded DNA molecules that can bind and inhibit with high affinity and specificity target molecules. Many aptamers are currently in development and have already undergone clinical trials as promising therapeutic tools against different diseases including cancer [92].

With the aim of inhibiting the antiapoptotic activity of NPM1, an effort to select functional aptamers was paid by using the technique known as "systematic evolution of ligands by exponential enrichment" (SELEX) [93]. This procedure, performed against full-length NPM1, led to the identification of the 1A1 RNA aptamer and its truncated 40-mer form 1A1(1-40) (Figure 2B), both able to bind NPM1 with K_D=33 nM and K_D=30 nM, respectively (Table 1) [94]. Subsequent efforts were paid to identify the exact NPM1 domain targeted by the aptamer *in vitro*. These studies showed that the isolated NPM1 central region (residues 114-187) was bound with same affinity as full-length NPM1 while the N-terminal (1-113) and C-terminal (188-294) domains were not bound by the aptamer (Figure 3). Further experiments performed *in vitro* with full-length NPM1 demonstrated that the aptamer interferes with NPM1 oligomerization. It is well-known that the isolated N-terminal domain of the protein, which is not bound by the aptamer in isolation, is *per se* sufficient to form very stable pentamers and dimers of pentamers [27, 32, 33]. Therefore, further experiments are necessary to understand the dependence of NPM1 oligomers stability on the interplay among different domains in the context of the whole protein.

The effect of aptamers was assayed in different cancer cell lines including MCF-7 (human breast adenocarcinoma). It was confirmed that aptamers interact with NPM1 also *in vivo* and promote monomer accumulation and oligomer depletion. Importantly, immunofluorescence studies indicated that, upon aptamers expression, NPM1 delocalizes from nucleoli to the nucleoplasm [94]. Since it is known that the nucleolar localization signal is localized at the very C-termini of the protein (Trp288 and Trp290), an area far from the putative aptamers binding site, also this effect awaits for a structural explanation; it is possible that currently unknown interactions between NPM1 domains may be destabilized by aptamer binding and interfere with NPM1 ability to associate with nucleoli. From the cell viability point of view, the expression of the aptamers caused an increase of apoptotic cells, comparable to what seen with siRNA mediated NPM1 down-regulation. A synergistic effect in causing apoptosis with the DNA damaging drugs etoposide and cisplatin was also observed. Mechanistically, it was shown that aptamer expression is followed by p14ARF accumulation in the nucleoplasm, p53 increased levels and p21 expression, similarly to what observed with NSC348883 [82].

CIGB-300

Many protein kinases are established targets in cancer therapy and several kinase inhibitors already entered the clinic or are undergoing clinical trials. Among the protein kinases that could be targeted one is the Ser/Thr Casein kinase 2 (CK2). In fact, high levels CK2 have been found in different cancer cells [95, 96] especially those which show remarkable resistance to death, being this protein a major player in apoptosis suppression [97].

CIGB-300 is a cyclic peptide fused, at the N-terminus, to a cell-penetrating peptide derived from the HIV Tat protein (GRKKRRQRRRPPQ-β-ala-CWMSPRHLGTC with a disulphide bond between the two cysteine residues). This compound was derived by screening a random cyclic peptide phage library against the HPV-16 E7 oncoprotein site targeted by CK2 for phosphorylation [98]. Therefore, rather than being a direct CK2 inhibitor, CIGB-300 was selected for interfering with CK2 phosphorylation by interacting with one of its targets. Consistently, it was shown that treatment with CIGB-300 induces apoptosis in several tumour cell lines (Table 1). Furthermore CIGB-300 was shown to significantly reduce tumour growth in syngeneic C57BL6 mice implanted with TC-1 lung epithelial tumour [98]. This initial observation stimulated further studies aimed at identifying the protein(s) targeted by CIGB-300 in vivo. Pull-down experiments in NCI-H82 cells (small cell lung cancer) identified 20 proteins: two known CK2 substrates, NPM1 and nucleolin as well as several ribosomal proteins. Subsequent experiments demonstrated that CIGB-300 directly binds NPM1 and not nucleolin *in vivo,* and suggested that nucleolin and ribosomal proteins pull-down may be mediated by NPM1 [99]. Using an *in vitro* CK2 phosphorylation assay, it was shown that NPM1 phosphorylation inhibition is due to direct binding of CIGB-300 to NPM1. In particular, CIGB-300 inhibits CK2 phosphorilation of NPM1 at Ser125 and rapidly co-localize with NPM1 at nucleoli after administration (Figure 3). Nucleolar disassembly was also observed after treatment with CIGB-300 as monitored by the rapid nucleoplasmic redistribution of both NPM1 and fibrillarin. CK2 was previously shown to be a master regulator of ribosome biogenesis [100]. Furthermore, CK2 inhibition by DRB was shown to trigger nucleolar breakdown and interfere with ribosome biogenesis. Notably similar effects were also observed by mutating NPM1 Ser125 [101]. These observations suggest that the activity of

Table 1: Molecules that interact with NPM1

Molecule	Efficacy	Target Cells	Clinical Trails	Site of interaction	Key ref.
NSC34884	IC_{50} between 1,4-4,0 μM	LNCaP (prostate adenocarcinoma) HCT116 (colorectal carcinoma) Granta (mantle cell lymphoma) HepG2 (hepatocellular carcinoma) OCI-AML3 (AML with NPM1c+)		oligomerization surface of the N-terminal domain	82,85,86
Rev37-47	GI_{50}=95,1 μM	Ras-3T3 (Ras transformed mouse fibroblasts)		N-terminal domain	91
RNA aptamers 1A1 and 1A1(40)	K_D= 30-33 nM (IC_{50} or GI_{50} not available)	HeLa (cervical carcinoma) MCF-7 (breast cancer) SGC7901 (gastric cancer)		Central Region (114-187)	94
CIGB-300	IC_{50} between 20-180 μM	H-125 (non-small cell lung) H-82 (small cell lung cancer) HeLa (cervical carcinoma) Jurkat (T-cell leukemia) MCF-7 (breast cancer) Hep-2 (laryngeal carcinoma) TC-1 (murine lung carcinoma)	Phase I (locally advanced cervical cancer) Phase II (squamous cell carcinoma or cervix adeno carcinoma)	Binds Ser125 and inhibits phosphorilation by CKII	98,99 106,107
Avrainvillamide	GI_{50} between 0,33-0,52 μM	T-47D (breast carinoma) LnCaP (prostate adenocarcinoma) OCI-AML2 (AML with wtNPM1) OCI-AML3 (AMl with NPM1c+)		C-terminal domain, alkylates Cys275	108,109 110
TmPyP4	IC_{50}=0,2 μM (with halogen (irradiation at 7,2 J/cm^2) GI_{50}≈15 μM	B78-H1 (murine melanoma) SW480 (colon carcinoma)		binds rDNA G-quadruplexes recognized by C-terminal domain	113,114 116,117
ATRA/ATO	IC50≈1 μM	OCI-AML3 (AMLwith NPM1c+) IMS-M2 (AMLwith NPM1c+) AML patients primary blasts		promote degradation of NPM1c+ by oxidation of Cys288	120,121
Deguelin	IC_{50}=1,49 μM	OCI-AML3 (AMLwith NPM1c+)		promote degradation of NPM1c+ with unknown mechanism	124
EPTG	IC_{50}=5-20 μM (according to initial cell density)	IMS-M2 (AMLwith NPM1c+)		promote degradation of NPM1c+ with unknown mechanism	125
NucAnt 6L	GI50=5-38 μM K_D= 1nM	T29 (murine T-cell lymphoma) MOLT-4 (ALL) Jurkat (T-cell leukemia) A20 (murine B-cell lymphoma) HL-60 (APL) HCT116 (colorectal carcinoma) MDA-MB231(mammary gland Adenocarcinoma) MDA-MB435 (melanoma) U87-MG (glioblastoma) U373-MG (glioblastoma)	Phase I/IIa (multiple solid tumors)	binds full-length NPM1 (individual domains not tested)	126,127
YTR107	not available	HT29 (colorectal carcinoma) D54 (glioblastoma) PANC-1 (pancreatic duct adenocarcinoma)		oligomerization surface of the N-terminal domain	129,130 131

CK2 as a master regulator of nucleolar assembly and ribogenesis is operated through NPM1 as a downstream effector. CK2 mediated phosphorylation of NPM1 has also been shown to regulate its chaperone activity [102]. NPM1 inhibition might therefore interfere with the proper folding of ribosomal proteins, their correct loading on the nascent ribosome, but also with histone assembly at nucleolar rDNA. Even though the precise mechanism connecting nucleolar breakdown to apoptosis is not clear, data obtained with CIGB-300 support the idea that NPM1 phosphorylation may be a target for cancer treatment in those tumours were high levels of NPM1 are present. CIGB-300 exerts a broad antiproliferative effect on cell lines derived from breast, cervical, lung, colon, prostate cancer and chronic lymphocytic leukemia (CLL) while a robust antitumour effect was also observed *in vivo* in mouse models of cervical and lung cancer as well as CLL [98, 103, 104]. Recently it was also observed that the concomitant administration of CIGB-300 and drugs like cisplatin, paclitaxel, doxorubicin or 5-fluorouracil cisplatin gives rise to synergic chemotherapeutic effects in lung and cervical cancer models [105].

In clinical research, CIGB-300 has been established to provide some benefits and to be safe and tolerable by local injection in cervical malignancies [106]. More recently, another phase I clinical study on patients with stage 1B2/II cervical cancer allowed the estimation of the maximum tolerated dose (MTD) and the pharmacokinetics/biodistribution profiles for CIGB-300 following local administration [107]. Collectively these

data make of CIGB-300 the NPM1-targeting drug that has been more deeply investigated so far.

AVRAINVILLAMIDE

(+)-avrainvillamide is an alkaloid (hereafter avrainvillamide; Figure 2C), firstly isolated from a marine fungal strain of *Aspergillus sp*. and initially shown to display antiproliferative activity on a panel of cancer cell lines [108]. In 2007 it was reported that avrainvillamide was able to form tight complexes with NPM1 *in vivo* [109]. Avrainvillamide is thought to act as an electrophile centre subjected to the nucleophilic addition of a sulphur group to its unsaturated nitrone moiety. Consistently, the NPM1-avrainvillamide complex was disrupted by iodoacetamide treatment. Binding assays on site-directed mutants revealed that avrainvillamide binds NPM1 primarily by alkylating Cys275, which is located in helix H2 of the C-terminal three-helix bundle (Figure 1 and Figure 3). This makes this alkaloid the first and only small molecule inhibitor known to directly bind the NPM1 C-terminal DNA-binding domain (Table 1). Cellular studies reported that HeLa S3 cells transfected with a siRNA targeting NPM1 exhibited enhanced sensitivity to avrainvillamide, providing a correlation between the antiproliferative effects of avrainvillamide and levels of NPM1. A following recent paper thoroughly investigated the effect of avrainvillamide-NPM1 association in AML [110]. It was shown that avrainvillamide binds both NPM1wt and NPM1c+ in vitro, with increased efficiency for the latter, probably due to its unfolded C-terminal domain. The effect of avrainvillamide on NPM1 cellular localization was studied both in cells carrying wild-type and predominantly nucleolar NPM1 (OCI-AML2) and in cells carrying heterozygous NPM1 mutation and the NPM1c+ phenotype (OCI-AML3). Treatment with sublethal doses of avrainvillamide (250 nM) was not effective in displacing NPM1 from nucleoli in OCI-AML2 cells. Conversely, NPM1c+ was observed to partially relocalize in nucleoli when OCI-AML3 cells were challenged with the same dose of avrainvillamide. Higher doses of avrainvillamide (4 μM) caused an apoptotic phenotype in both cell lines with condensed or fragmented nuclei. The cellular localization of NPM1 was however markedly different in the two cell lines. While NPM1 was displaced from nucleoli in OCI-AML2 apoptotic cells, it appeared to co-localize with the highly condensed nuclei of OCI-AML3 cells. Interestingly, the nucleolar relocalization effect of avrainvillamide on the mutant form of NPM1 was confirmed by transfecting cells with an eGFP-NPM1c+ construct. Conversely, treatment with the Crm1-exportin1 inhibitor leptomycin, resulted in the nucleoplasmic but not nucleolar accumulation of the eGFP-NPM1c+ construct. It was also shown that avrainvillamide binding to the mutated C-terminal domain at Cys275 does not result in domain refolding, which would have explained the re-localization effect. Therefore, avrainvillamide appears to act as a surrogate for the compromised nucleolar localization signal (NoLS) in NPM1c+ even in the presence of an unfolded C-terminal domain [110], through a mechanism that is currently unknown.

TMPYP4

TmPyP4 (tetra-N-methyl-pyridyl porphyrin) (Figure 2D) is a positively charged porphyrin that binds G-quadruplex DNA with high affinity [111]. The potential of this molecule in cancer therapy has been recently investigated by different groups (Table 1) [111, 112, 113, 114].

NPM1 was recently recognized as a G-quadruplex binding protein and initially implicated in binding G-quadruplex sequences from the *SOD2* and *c-MYC* gene promoters [29, 42, 43]. These initial findings prompted investigations aimed at assessing whether there might be a link between DNA binding, especially at G-quadruplex regions, and nucleolar localization. Nucleoli are nuclear regions enriched in proteins and RNAs that are organized in correspondence of and around tandemly duplicated ribosomal DNA genes; bioinformatics and *in vitro* analysis of rDNA suggested that this gene contains as much as 14 different G-quadruplex regions in the non-template strand [115].

NPM1 was initially found to be associated to the rDNA throughout the whole gene [9] and further studies revealed that rDNA G-quadruplexes are effectively bound by NPM1 both *in vitro* and *in vivo* [116]. This activity is played by the C-terminal domain of the protein and depends critically on its folded state. It was shown that the C-terminal domain of the AML-linked NPM1 mutant A form, which lacks both Trp288 and Trp290, is unfolded and unable to bind rDNA G-quadruplexes, while the reinsertion of the two tryptophan residues was sufficient to restore the correct folding, G-quadruplex binding and nucleolar localization [40, 116]. Among the possible strategies for treating AML with mutated NPM1 a so-called "NPM1 nucleolar starvation" hypothesis was suggested [7, 71]. This is based on the observation that mutations are always heterozygous and a small but detectable fraction of the wild-type protein always resides in nucleoli of leukemic blasts, possibly because it is necessary for nucleolar functions such as ribogenesis and/or for the correct assembly of nucleoli. Therefore depriving nucleoli of NPM1 might cause nucleolar stress and induce an apoptotic response. A lower dosage of such NPM1-nucleoli depleting agents might be toxic for AML blasts with respect to healthy cells given their limited residual NPM1 pool. This hypothesis was tested using TmPyP4 as an investigational tool on OCI-AML2 leukemia cells and OCI-AML3 cells expressing NPM1c+. First it was shown, on both OCI-AML2 and OCI-AML3 cells, that TmPyP4

effectively displaces NPM1 from nucleoli, even at sub-cytotoxic amounts [116, 117]. Indeed, nucleolar protein content is heavily affected by TmPyP4 since two other important nucleolar proteins, i.e. nucleolin and fibrillarin, are also displaced. However, TmPyP4 was found to be more toxic on OCI-AML2 than on OCI-AML3 cells, both as a function of dose and time, despite the limited amount of nucleolar NPM1wt in OCI-AML3 cells as compared to OCI-AML2 [117]. Mechanistically, it was shown i) that OCI-AML3 cells contain reduced levels of both NPM1 and p53 as compared to OCI-AML2, ii) that levels of p53 in both cell lines decreased in the presence of TmPyP4 and iii) that p53 was activated, as monitored by elevation of p21 mRNA levels, in OCI-AML2 but not in OCI-AML3 cells [117]. This is possibly linked to p14ARF inhibition of HDM2 in OCI-AML2 cells following TmPyP4 treatment; an event that may not happen in OCI-AML3 cells since p14ARF is delocalized in the cytoplasm by NPM1c+ and degraded [78].

ATRA/ARSENIC OXIDE, DEGUELIN AND (-)-EPIGALLOCATECHIN-3-GALLATE

All-trans-retinoic-acid (ATRA) (Figure 2E), alone or in combination with arsenic trioxide (ATO; As_2O_3), is currently the frontline treatment for acute promyelocytic leukemia (APL) harbouring the PML-RARα gene rearrangement [118, 119]. Since the use of these agents has provided excellent results in APL, the therapeutic potency of these two compounds either in association with one another or with other cytotoxic agents has been also recently investigated in the most common type of AML i.e. the one with mutated *NPM1* gene and NPM1c+ expression [120, 121]. Initially it was shown that ATO, but not ATRA, is effective in inducing apoptosis in AML cell lines carrying NPM1 mutation A (NPM1c+ phenotype), both OCI-AML3 and IMS-M2, as compared to AML cells with NPM1wt only. When ATO was combined with ATRA a striking cooperative action in inducing apoptosis was detected in OCI-AML3 cells. These results were also replicated using patients' primary blasts. The primary blasts carrying the concurrent *FLT3-ITD* mutation were more susceptible than those where *FLT3* was not affected and this is particularly important given the worse prognosis of these patients with respect to those carrying *NPM1* mutation only. Furthermore pre-treatment with ATO/ATRA was shown to greatly sensitize cells to daunorubicin, which is currently used in AML induction therapy [120]. Similar results were also obtained by El Hajj and coworkers [121]. ATO is known to target the promyelocitic leukemia protein (PML) and this was confirmed also in AML blasts carrying NPM1c+. Importantly it was also shown that ATO treatment results in decreased levels of NPM1c+ while levels of NPM1wt remain unchanged [120]. Surprisingly this effect was exerted also by ATRA and a synergistic behaviour in the combined treatment was observed. Pre-treatment with the proteasome inhibitor MG132 reverted this phenotype, suggesting that NPM1c+ disappearance after ATO/ATRA treatment is due to proteasomal degradation. Further experiments suggested that NPM1c+ specific degradation may be triggered by the oxidative stress induced by ATO. In particular, oxidation of NPM1c+ Cys288 (which replaces the tryptophan residue of the wild type protein) could be the reason that makes the mutant protein more susceptible to proteasomal-mediated degradation with respect to wild-type [120].

These results are very promising, especially if we consider that both ATRA and ATO are already in the clinic, but further studies are required to reach a thorough biochemical characterization of the effects played by these drugs in AML with *NPM1* mutations.

Therapeutic strategies focused on NPM1c+ specific degradation have also been invoked in studies centred on two natural compounds, i.e. deguelin (Figure 2F) and (-)-epigallocatechin-3-gallate (ECGT) (Figure 2G). Deguelin is a rotenoid molecule initially isolated from the African plant *Mundulea sericea* (Leguminosae) and later from other plants, which displayed potent and selective apoptotic and antiangiogenic effects on a variety of cancer cells (lung, prostate, gastric, and breast cancer) [122]. ECGT is the major polyphenol extracted from green tea, widely investigated for its antioxidant properties and for its possible effects in cancer prevention [123]. In both cases it was shown that treatment of OCI-AML3 cells with deguelin [124] and of IMS-M2 cells (also expressing NPM1c+) with ECGT [125], was effective in reducing NPM1c+ but not NPM1wt levels and in inducing apoptosis. NPM1 levels were instead not consistently affected when treating AML cell lines expressing NPM1wt only [124]. Nothing is known however about the mechanistic basis of NPM1c+ specific degradation triggered by either deguelin or ECGT treatment.

NUCANT (N6L)

Recently, Destouches and collaborators have identified NPM1 as one of the targets of a promising anticancer compound, NucAnt 6L (N6L) [126]. N6L is a synthetic pseudopeptide consisting of a central 3_{10}-helical template formed by a repetition of 6 pseudopeptide residues (Lys-Aib-Gly) in which, each of the six lysines is linked to other three pseudopeptides (Lys [CH_2N] Pro-Arg) (Figure 2H). This molecule has been shown to inhibit the formation of colonies of various cancer cell lines in soft agar assays with IC_{50} values ranging from 5 to 40 μM, and also to significantly inhibit invasion capacity in a metastatic melanoma cell model (Table 1) [127]. *In vivo*, upon administration, N6L rapidly localizes to tumour tissue in tumour-bearing mice inducing significant inhibition of tumour growth without evident toxicity [126].

The anticancer activity was initially ascribed to the interaction of N6L with nucleolin, however, further investigation also identified NPM1 as a N6L interaction partner. The interaction between NPM1 and N6L was deemed to be monophasic with a dissociation constant of 1 nM as inferred by Surface Plasmon Resonance (SPR) experiments with purified NPM1 [126]. *In vitro* experiments were conducted only using full-length NPM1 and therefore it is presently unknown which NPM1 domain is actually targeted by N6L. Since N6L is rich in lysine and arginine residues, it resembles some properties of positively charged peptides that also bind NPM1, such as those belonging to the HIV Rev and Tat proteins. Therefore N6L, through NPM1 binding, may interfere with some protein-protein associations played by the protein (Figure 3). Further studies are needed to dissect this interaction and to investigate the effect of the drug on cell lines carrying NPM1c+ mutation.

YTR107

Recent evidences have shown that NPM1, when phosphorylated on threonine 199 (pT199NPM1), is a key component of the DNA double strand break (DSB) repair machinery [128]. In particular, in response to the formation of DNA DSBs, pT199NPM1 is recruited to

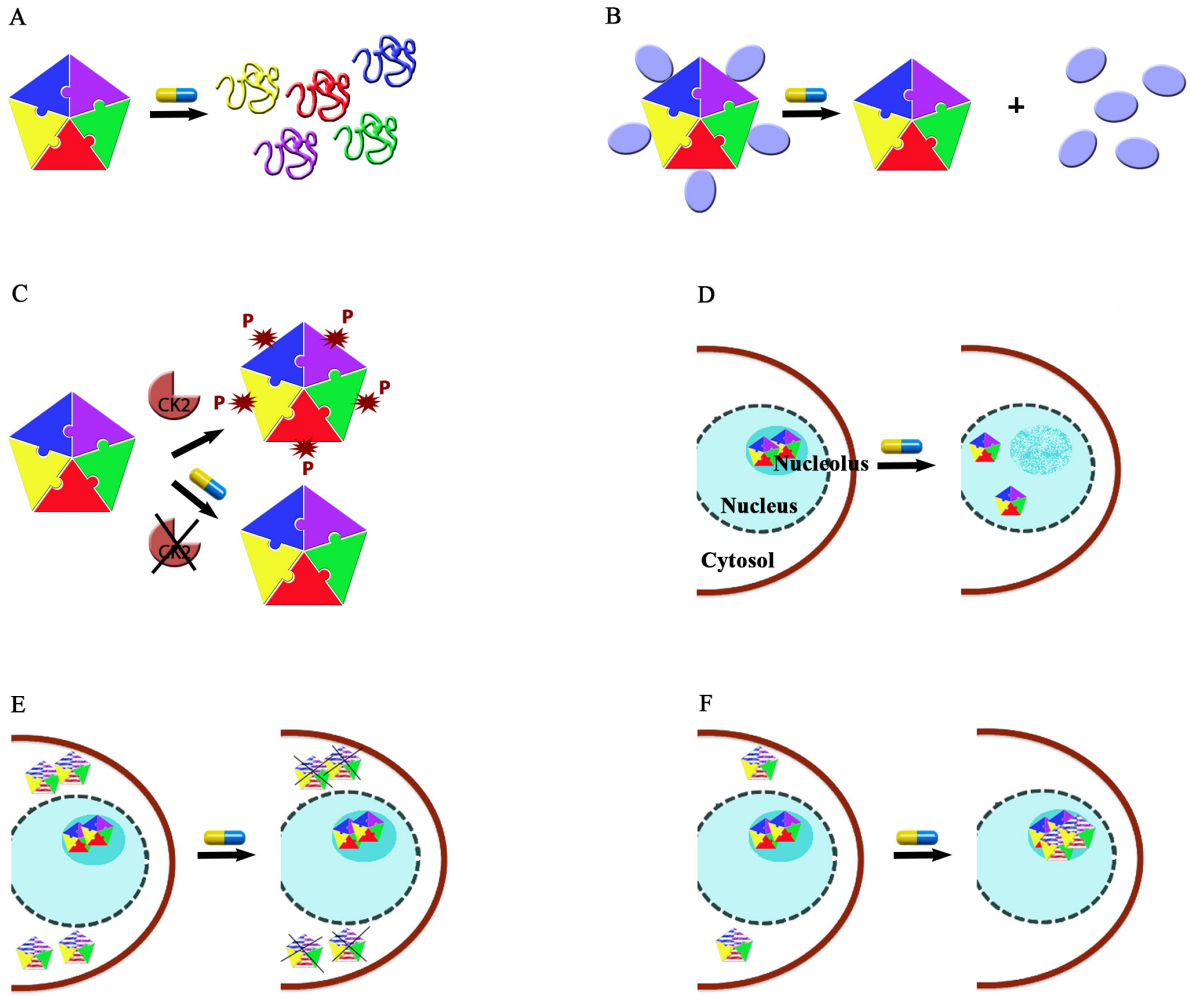

Figure 4: Strategies for NPM1 targeting. A. Interfering with monomer-monomer interactions at the N-terminal domain would cause domain unfolding and the impairment of its functions. **B.** Interfering with the protein-protein interaction surface at the N-terminal domain might prevent most of the NPM1 anti-apoptotic activities. **C.** Interfering with NPM1 post-translational modifications such as CK2-mediated phosphorilation. **D.** Interfering with NPM1 C-terminal domain interactions with nucleic acids would result in nucleolar stress due to the dissociation of the protein from nucleoli (nucleolar starvation hypothesis). **E.** Selective degradation of NPM1c+ while leaving wild-type NPM1 unaffected may be pursued in AML with *NPM1* mutation. **F.** Pharmacological chaperone strategy aimed at refolding the mutated C-terminal domain in NPM1c+ would relocate the protein in nucleoli thus counteracting NPM1c+ anti-apoptotic activities. In panels E and F, mutated NPM1 is distinguished from wild-type NPM1 because represented with horizontal white and coloured lines.

the site of damage and binds ubiquitinated chromatin in a RNF8/RNF168 dependent manner, forming irradiation-induced foci (IRIF) that promote the repair of DNA DSBs [128]. The use of radiation therapy in cancer treatment is limited by the intrinsic resistance acquired by cancer cells through the increased efficacy of their DNA damage repair processes, thus the inhibition of DNA repair mechanisms in cancer cells exposed to ionizing radiation may represent a valid therapeutic approach and, in this context, NPM1 is a new promising target.

Sekhar and colleagues [129], through affinity-based solid-phase resin capture and liquid chromatography/tandem mass spectrometry (LC/MS-MS), have identified NPM1 as the biological target of YTR107, a potent radiosensitizing compound previously identified [130]. YTR107 ((Z) -5 - ((N-benzyl-1H-indol-3-yl) methylene) pyrimidine-2, 4, 6 (1H, 3H, 5H) trione) (Figure 2I) interferes with DNA damage repair mechanisms and is therefore capable of sensitizing to radiation different tumour cell lines, including HT29 colorectal adenocarcinoma cells, D54 glioblastoma cells, PANC1 pancreatic cancer cells, different breast cancer cell models and NSCLC cell lines (Table 1) [127]. Moreover, YTR107 significantly potentiated radiation-induced growth delay in HT29 tumour xenografts [129].

Evidence suggesting that the radiosensitization induced by YTR107 is mediated by NPM1 was also reported [131]. First it was shown that NPM1-null mouse embryonic fibroblasts (MEFs) but not NPM1-wt MEFs are deficient in DNA repair and are radiosensitive. Then, it was shown that treatment with YTR107 of NPM1wt MEFs, but not of NPM1-null MEFs, impaired the formation of pNPM1 irradiation-induced foci and triggered a significant dose-dependent radiosensitization [131]. YTR107 was also shown to bind to the N-terminal region of NPM1 (residues 1-122) responsible for protein oligomerization and to promote NPM1 monomerization (Figure 3) [131]. Very recently, the synthesis of YTR107 analogues with increased efficacy on several cell lines, including OCI-AML3, has also been reported [132].

Strategies in NPM1 targeting

The molecules that we have described above all bind NPM1 or interfere with specific functions played by the protein and, collectively, exemplify six different strategies that can be adopted in NPM1 targeting, summarized in Figure 4.

A first strategy consists in interfering with NPM1 oligomerization (Figure 4A). This effect is best achieved through molecules that target the NPM1 N-terminal domain dimerization surface. This is for instance the case of NSC348884 and, possibly, of YTR107, which were both shown to promote monomer formation, destabilizing the pentameric ring. Such molecules were tested against a significant panel of cancer cell lines where NPM1 is overexpressed and showed significant activity, especially in combination with other drugs (NSC34884) or radiation (YTR107). Not surprisingly, given the high hydrophobicity of the NPM1 dimerization surface, both drugs are poorly water-soluble (Figure 2A and 2I). This may constitute an obstacle for all drugs targeting this NPM1 surface that may be possibly overcome by the development of appropriate delivery systems such as nanocarriers.

A second strategy consists in interfering with the NPM1 protein-protein interaction network (Figure 4B). Data available in the literature [28, 90] and unpublished data from our laboratory, all suggest that protein epitopes enriched in positively charged residues are specifically recognized by the NPM1 N-terminal domain. Importantly, it is highly probable that NPM1 recognizes its protein partners' epitopes with the same surface. Therefore, targeting this surface would hamper several anti-apoptotic functions played by the protein simultaneously. This strategy is so far exemplified by the Rev-NLS peptide, which was effective against the Ras-3T3 cell line, also engrafted in mice, through a p53-dependent apoptotic response. Moreover, the pseudopeptide N6L was found to be effective against a large panel of cancer cell lines and structurally resembles NPM1 interacting peptides. Further design or optimization of molecules that display high affinity and specificity for the NPM1 protein interacting surface would be greatly facilitated if we had structures of the complexes between the NPM1 N-terminal domain and different protein partners.

A third strategy consists in interfering with NPM1 post-translational modifications (Figure 4C). This is exemplified by the cyclic peptide CIGB-300 that targets NPM1 Ser125 and prevents its phosphorylation by CK2, causing nucleolar stress and nucleolar breakdown followed by apoptotic cell death. CIGB300 has demonstrated pharmacological activity against a large panel of tumour cell lines where NPM1 is overexpressed and in mice xenografts. Its mechanism of action is unique among NPM1 interacting drugs however, since Ser125 is located just at the end of NPM1 N-terminal domain, it cannot be excluded that its action may also be that of impairing some NPM1 protein-protein association.

A fourth strategy is based on the so-called "nucleolar starvation hypothesis" [7, 72] according to which the selective displacement of NPM1 from nucleoli might cause nucleolar stress followed by apoptotic cell death (Figure 4D). TmPyP4 and Avrainvillamide actions may be both categorized under this strategy. Both compounds target the NPM1 C-terminal domain structure (avrainvillamide) or nucleic acid binding activity (TmPyP4) and are the only two drugs that are meant to interfere with this protein's domain activities. TmPyP4 effectively displaced NPM1 from nucleoli of AML cells and showed toxicity in a cell line where NPM1 is wild-type, due to p53 activation, while was relatively ineffective in a cell line carrying NPM1c+. Avrainvillamide exerted similar toxicity in the two cell

lines but with a remarkable difference as regards to the NPM1 status: wild-type NPM1 was found displaced from nucleoli while NPM1c+ regained nucleolar localization (which let us categorize avrainvillamide also under a different strategy, see below). The NPM1 C-terminal domain surface involved in nucleic acid binding, and thus responsible for nucleolar localization, has been structurally elucidated and this offers a remarkable opportunity for the design of further molecules specifically aimed at interfering with this surface. Interestingly, while the nucleolar starvation strategy was initially proposed for the treatment of AML with NPM1c+, evidences obtained so far suggest that could it be effective also in tumours where wild-type NPM1 is overexpressed.

The strategies described above are relevant for targeting both solid tumours where NPM1 is overexpressed and AML with NPM1c+. The last two strategies instead may be considered specific for the latter.

The fifth strategy consists in the use of drugs causing the selective destruction of the mutated form of NPM1 (NPM1c+) while leaving wild-type NPM1 levels relatively unchanged (Figure 4E). Recent results suggest that the combined treatment with ATRA/ATO or the use of natural compounds like deguelin or EPGT may reach this goal. The underlying molecular mechanisms are totally unknown with respect to deguelin and EPGT. As to ATRA/ATO, it has been suggested that NPM1c+ might be more sensitive than wild-type NPM1 to the oxidative stress caused by ATO. This strategy is definitively worth further investigation especially in view of its high specificity for NPM1c+.

The final sixth strategy is the least investigated so far and consists in the so-called "pharmacological chaperone" approach [7]. In NPM1c+, the C-terminal domain of the protein is largely destabilized or totally unfolded due to the loss of Trp288 and Trp290. As a consequence the protein loses its nucleolar localization. According to this strategy any drug capable of refolding the mutated C-terminal domain should restore NPM1c+ nucleolar localization thus counteracting NPM1c+ cytosolic antiapoptotic activities (Figure 4F). Interestingly, avrainvillamide was shown to act as a surrogate of a pharmacological chaperone, being capable of inducing NPM1c+ nucleolar relocalization, but without refolding the protein's C-terminal domain.

Finally, always with reference to AML with *NPM1* mutations, it is worth mentioning an additional strategy consisting in the use of nucleolar stress inducers that already are in the clinic. For instance, promising results were obtained by treating patients with actinomycin D, a well known RNA polymerase I inhibitor [133].

CONCLUSIONS

The studies we have reviewed here have shown a therapeutic potential for many molecules that interact with NPM1. Moreover, a striking synergy was observed in many cases when NPM1-targeting compounds were administered in combination with different chemotherapeutic agents or radiotherapy. This suggests that interfering with NPM1 status or functions may be a general way to sensitize cancer cells. Even though it cannot be excluded that, at least in some cases, cellular responses could be due not only to a direct effect on NPM1 and its interactors but also to indirect effects like those following DNA damage, it is clear from these studies that NPM1 targeting may be a powerful strategy for treating a number of tumours of diverse histological origin.

This is a field still in its infancy. Very few tests in animal models were performed and only two compounds (CIGB-300 and Nucant N6L) have entered clinical trials (Table 1). Indeed, many of the compounds we have reviewed here have been discovered or recognized for their effect on NPM1 only in the last two or three years. Furthermore, they not always display chemical features suitable for their development as drugs. Therefore, we anticipate that new molecules will be discovered, pursuing any of the different strategies that we delineated above.

One important issue that arises from the analysis of the available molecules is that, in some cases, their influence on the NPM1 status and localization is not immediately understandable on the basis of our current knowledge of NPM1 structural features. For instance, even though it is well established that the C-terminal domain of the protein contains the NoLS and therefore is responsible for NPM1 nucleolar localization, molecules targeting the NPM1 oligomerization surface at the N-terminal domain were equally able to displace the protein from nucleoli. Moreover, molecules targeting the central region caused loss of oligomerization, a property that is currently ascribed to the N-terminal domain. These and other data indicate the need for a better description of NPM1 structure and interactions. In particular the structures of the NPM1 N-terminal and C-terminal domains have been determined in isolation but nothing is known about the central domain and if, how and when the three domains interact with each other and structurally cooperate to enable NPM1 to fulfil its functions. Furthermore, even though NPM1 interacts with a plethora of different proteins, none of these interactions has been structurally characterized. If we aim at successfully targeting NPM1 for cancer treatment we will also need to address these important issues.

ACKNOWLEDGMENTS

We thank Maurizio Brunori for his useful advice in developing research on NPM1.

FUNDING

This work was supported by a grant from Associazione Italiana Ricerca sul Cancro (IG-15197 to

L.F.).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Grisendi S, Mecucci C, Falini B, Pandolfi PP. Nucleophosmin and cancer. Nat Rev Cancer. 2006; 6:493-505.
2. Wang D, Umekawa H, Olson MO. Expression and subcellular locations of two forms of nucleolar protein B23 in rat tissues and cells. Cell Mol Biol Res. 1993; 39: 33-42.
3. Emmott E, Hiscox JA. Nucleolar targeting: the hub of the matter. EMBO Rep. 2009; 10:231-238.
4. Marasco D, Scognamiglio PL. Identification of inhibitors of biological interactions involving intrinsically disordered proteins. Int J Mol Sci. 2015; 16:7394-7412.
5. Colombo E, Alcalay M, Pelicci PG. Nucleophosmin and its complex network: a possible therapeutic target in hematological diseases. Oncogene. 2011; 30:2595-2609.
6. Lindstrom, MS. NPM1/B23: A Multifunctional Chaperone in Ribosome Biogenesis and Chromatin Remodeling. Biochem Res Int. 2011; 2011:195209.
7. Federici L, Falini B. Nucleophosmin mutations in acute myeloid leukemia: a tale of protein unfolding and mislocalization. Protein Sci. 2013; 22:545-56.
8. Herrera, JE, Savkur R, Olson MO. The ribonuclease activity of nucleolar protein B23. Nucleic Acids Res. 1995; 23:3974-3979.
9. Murano K, Okuwaki M, Hisaoka M, Nagata K. Transcription regulation of the rRNA gene by a multifunctional nucleolar protein, B23/nucleophosmin, through its histone chaperone activity. Mol Cell Biol. 2008; 28:3114-3126.
10. Savkur RS, Olson MO. Preferential cleavage in pre-ribosomal RNA by protein B23 endoribonuclease. Nucleic Acids Res. 1998; 26:4508-4515.
11. Maggi LB Jr, Kuchenruether M, Dadey DY, Schwope RM, Grisendi S, Townsend RR, Pandolfi PP, Weber JD. Nucleophosmin serves as a rate-limiting nuclear export chaperone for the mammalian ribosome. Mol Cell Biol. 2008; 28:7050-7065.
12. Okuda M. The role of nucleophosmin in centrosome duplication. Oncogene. 2002; 21:6170-6174.
13. Wang W, Budhu A, Forgues M, Wang XW. Temporal and spatial control of nucleophosmin by the Ran-Crm1 complex in centrosome duplication. Nat Cell Biol. 2005; 7:823-830.
14. Fantini D, Vascotto C, Marasco D, D'Ambrosio C, Romanello M, Vitagliano L, Pedone C, Poletto M, Cesaratto L, Quadrifoglio F, Scaloni A, Radicella JP, Tell G. Critical lysine residues within the overlooked N-terminal domain of human APE1 regulate its biological functions. Nucleic Acids Res. 2010; 38:8239-8256.
15. Ziv O, Zeisel A, Mirlas-Neisberg N, Swain U, Nevo R, Ben-Chetrit N, Martelli MP4, Rossi R, Schiesser S, Canman CE, Carell T, Geacintov NE, Falini B, Domany E, Livneh Z. Identification of novel DNA-damage tolerance genes reveals regulation of translesion DNA synthesis by nucleophosmin. Nat Commun. 2014; 5:5437.
16. Okuwaki M, Matsumoto K, Tsujimoto M, Nagata K. Function of nucleophosmin/B23, a nucleolar acidic protein, as a histone chaperone. FEBS Lett. 2001; 506:272-276.
17. Szebeni A, Olson MO. Nucleolar protein B23 has molecular chaperone activities. Protein Sci. 1999; 8:905-912.
18. Eitoku M, Sato L, Senda T, Horikoshi M. Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. Cell Mol Life Sci. 2008; 65:414-444.
19. Colombo E, Marine JC, Danovi D, Falini B, Pelicci PG. Nucleophosmin regulates the stability and transcriptional activity of p53. Nat Cell Biol. 2002; 4:529-533.
20. Itahana K, Bhat KP, Jin A, Itahana Y, Hawke D, Kobayashi R, Zhang Y. Tumour suppressor ARF degrades B23, a nucleolar protein involved in ribosome biogenesis and cell proliferation. Molecular Cell. 2003; 12:1151-1164.
21. Kurki S, Peltonen K, Latonen L, Kiviharju TM, Ojala PM, Meek D, Laiho M. Nucleolar protein NPM interacts with HDM2 and protects tumour suppressor protein p53 from HDM2-mediated degradation. Cancer Cell. 2004; 5:465-475.
22. Bertwistle D, Sugimoto M, Sherr CJ. Physical and functional interactions of the Arf tumour suppressor protein with nucleophosmin/B23. Mol Cell Biol. 2004; 24:985-996.
23. Kuo ML, den Besten W, Bertwistle D, Roussel MF, Sherr CJ. N-terminal polyubiquitination and degradation of the Arf tumour suppressor. Genes Dev. 2004; 18: 1862-1874.
24. Li Z, Boone D, Hann SR. Nucleophosmin interacts directly with c-Myc and controls c-Myc-induced hyperproliferation and transformation. Proc Natl Acad Sci U S A. 2008; 105:18794-18799.
25. Li Z, Hann SR. Nucleophosmin is essential for c-Myc nucleolar localization and c-Myc-mediated rDNA transcription. Oncogene. 2013; 32:1988-1994.
26. Szebeni A, Olson MO. Nucleolar protein B23 has molecular chaperone activities. Protein Sci. 1999; 8:905-912.
27. Hingorani K, Szebeni A, Olson MO. Mapping the functional domains of nucleolar protein B23. J Biol Chem 2000; 275:24451-24457
28. Mitrea DM, Grace CR, Buljan M, Yun MK, Pytel NJ, Satumba J, Nourse A, Park CG, Madan Babu M, White SW, Kriwacki RW. Structural polymorphism in the N-terminal oligomerization domain of NPM1. Proc Natl Acad Sci U S A. 2014; 111:4466-4471.

29. Federici L, Arcovito A, Scaglione GL, Scaloni F, Lo Sterzo C, Di Matteo A, Falini B, Giardina B, Brunori M. Nucleophosmin C-terminal leukemia-associated domain interacts with G-rich quadruplex forming DNA. J Biol Chem. 2010; 285:37138-37149.

30. 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, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. N Engl J Med. 2005; 352:254-266.

31. Eirín-López JM, Frehlick LJ, Ausió J. Long-term evolution and functional diversification in the members of the nucleophosmin/nucleoplasmin family of nuclear chaperones. Genetics. 2006; 173:1835-1850

32. Lee HH, Kim HS, Kang JY, Lee BI, Ha JY, Yoon HJ, Lim SO, Jung G, Suh SW. Crystal structure of human nucleophosmin-core reveals plasticity of the pentamer-pentamer interface. Proteins. 2007; 69:672-678.

33. Namboodiri VM, Dutta S, Akey IV, Head JF, Akey CW. The crystal structure of Drosophila NLP-core provides insight into pentamer formation and histone binding. Structure. 2003; 11:175-186.

34. Namboodiri VM, Akey IV, Schmidt-Zachmann MS, Head JF, Akey CW. The structure and function of Xenopus NO38-core, a histone chaperone in the nucleolus. Structure. 2004; 12:2149-2160.

35. Negi SS, Olson MO (2006) Effects of interphase and mitotic phosphorylation on the mobility and location of nucleolar protein B23. J Cell Sci. 2006; 119:3676-3685.

36. Wang W, Budhu A, Forgues M, Wang XW. Temporal and spatial control of nucleophosmin by the Ran-Crm1 complex in centrosome duplication. Nat Cell Biol. 2005; 7:823-830.

37. Marasco D, Ruggiero A, Vascotto C, Poletto M, Scognamiglio PL, Tell G, Vitagliano L. Role of mutual interactions in the chemical and thermal stability of nucleophosmin NPM1 domains. Biochem Biophys Res Commun. 2013; 430:523-528.

38. Herrera JE, Correia JJ, Jones AE, Olson MO. Sedimentation analyses of the salt- and divalent metal ion-induced oligomerization of nucleolar protein B23. Biochemistry. 1996;35:2668-2673.

39. Grummitt CG, Townsley FM, Johnson CM, Warren AJ, Bycroft M. Structural consequences of nucleophosmin mutations in acute myeloid leukemia. J Biol Chem. 2008; 283:23326-23332.

40. Falini B, Bolli N, Shan J, Martelli MP, Liso A, Pucciarini A, Bigerna B, Pasqualucci L, Mannucci R, Rosati R, Gorello P, Diverio D, Roti G, et al. Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML. Blood. 2006; 107:4514-4523.

41. Wang D, Baumann A, Szebeni A, Olson MO. The nucleic acid binding activity of nucleolar protein B23.1 resides in its carboxyl-terminal end. J Biol Chem. 1994; 269:30994-30998.

42. Xu Y, Fang F, Dhar SK, St Clair WH, Kasarskis EJ, St Clair DK. The role of a single-stranded nucleotide loop in transcriptional regulation of the human sod2 gene. J Biol Chem. 2007; 282:15981-15994.

43. Gallo A, Lo Sterzo C, Mori M, Di Matteo A, Bertini I, Banci L, Brunori M, Federici L. Structure of nucleophosmin DNA-binding domain and analysis of its complex with a G-quadruplex sequence from the c-MYC promoter. J Biol Chem. 2012; 287:26539-26548.

44. Arcovito A, Chiarella S, Della Longa S, Di Matteo A, Lo Sterzo C, Scaglione GL, Federici L. Synergic role of nucleophosmin three-helix bundle and a flanking unstructured tail in the interaction with G-quadruplex DNA. J Biol Chem. 2014; 289:21230-21241.

45. Scognamiglio PL, Di Natale C, Leone M, Poletto M, Vitagliano L, Tell G, Marasco D. G-quadruplex DNA recognition by nucleophosmin: new insights from protein dissection. Biochim Biophys Acta. 2014; 1840:2050-2059.

46. Shandilya J, Swaminathan V, Gadad SS, Choudhari R, Kodaganur GS, Kundu TK. Acetylated NPM1 localizes in the nucleoplasm and regulates transcriptional activation of genes implicated in oral cancer manifestation. Mol Cell Biol. 2009; 29:5115-5127.

47. Léotoing L, Meunier L, Manin M, Mauduit C, Decaussin M, Verrijdt G, Claessens F, Benahmed M, Veyssière G, Morel L, Beaudoin C. Influence of nucleophosmin/B23 on DNA binding and transcriptional activity of the androgen receptor in prostate cancer cell. Oncogene. 2008; 27:2858-2867.

48. Yun JP, Miao J, Chen GG, Tian QH, Zhang CQ, Xiang J, Fu J, Lai PB. Increased expression of nucleophosmin/B23 in hepatocellular carcinoma and correlation with clinicopathological parameters. Br J Cancer. 2007; 96:477-484.

49. Pianta A, Puppin C, Franzoni A, Fabbro D, Di Loreto C, Bulotta S, Deganuto M, Paron I, Tell G, Puxeddu E, Filetti S, Russo D, Damante G. Nucleophosmin is overexpressed in thyroid tumours. Biochem Biophys Res Commun. 2010; 397:499-504.

50. Nozawa Y, Van Belzen N, Van der Made AC, Dinjens WN, Bosman FT. Expression of nucleophosmin/B23 in normal and neoplastic colorectal mucosa. J Pathol. 1996; 178:48-52.

51. Tanaka M, Sasaki H, Kino I, Sugimura T, Terada M. Genes preferentially expressed in embryo stomach are predominantly expressed in gastric cancer. Cancer Res. 1992; 52:3372-3377.

52. Zhu Y, Shi M, Chen H, Gu J, Zhang J, Shen B, Deng X, Xie J, Zhan X, Peng C. NPM1 activates metabolic changes by inhibiting FBP1 while promoting the tumourigenicity of pancreatic cancer cells. Oncotarget. 2015; 6:21443-21451. doi: 10.18632/oncotarget.4167.

53. Chen J, Sun J, Yang L, Yan Y, Shi W, Shi J, Huang Q, Chen J, Lan Q. Upregulation of B23 promotes tumour cell proliferation and predicts poor prognosis in glioma. Biochem Biophys Res Commun. 2015; 466:124-30.
54. Holmberg Olausson K, Elsir T, Moazemi Goudarzi K, Nistér M, Lindström MS. NPM1 histone chaperone is upregulated in glioblastoma to promote cell survival and maintain nucleolar shape. Sci Rep. 2015; 5:16495.
55. Kuo YH, Chen YT, Tsai HP, Chai CY, Kwan AL. (2015) Nucleophosmin overexpression is associated with poor survival in astrocytoma. APMIS. 2015; 123:515-522.
56. Yung BY. Oncogenic role of nucleophosmin/B23. Chang Gung Med J. 2007; 30:285-293.
57. Zeller KI, Haggerty TJ, Barrett JF, Guo Q, Wonsey DR, Dang CV. Characterization of nucleophosmin (B23) as a Myc target by scanning chromatin immunoprecipitation. J Biol Chem. 2001; 276:48285-48291.
58. Roussel P, Hernandez-Verdun D. Identification of Ag-NOR proteins, markers of proliferation related to ribosomal gene activity. Exp Cell Res. 1994; 214: 465-472.
59. Poletto M, Malfatti MC, Dorjsuren D, Scognamiglio PL, Marasco D, Vascotto C, Jadhav A, Maloney DJ, Wilson DM 3rd, Simeonov A, Tell G. Inhibitors of the apurinic/apyrimidinic endonuclease 1 (APE1)/nucleophosmin (NPM1) interaction that display anti-tumor properties. Mol Carcinog. 2015; doi: 10.1002/mc.22313.
60. Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, Saltman DL, Look AT. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science. 1994; 263:1281-1284.
61. Falini B, Nicoletti I, Bolli N, Martelli MP, Liso A, Gorello P, Mandelli F, Mecucci C, Martelli MF. Translocations and mutations involving the nucleophosmin (NPM1) gene in lymphomas and leukemias. Haematologica. 2007; 92:519-532.
62. Chiarle R, Gong JZ, Guasparri I, Pesci A, Cai J, Liu J et al. NPM-ALK transgenic mice spontaneously develop T-cell lymphomas and plasma cell tumours. Blood. 2003; 101:1919-1927.
63. Redner RL, Chen JD, Rush EA, Li H, Pollock SL. The t(5;17) acute promyelocytic leukemia fusion protein NPM-RAR interacts with co-repressor and co-activator proteins and exhibits both positive and negative transcriptional properties. Blood. 2000; 95:2683-2690.
64. Okazuka K, Masuko M, Seki Y, Hama H, Honma N, Furukawa T, Toba K, Kishi K, Aizawa Y. Successful all-trans retinoic acid treatment of acute promyelocytic leukemia in a patient with NPM/RAR fusion. Int J Hematol. 2007; 86 246-249.
65. Yoneda-Kato N, Look AT, Kirstein MN, Valentine MB, Raimondi SC, Cohen KJ, Carroll AJ, Morris SW. The t(3;5)(q25.1;q34) of myelodysplastic syndrome and acute myeloid leukemia produces a novel fusion gene, NPM-MLF1. Oncogene. 1996; 12:265-275.
66. Berger R, Busson M, Baranger L, Hélias C, Lessard M, Dastugue N, Speleman F. Loss of the NPM1 gene in myeloid disorders with chromosome 5 rearrangements. Leukemia. 2006; 20:319-321.
67. Falini B, Nicoletti I, Martelli MF, Mecucci C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. Blood. 2007; 109:874-85.
68. Meani N, Alcalay M. Role of nucleophosmin in acute myeloid leukemia. Expert Rev Anticancer Ther. 2009; 9:1283-94.
69. Scaloni F, Gianni S, Federici L, Falini B, Brunori M. Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. FASEB J. 2009; 23:2360-2365.
70. Scaloni F, Federici L, Brunori M, Gianni S. Deciphering the folding transition state structure and denatured state properties of Nucleophosmin C-terminal domain. Proc Natl Acad Sci USA. 2010; 107:5447-5452.
71. Falini B, Bolli N, Liso A, Martelli MP, Mannucci R, Pileri S, Nicoletti I. Altered nucleophosmin transport in acute myeloid leukaemia with mutated NPM1: molecular basis and clinical implications. Leukemia. 2009; 23:1731-1743.
72. Falini B, Gionfriddo I, Cecchetti F, Ballanti S, Pettirossi V, Martelli MP. Acute Myeloid Leukemia with mutated nucleophosmin (NPM1): any hope for a targeted therapy? Blood Rev. 2011; 25:247-254.
73. Sportoletti P, Varasano E, Rossi R, Mupo A, Tiacci E, Vassiliou G, Martelli MP, Falini B. Mouse models of NPM1-mutated acute myeloid leukemia: biological and clinical implications. Leukemia. 2015; 29:269-278.
74. Arber D, Brunning RD, Le Beau MM, Falini B, Vardiman JW, Porwit A, et al. Acute myeloid leukaemia with recurrent genetic abnormalities. In: Swerdlow SE, Campo E, Harris NL, et al, editors. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon, France: International Agency for Research on Cancer (IARC). 2008; 110-123.
75. Ofran Y, Rowe JM. Introducing minimal residual disease in AML. Curr Opin Hematol. 2015; 22:139-145.
76. Di Fiore PP. Playing both sides: nucleophosmin between tumour suppression and oncogenesis. J Cell Biol. 2008;182:7-9.
77. Bonetto P, Davoli T, Sironi C, Amati B, Pelicci PG, Colombo E. Nucleophosmin and its AML-associated mutant regulate c-Myc turnover through Fbw7 gamma. J Cell Biol. 2008; 182:19-26.
78. Colombo E, Martinelli P, Zamponi R, Shing DC, Bonetti P, Luzi L, Volorio S, Bernard L, Pruneri G, Alcalay M, Pelicci PG. Delocalization and destabilization of the Arf tumour suppressor by the leukemia-associated NPM mutant. Cancer Res. 2006; 66:3044-3050.

79. Leong SM, Tan BX, Bte Ahmad B, Yan T, Chee LY, Ang ST, Tay KG, Koh LP, Yeoh AE, Koay ES, Mok YK, Lim TM. Mutant nucleophosmin deregulates cell death and myeloid differentiation through excessive caspase-6 and -8 inhibition. Blood. 2010; 116:3286-96.

80. Noguera NI, Song MS, Divona M, Catalano G, Calvo KL, García F, Ottone T, Florenzano F, Faraoni I, Battistini L, Colombo E, Amadori S, Pandolfi PP, Lo-Coco F. Nucleophosmin/B26 regulates PTEN through interaction with HAUSP in acute myeloid leukemia. Leukemia. 2013; 27:1037-43.

81. Di Natale C, Scognamiglio PL, Cascella R, Cecchi C, Russo A, Leone M, Penco A, Relini A, Federici L, Di Matteo A, Chiti F, Vitagliano L, Marasco D. Nucleophosmin contains amyloidogenic regions that are able to form toxic aggregates under physiological conditions. FASEB J. 2015; 29:3689-3701.

82. Qi W, Shakalya K, Stejskal A, Goldman A, Beeck S, Cooke L, Mahadevan D. NSC348884, a nucleophosmin inhibitor disrupts oligomer formation and induces apoptosis in human cancer cells. Oncogene. 2008; 27:4210-4220.

83. Maiguel DA, Jones L, Chakravarty D, Yang C, Carrier F. Nucleophosmin sets a threshold for p53 response to UV radiation. Mol Cell Biol. 2004; 24:3703-3711.

84. Lindström MS, Zhang Y. B23 and ARF: friends or foes? Cell Biochem Biophys. 2006; 46:79-90.

85. Zhang J, Zhao HL, He JF, Li HY. Inhibitory effect of NSC348884, a small molecular inhibitor of nucleophosmin, on the growth of hepatocellular carcinoma cell line hepG2. [Article in Chinese]. Zhongguo Yi Xue Ke Xue Yuan Xue Bao. 2012; 34:58-61.

86. Balusu R, Fiskus W, Rao R, Chong DG, Nalluri S, Mudunuru U, Ma H, Chen L, Venkannagari S, Ha K, Abhyankar S, Williams C, McGuirk J, Khoury HJ, Ustun C, Bhalla KN. Targeting levels or oligomerization of nucleophosmin 1 induces differentiation and loss of survival of human AML cells with mutant NPM1. Blood. 2011; 118:3096-3106.

87. Emmott E, Hiscox JA. Nucleolar targeting: the hub of the matter. EMBO Rep. 2009; 10:231-8.

88. Fankhauser C, Izaurralde E, Adachi Y, Wingfield P, Laemmli UK. Specific complex of human immunodeficiency virus type 1 rev and nucleolar B23 proteins: dissociation by the Rev response element. Mol. Cell. Biol. 1991; 11:2567-2575.

89. Umekawa H, Chang JH, Correia JJ, Wang D, Wingfield PT, Olson MO. Nucleolar protein B23: bacterial expression, purification, oligomerization and secondary structures of two isoforms. Cell Mol Biol Res. 1993; 39:635-45.

90. Szebeni A, Herrera JE, Olson MO. Interaction of nucleolar protein B23 with peptides related to nuclear localization signals. Biochemistry. 1995; 34:8037-8042.

91. Chan HJ, Weng JJ, Yung BY. Nucleophosmin/B23-binding peptide inhibits tumour growth and up-regulates transcriptional activity of p53. Biochem Biophys Res Commun. 2005; 333:396-403.

92. Keefe AD, Pai S, Ellington A. Aptamers as therapeutics. Nat Rev Drug Discov. 2010; 9:537-550.

93. Nimjee SM, Rusconi CP, Sullenger BA. APTAMERS: an emerging class of therapeutics. Annu Rev Med. 2005; 56: 555-583.

94. Jian Y, Gao Z, Sun J, Shen Q, Feng F, Jing Y, Yang C. RNA aptamers interfering with nucleophosmin oligomerization induce apoptosis of cancer cells. Oncogene. 2009; 28:4201-4211.

95. Trembley JH, Wang G, Unger G, Slaton J, Ahmed K. Protein kinase CK2 in health and disease: CK2: a key player in cancer biology. Cell Mol Life Sci. 2009; 66:1858-1867.

96. Ruzzene M, Pinna LA. Addiction to protein kinase CK2: a common denominator of diverse cancer cells? Biochim Biophys Acta. 2010;1804:499-504.

97. Ahmad KA, Wang G, Unger G, Slaton J, Ahmed K. Protein kinase CK2—a key suppressor of apoptosis. Adv Enzyme Regul. 2008; 48:179-187.

98. Perea SE, Reyes O, Baladron I, Perera Y, Farina H, Gil J, Rodriguez A, Bacardi D, Marcelo JL, Cosme K, Cruz M, Valenzuela C, López-Saura PA, et al. CIGB-300, a novel proapoptotic peptide that impairs the CK2 phosphorylation and exhibits anticancer properties both in vitro and in vivo. Mol Cell Biochem. 2008; 316:163-167.

99. Perera Y, Farina HG, Gil J, Rodriguez A, Benavent F, Castellanos L, Gómez RE, Acevedo BE, Alonso DF, Perea SE. Anticancer peptide CIGB-300 binds to nucleophosmin/B23, impairs its CK2-mediated phosphorylation, and leads to apoptosis through its nucleolar disassembly activity. Mol Cancer Ther. 2009; 8:1189-1196.

100. Louvet E, Junéra HR, Le Panse S, Hernandez-Verdun D. Dynamics and compartmentation of the nucleolar processing machinery. Exp Cell Res. 2005; 304:457-470.

101. Louvet E, Junéra HR, Berthuy I, Hernandez-Verdun D. Compartmentation of the nucleolar processing proteins in the granular component is a CK2-driven process. Mol Biol Cell. 2006; 17:2537-2546.

102. Szebeni A, Hingorani K, Negi S, Olson MO. Role of protein kinase CK2 phosphorylation in the molecular chaperone activity of nucleolar protein b23. J Biol Chem. 2003; 278:9107-9115.

103. Perera Y, Costales HC, Diaz Y, Reyes O, Farina HG, Mendez L, Gómez RE, AcevedoBE, Gomez DE, Alonso DF, Perea SE. Sensitivity of tumour cells towards CIGB-300 anticancer peptide relies on its nucleolar localization. J Pept Sci. 2012; 18:215-223.

104. Martins LR, Perera Y, Lúcio P, Silva MG, Perea SE, Barata JT. Targeting chronic lymphocytic leukemia using CIGB-300, a clinical-stage CK2-specific cell-permeable peptide inhibitor. Oncotarget. 2014; 5:258-263. doi: 10.18632/oncotarget.1513.

105. Perera Y, Toro ND, Gorovaya L, Fernandez-DE-Cossio J, Farina HG, Perea SE. Synergistic interactions of the anti-casein kinase 2 CIGB-300 peptide and chemotherapeutic agents in lung and cervical preclinical cancer models. Mol Clin Oncol. 2014; 2:935-944.

106. Solares AM, Santana A, Baladrón I, Valenzuela C, González CA, Díaz A, Castillo D, Ramos T, Gómez R, Alonso DF, Herrera L, Sigman H, Perea SE, et al. Safety and preliminary efficacy data of a novel casein kinase 2 (CK2) peptide inhibitor administered intralesionally at four dose levels inpatients with cervical malignancies. BMC Cancer. 2009; 9:146.

107. Sarduy MR, García I, Coca MA, Perera A, Torres LA, Valenzuela CM, Baladrón I, Solares M, Reyes V, Hernández I, Perera Y, Martínez YM, Molina L, et al. Optimizing CIGB-300 intralesional delivery in locally advanced cervical cancer. Br J Cancer. 2015; 112:1636-1643.

108. Fenical W, Jensen PR, Cheng XC. Avrainvillamide, a Cytotoxic Marine Natural Product, and Derivatives Thereof. US Patent 6, 066, 635. 2000

109. Wulff JE, Siegrist R, Myers AG. The natural product avrainvillamide binds to the oncoprotein nucleophosmin. J Am Chem Soc. 2007; 129:14444-14451.

110. Mukherjee H, Chan KP, Andresen V, Hanley ML, Gjertsen BT, Myers AG. Interactions of the natural product (+)-avrainvillamide with nucleophosmin and exportin-1 mediate the cellular localization of nucleophosmin and its AML-associated mutants. ACS Chem Biol. 2015; 10:855-863.

111. Freyer MW, Buscaglia R, Kaplan K, Cashman D, Hurley LH, Lewis EA. Biophysical Studies of the c-MYC NHE III1 Promoter: Model Quadruplex Interactions with a Cationic Porphyrin. Biophys J. 2007; 92:2007-2015.

112. Siddiqui-Jain, 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. 2002; 99:11593-11598.

113. Rapozzi V, Zorzet S, Zacchigna M, Della Pietra E, Cogoi S, Xodo LE (2014) Anticancer activity of cationic porphyrins in melanoma tumour-bearing mice and mechanistic in vitro studies. Mol Cancer. 2014;13:75.

114. Zhang YQ, Zhang YH, Xie J, Li MN, Liu ZR, Shen JY, Shi SS, Lan XY, Wang S, Cheng NL. TMPyP4-regulated cell proliferation and apoptosis through the Wnt/-catenin signaling pathway in SW480 cells. J Recept Signal Transduct Res. 2016; 36:167-172.

115. Drygin D1, Siddiqui-Jain A, O'Brien S, Schwaebe M, Lin A, Bliesath J, Ho CB, Proffitt C, Trent K, Whitten JP, Lim JK, Von Hoff D, Anderes K, Rice WG. Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. Cancer Res. 2009; 69:7653-7661.

116. Chiarella S, De Cola A, Scaglione GL, Carletti E, Graziano V, Barcaroli D, Lo Sterzo C, Di Matteo A, Di Ilio C, Falini B, Arcovito A, De Laurenzi V, Federici L. Nucleophosmin mutations alter its nucleolar localization by impairing G-quadruplex binding at ribosomal DNA. Nucleic Acids Res. 2013; 41:3228-3239.

117. De Cola A, Pietrangelo L, Forlì F, Barcaroli D, Budani MC, Graziano V, Protasi F, Di Ilio C, De Laurenzi V, Federici L. AML cells carrying NPM1 mutation are resistant to nucleophosmin displacement from nucleoli caused by the G-quadruplex ligand TmPyP4. Cell Death Dis. 2014; 5:e1427.

118. Lo-Coco F, Avvisati G, Vignetti M, Thiede C, Orlando SM, Iacobelli S, Ferrara F, Fazi P, Cicconi L, Di Bona E, Specchia G, Sica S, Divona M, Levis A, et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. N Engl J Med. 2013; 369:111-121.

119. Sanz MA, Grimwade D, Tallman MS, Lowenberg B, Fenaux P, Estey EH, Naoe T, Lengfelder E, Büchner T, Döhner H, Burnett AK, Lo-Coco F. Management of acute promyelocytic leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. Blood. 2009; 113:1875-1891.

120. Martelli MP, Gionfriddo I, Mezzasoma F, Milano F, Pierangeli S, Mulas F, Pacini R, Tabarrini A, Pettirossi V, Rossi R, Vetro C, Brunetti L, Sportoletti P, Tet al. Arsenic trioxide and all-trans retinoic acid target NPM1 mutant oncoprotein levels and induce apoptosis in NPM1-mutated AML cells. Blood. 2015; 125:3455-3465.

121. El Hajj H, Dassouki Z, Berthier C, Raffoux E, Ades L, Legrand O, Hleihel R, Sahin U, Tawil N, Salameh A, Zibara K, Darwiche N, Mohty M, Dombret H, Fenaux P, de Thé H, Bazarbachi A. Retinoic acid and arsenic trioxide trigger degradation of mutated NPM1, resulting in apoptosis of AML cells. Blood. 2015; 125:3447-3454.

122. Wang Y, Ma W, Zheng W (2013) Deguelin, a novel anti-tumourigenic agent targeting apoptosis, cell cycle arrest and anti-angiogenesis for cancer chemoprevention. Mol Clin Oncol. 2013; 1:215-219.

123. Zhang L, Wei Y, Zhang J. Novel mechanisms of anticancer activities of green tea component epigallocatechin-3-gallate. Anticancer Agents Med Chem. 2014; 14:779-786.

124. Yi S, Wen L, He J, Wang Y, Zhao F, Zhao J, Zhao Z, Cui G, Chen Y. Deguelin, a selective silencer of the NPM1 mutant, potentiates apoptosis and induces differentiation in AML cells carrying the NPM1 mutation. Ann Hematol. 2015; 94:201-210.

125. Chi HT, Ly BT, Vu HA, Sato Y, Dung PC, Xinh PT. Down-regulated expression of NPM1 in IMS-M2 cell line by (-)-epigallocatechin-3-gallate. Asian Pac J Trop Biomed. 2014; 4:570-574.

126. Destouches D, Page N, Hamma-Kourbali Y, Machi V, Chaloin O, Frechault S, Birmpas C, Katsoris P, Beyrath J, Albanese P, Maurer M, Carpentier G, Strub JM, Van Dorsselaer A, Muller S, Bagnard D, Briand JP, Courty J. A simple approach to cancer therapy afforded by multivalent

pseudopeptides that target cell-surface nucleoproteins. Cancer Res. 2011; 71:3296-3305.

127. Destouches D, Huet E, Sader M, Frechault S, Carpentier G, Ayoul F, Briand JP, Menashi S, Courty J. Multivalent pseudopeptides targeting cell surface nucleoproteins inhibit cancer cell invasion through tissue inhibitor of metalloproteinases 3 (TIMP-3) release. J Biol Chem. 2012; 287:43685-43693.

128. Koike A, Nishikawa H, Wu W, Okada Y, Venkitaraman AR, Ohta T. Recruitment of phosphorylated NPM1 to sites of DNA damage through RNF8-dependent ubiquitin conjugates. Cancer Res. 2010; 70:6746-6756.

129. Sekhar KR, Reddy YT, Reddy PN, Crooks PA, Venkateswaran A, McDonald WH, GengL, Sasi S, Van Der Waal RP, Roti JL, Salleng KJ, Rachakonda G, Freeman ML. The novel chemical entity YTR107 inhibits recruitment of nucleophosmin to sites of DNA damage, suppressing repair of DNA double-strand breaks and enhancing radiosensitization. Clin Cancer Res. 2011; 17:6490-6499.

130. Reddy YT, Sekhar KR, Sasi N, Reddy PN, Freeman ML, Crooks PA. Novel substituted (Z)-5-((N-benzyl-1H-indol-3-yl)methylene)imidazolidine-2, 4-diones and5-((N-benzyl-1H-indol-3-yl)methylene)pyrimidine-2, 4, 6(1H, 3H, 5H)-triones as potent radio-sensitizing agents. Bioorg Med Chem Lett. 2010; 20:600-602.

131. Sekhar KR, Benamar M, Venkateswaran A, Sasi S, Penthala NR, Crooks PA, Hann SR, Geng L, Balusu R, Abbas T, Freeman ML. Targeting nucleophosmin 1 represents a rational strategy for radiation sensitization. Int J Radiat Oncol Biol Phys. 2014; 89:1106-1114.

132. Penthala NR, Ketkar A, Sekhar KR, Freeman ML, Eoff RL, Balusu R, Crooks PA. 1-Benzyl-2-methyl-3-indolylmethylene barbituric acid derivatives: Anti-cancer agents that target nucleophosmin 1 (NPM1). Bioorg Med Chem. 2015; 23:7226-7233.

133. Falini B, Brunetti L, Martelli MP. Dactinomycin in NPM1-Mutated Acute Myeloid Leukemia. N Engl J Med. 2015; 373:1180-1182.

OPEN

Citation: Oncogenesis (2017) **6**, e379; doi:10.1038/oncsis.2017.78

www.nature.com/oncsis

ORIGINAL ARTICLE

Structural investigation of nucleophosmin interaction with the tumor suppressor Fbw7γ

A Di Matteo[1,9], M Franceschini[2,3,9], A Paiardini[4], A Grottesi[5], S Chiarella[2,3], S Rocchio[6], C Di Natale[7], D Marasco[7], L Vitagliano[8], C Travaglini-Allocatelli[6] and L Federici[2,3]

Nucleophosmin (NPM1) is a multifunctional nucleolar protein implicated in ribogenesis, centrosome duplication, cell cycle control, regulation of DNA repair and apoptotic response to stress stimuli. The majority of these functions are played through the interactions with a variety of protein partners. NPM1 is frequently overexpressed in solid tumors of different histological origin. Furthermore NPM1 is the most frequently mutated protein in acute myeloid leukemia (AML) patients. Mutations map to the C-terminal domain and lead to the aberrant and stable localization of the protein in the cytoplasm of leukemic blasts. Among NPM1 protein partners, a pivotal role is played by the tumor suppressor Fbw7γ, an E3-ubiquitin ligase that degrades oncoproteins like c-MYC, cyclin E, Notch and c-jun. In AML with NPM1 mutations, Fbw7γ is degraded following its abnormal cytosolic delocalization by mutated NPM1. This mechanism also applies to other tumor suppressors and it has been suggested that it may play a key role in leukemogenesis. Here we analyse the interaction between NPM1 and Fbw7γ, by identifying the protein surfaces implicated in recognition and key aminoacids involved. Based on the results of computational methods, we propose a structural model for the interaction, which is substantiated by experimental findings on several site-directed mutants. We also extend the analysis to two other NPM1 partners (HIV Tat and CENP-W) and conclude that NPM1 uses the same molecular surface as a platform for recognizing different protein partners. We suggest that this region of NPM1 may be targeted for cancer treatment.

Oncogenesis (2017) **6,** e379; doi:10.1038/oncsis.2017.78; published online 18 September 2017

INTRODUCTION

Nucleophosmin (NPM1) is an abundant and ubiquitous protein[1] mainly localized in nucleoli, where it contributes to their structure and organization,[2,3] but also shuttles between nucleolus and cytoplasm to perform its functions.[4–6] NPM1 has a primary role in ribosome biogenesis and transport[7,8] but also contributes to the maintenance of genomic stability and DNA repair,[9,10] histones assembly,[11,12] centrosome duplication,[13,14] cell cycle regulation and response to stress stimuli.[5] The pleiotropic behavior of NPM1 is due to its modular structure consisting of: (i) an N-terminal oligomerization domain involved in protein–protein interactions and containing two nuclear export signals (NES);[1,4] (ii) an intrinsically unstructured central region which contains a bipartite nuclear localization signal (NLS) and (iii) a C-terminal nucleic acid binding domain where the nucleolar localization signal (NoLS) is located.[6] Multiple post-translational modifications such as phosphorylation, acetylation and glutathionylation regulate NPM1 localization and activities.[4,5,15]

NPM1 is overexpressed in several tumors, including prostate, liver, gastric, colon, pancreas, glioma and glioblastoma, astrocytoma and others.[16] Its overexpression often correlates with mitotic index and metastatization and it was proposed as an adverse prognostic marker.[17,18] The *NPM1* gene is also frequently altered in hematological malignancies arising from chromosomal translocations. Here, the N-terminal domain of NPM1 is fused to protein partners such as ALK, RARα and MLF1, giving rise to oncogenic proteins and haploinsufficency for wild-type NPM1.[1] Finally, the *NPM1* gene is the most frequently mutated in acute myeloid leukemia (AML), accounting for 30% of patients.[19] AML mutations are localized at the C-terminal domain of the protein and cause: (i) the loss of the NoLS, (ii) a severe destabilization or the complete unfolding of the domain and (iii) the appearance of a new NES.[19] As a result, mutated NPM1 loses affinity for nucleoli and is found stably and aberrantly in the cytoplasm.[19]

NPM1 interacts with several protein partners, modulating their stability and, importantly, it seems to have a fundamental role in their nucleolar localization. Indeed, most of NPM1 interacting proteins contain multivalent arginine-rich motifs[3] generally found in NoLS.[20] Furthermore, the reduction of NPM1 levels is associated with the alteration of nucleolar structure.[18] All these aspects substantiate the hypothesis that NPM1 behaves as a hub protein in nucleoli.[2,3] Relevant examples of NPM1 partners include ribosomal proteins (RPL5, RPS9, RPL23), viral proteins (Rev, Tat) and many tumor suppressors, including p14ARF, p53 and Fbw7γ.[4,16]

NPM1 is required for the nucleolar localization and stabilization of the isoform γ of Fbw7.[21] Fbw7 belongs to the SCF (Skp1, Cullin-1, Fbox protein) class of E3-ubiquitin ligases[22] and has a modular organization comprising: (i) the D dimerization domain, (ii) the Fbox domain that binds Skp1 of the SCF complex and (iii) the

[1]Istituto di Biologia e Patologia Molecolari – Consiglio Nazionale delle ricerche, Roma, Italy; [2]Dipartimento di Scienze Mediche, Orali e Biotecnologiche, Chieti, Italy; [3]CeSI-Met – Università di Chieti-Pescara 'G d'Annunzio', Chieti, Italy; [4]Dipartimento di Biologia e Biotecnologie 'C Darwin' – Sapienza Università di Roma, Roma, Italy; [5]CINECA Consorzio Interuniversitario, Sede di Roma, Roma, Italy; [6]Dipartimento di Scienze Biochimiche 'A Rossi Fanelli' - Sapienza Università di Roma, Roma, Italy; [7]Dipartimento di Farmacia,– Università di Napoli 'Federico II', Napoli, Italy and [8]Istituto di Biostrutture e Bioimmagini – Consiglio Nazionale delle Ricerche, Napoli, Italy. Correspondence: Professor L Federici, Dipartimento di Scienze Mediche, Orali e Biotecnologiche, Università di Chieti-Pescara 'G d'Annunzio', Via dei Vestini 31, 66100 Chieti, Italy.
E-mail: lfederici@unich.it or luca.federici@unich.it
[9]These authors contributed equally to this work.
Received 10 February 2017; revised 17 July 2017; accepted 19 July 2017

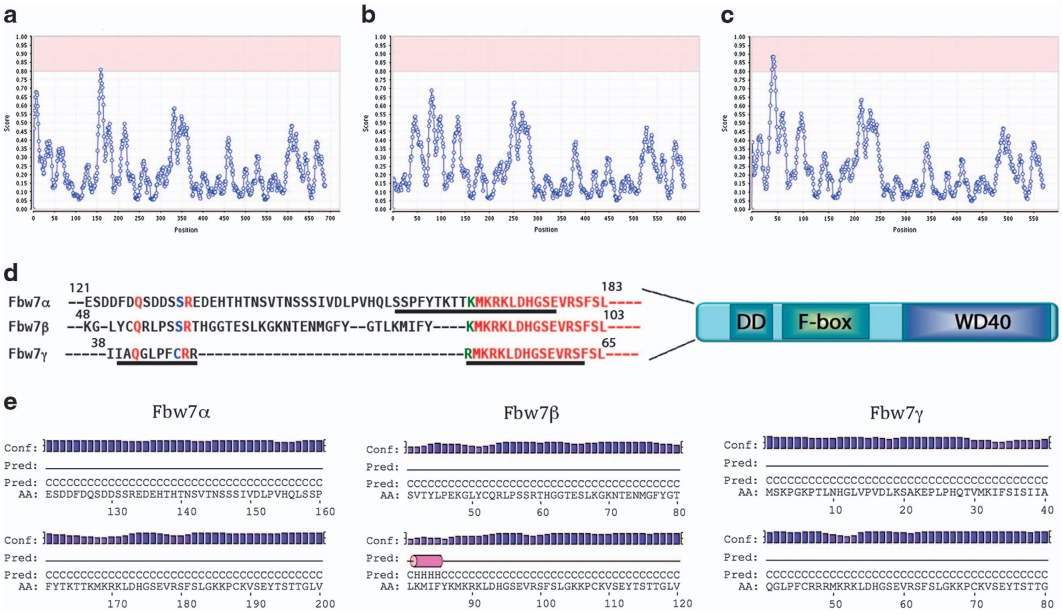

Figure 1. Identification of nucleolar localization signal (NoLS) in Fbw7. Fbw7 isoforms vary in their N-terminal sequence. The sequences of Fbw7α (**a**), Fbw7β (**b**) and Fbw7γ (**c**) were subjected to the NoD algorithm in order to identify putative NoLS (Scott *et al.*,[20]). The server identifies a full NoLS in Fbw7γ only (score above 0.8) while only a partial one in Fbw7α. Fbw7γ is known to be nucleolar while Fbw7α is located in the nucleoplasm. In **d** the underlined sequences correspond to the putative NoLS. PSIPRED secondary structure predictions for the three isoforms, in the regions of interest, are shown in **e**.

WD40 domain, which recognizes phosphorylated substrates.[22] The *Fbw7* gene codes for three protein isoforms (namely α, β and γ) differing in their N-terminal region and displaying distinct cellular localization: Fbw7α is nucleoplasmic, Fbw7β is cytoplasmic and Fbw7γ is nucleolar.[23] Many of Fbw7 targets are oncoproteins, including c-MYC, Notch, Cyclin E and c-Jun[22] and therefore isoforms localization may be instrumental in their regulation through the compartmentalization of substrates recognition and degradation. For instance, nucleolar c-Myc is specifically ubiquitinylated by Fbw7γ, thus regulating its growth promoting activity.[23]

The alteration of the NPM1/Fbw7γ/c-Myc circuitry was reported in AML with *NPM1* mutations.[21] First, it was shown that NPM1 is necessary for Fbw7γ nucleolar localization and stabilization. As a consequence, c-Myc ubiquitination and proteasome degradation is enhanced, thus lowering its levels. Conversely, c-Myc is stabilized in cells lacking NPM1 and, importantly, in AML blasts bearing the mutated form of NPM1.[21] Indeed, mutations causing NPM1 cytoplasmic delocalization, do not compromise the interaction of NPM1 with Fbw7γ which is also delocalized in the cytoplasm and degraded.[21] A similar delocalization/degradation mechanism was observed with the tumor suppressor p14ARF.[24,25] Overall, different apoptotic responses are compromised by the selective cytosolic degradation of NPM1 partners. These and other observations led to the suggestion that the NPM1 region/s implicated in protein partners recognition may be considered a target for cancer treatment.[16]

In this paper we investigate the interaction between NPM1 and Fbw7γ. We identify, in both proteins, the domains that are necessary for recognition and the aminoacids involved. We provide and validate a structural model for the interaction through protein-peptide docking and molecular dynamics simulations. We also extend this analysis to two other NPM1 interacting proteins, namely Tat[26] and CENP-W,[27,28] demonstrating that the same region of NPM1 recognizes all these proteins, substantiating the proposed role of NPM1 as a 'nucleolar hub'. We suggest that this protein region may be targeted for the treatment of AML with NPM1 mutations.

RESULTS

NPM1 interacts with the predicted NoLS sequence of Fbw7γ

In an effort to understand the molecular mechanism whereby Fbw7γ localizes in nucleoli, we carried out a bioinformatic analysis of Fbw7 isoforms using the NoD algorithm (Nucleolar Localization Signal Detector; http://www.compbio.dundee.ac.uk/www-nod/index.jsp) to identify putative nucleolar localization signals (NoLS) in these proteins (Figures 1a–c). This analysis, which relies on sequence only, is based on the observation that the NoLS of many proteins consists of a short motif rich in lysines and arginines positioned in variably spaced clusters.[20] The results showed that only the N-terminal region of Fbw7γ contains a putative NoLS (Figures 1c and d), while a partial signal is present in Fbw7α (nucleoplasmic) (Figures 1a and d) and absent in Fbw7β (cytosolic) (Figures 1b and d), in agreement with their observed cellular localization.[23] In all three isoforms, this region of the protein is predicted to be natively unstructured (Figure 1e) and a higher amount of positive charges within few residues can be appreciated in the γ isoform with respect to the other two. Indeed the α and β isoforms contain insertion regions, separating the two clusters of positive charges found in the γ isoform, endowed with a higher conformational entropy and smaller stability, which possibly preclude their recognition as NoLS.

Since Fbw7γ nucleolar localization depends on the presence in nucleoli of NPM1,[21] we hypothesized that its predicted NoLS is the epitope recognized by NPM1 and we investigated the molecular determinants of this interaction. To this purpose, we used both the N-terminal and C-terminal regions of NPM1 (Nter-NPM1, residues 16-123 and Cter-NPM1, residues 225-294, respectively), and a peptide encompassing residues 43-56 of the Fbw7γ sequence (hereafter Fbw7γ*), consisting of the central region of the Fbw7γ predicted NoLS and containing six positively charged residues arranged in two clusters (Table 1). The binding process

was monitored by equilibrium fluorescence spectroscopy taking advantage of a dansyl group attached to the peptide N terminus. Titrations showed an increase of fluorescence dansyl emission with a blue shift of the emission peak as a function of Nter-NPM1 concentration (see Figure 2a inset). Analysis of the data according to Equation (1) (Figure 2a) yielded an equilibrium dissociation constant $K_D = 3.2 \pm 0.6$ μM (Table 1). Data were well fitted with a 1:1 peptide/Nter-NPM1 monomer stoichiometry, even though Nter-NPM1 is pentameric at all concentrations tested in our experiments.

When the C-terminal domain of NPM1 (Cter-NPM1) was used, no variation in the emission peak was observed, indicating that it does not bind Fbw7γ* (Figure 2b). We also performed the same experiments using dansylated peptides (named unrelated long and unrelated short) whose sequences are not present in the Fbw7γ isoform and which are not predicted to be NoLS (Table 1). No interaction was observed when these peptides were titrated with Nter-NPM1.

Table 1. Dissociation constants for the complexes between Nter-NPM1 and peptides

Peptide	*Sequence*	*Nter-NPM1* K_D *(μM)*
Fbw7γ*	^{43}LPFCRRRMKRKLDH^{56}	3.2 ± 0.6
CENP-W*	^{14}KRKAPRGFLKRVFKRKK^{30}	6.2 ± 0.9
Tat*	^{47}AGRKKRRQRRRPPQ^{60}	2.4 ± 0.5
Unrelated long	DDEAQTLAKFVLSQK	Ni
Unreleted short	VLSQK	Ni

Abbreviation: Ni, no interaction.

Since it has been reported that NPM1 interacts also with CENP-W and Tat proteins,[26,27] we used the NoD algorithm to identify in these two proteins their NoLS. Then, dansylated peptides corresponding to the suggested NoLS regions of the proteins (peptides CENP-W* and Tat*, respectively) (Table 1) were tested for their interaction with Nter-NPM1 (Figures 2c and d) and Cter-NPM1 (Supplementary Figure 1). For both peptides, equilibrium titrations experiments showed an increase of fluorescence emission as a function of protein concentration only when the Nter-NPM1 domain was used (Figures 2c and d), demonstrating that the interaction specifically involves the N-terminal domain and that, also in these cases, the predicted NoLS is the binding epitope recognized by NPM1. The calculated dissociation constants parallel the one obtained for Fbw7γ, being $K_D = 6.2 \pm 0.9$ μM for the Nter-NPM1-CENP-W* interaction and $K_D = 2.4 \pm 0.5$ μM for the Nter-NPM1-Tat* interaction (Table 1).

Identification of Nter-NPM1 residues involved in Fbw7γ recognition

Nter-NPM1 monomer consists of eight antiparallel β-strands forming a β-barrel with jelly-roll topology. Five monomers tightly associate to form a crown shaped pentamer. Since the identified Nter-NPM1-interacting epitopes are all enriched in positive

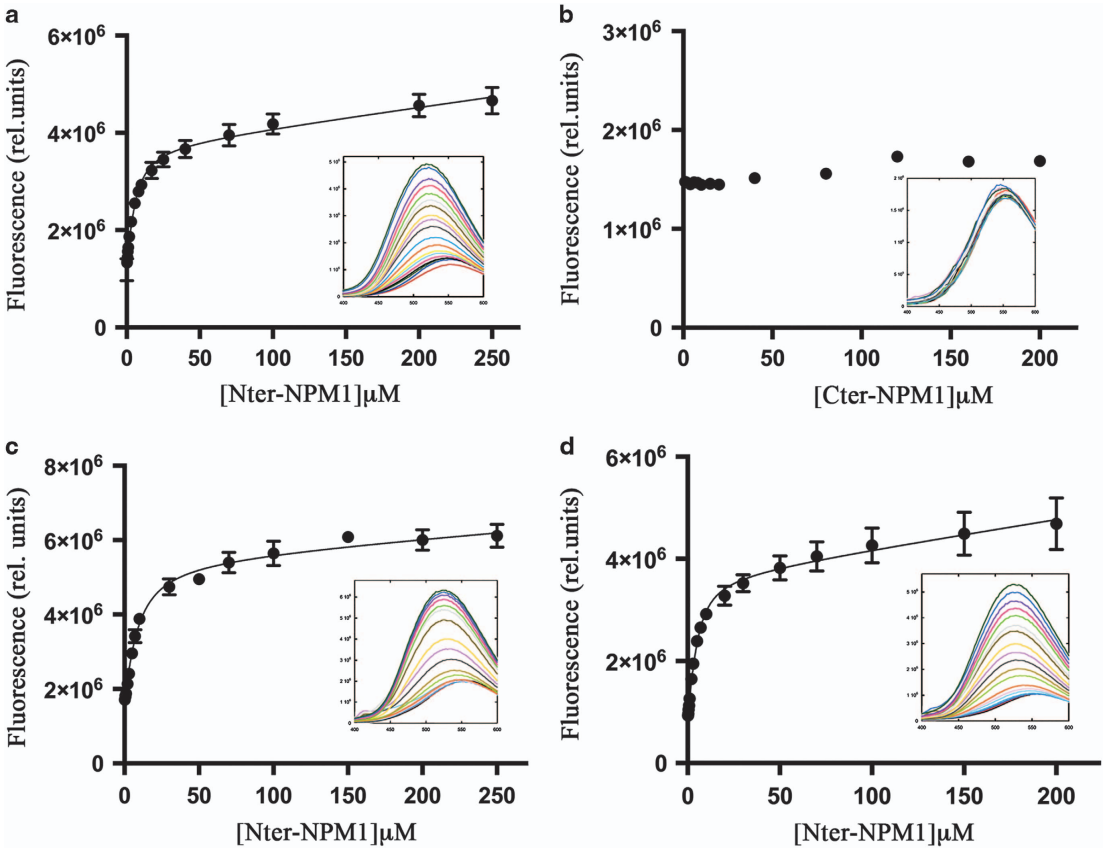

Figure 2. Interaction analysis of NoLS sequences. Peptides corresponding to the putative NoLS were dansylated at their N terminus and titrated with NPM1 constructs. The static fluorescence spectra are shown in insets, while the experimental maxima and their fit according to equation 1 (see Materials and Methods) are reported as a function of NPM1 concentrations in the main panels as follows: (**a**) Interaction between Fbw7γ* and Nter-NPM1. (**b**) Interaction between Fbw7γ* and Cter-NPM1. (**c**) Interaction between CENP-W* and Nter-NPM1. (**d**) Interaction between Tat* and Nter-NPM1. Peptides sequences are reported in Table 1.

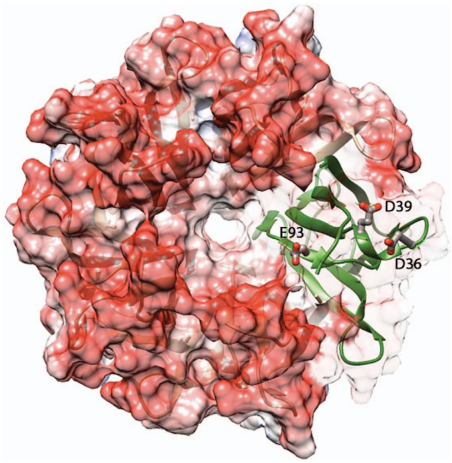

Figure 3. Electrostatic potential surface analysis of Nter-NPM1. The crystal structure of human Nter-NPM1 displays a pentameric organization and was subjected to calculation of the electrostatic potential surface through the APBS algorithm. Negative and positive charges are shown in red and blue, respectively. One of the monomers is shown in ribbon to better highlight the position of three important acidic residues (D36, D39 and E93), which are shown in sticks.

Table 2. Dissociation constants for the complexes between the Fbw7γ* peptide and Nter-NPM1 mutants

Protein	K_D *(μM)*
Nter-NPM1	3.2 ± 0.6
D36A	10.8 ± 2.6
E37A	12.5 ± 1.4
E39A	6.2 ± 1.5
E93A	5.0 ± 1.6
E121A	13.3 ± 3.8
D36A-E39A	13.5 ± 2.0
D36A-E93A	8.4 ± 1.5
E39A-E93A	7.6 ± 1.9
D36A-E39A-E93A	22.0 ± 3.0
D36A-E37A-E39A-E93A	151.5 ± 20.5
D36A-E39A-E93A-E121A	224.3 ± 35.9
D36A-E37A-E39A-E93A-E121A	653.7 ± 46.9

charges, we inspected the electrostatic potential surface of Nter-NPM1 in search for negatively charged patches. Indeed, as shown in Figure 3, a large negatively charged surface is found, extending from the pentamer external surface to its central cavity. Among the residues that contribute to this negatively charged surface, we focused our attention on three residues, namely D36, E39 and E93, because they were previously shown to play a role in the interaction of NPM1 with the tumor suppressor p14ARF.[29] To establish their involvement in binding Fbw7γ*, these residues were all mutated to alanine, as single or double mutants; a triple mutant was also prepared. The interaction was measured using the same protocol as for the wild-type protein (Supplementary Figure 2) and the resulting K_D are reported in Table 2. Mutation to alanine of D36 and E39 residues led to an increase of K_D between two and three fold. Mutation of E93 had instead a smaller effect. Consistently, when the double mutants D36A-E93A and E39A-E93A where tested, observed K_D were comparable to those obtained with the D36A and E39A single mutants, respectively. When the D36A-E39A double mutant was tested no dramatic additive effect of the two mutations in destabilizing the interaction was obtained with $K_D = 13.5 \pm 2.0$ μM. Finally, the triple mutant D36A-E39A-E93A resulted to bind the peptide with a $K_D = 22 \pm 3$ μM, around seven fold higher than wild type. These data suggest that though residues D36, E39 and to a minor extent E93, contribute to the binding energy, the overall binding process is not entirely dependent on their interactions. Additional residues are likely involved.

In order to obtain a better description of the interaction and identify additional residues involved, we performed a molecular docking analysis of the Nter-NPM1- Fbw7γ* complex. Given the complexity of docking a long and flexible peptide, the procedure adopted here was based on a 'divide and conquer' approach, starting from a combinatorial merging of energy-favorable tripeptides, which were then used as templates for biased-guided docking (see Materials and Methods for a detailed description). The top ten scoring docking poses cluster in the same binding site, with a mean RMSD between poses of 4.3 Å (Supplementary Figure 3). This is principally the result of the variable conformations adopted by the C-terminal half of the peptide, while the N-terminal half appears more fixed. Only a single peptide was docked onto the Nter-NPM1 pentamer, nevertheless equivalent docking surfaces are available for the additional peptides. It is possible that different peptide conformations, among those selected by the docking procedure, would be adopted when all Nter-NPM1 monomers are engaged by peptides.

Figures 4a and b show the best scoring docking pose and, according to this model, Fbw7γ* adopts an extended conformation and lies with its N-terminal end along the external surface of the protein while its C-terminal end protrudes into the central cavity of the pentamer. The majority of interactions are established with residues belonging to a single Nter-NPM1 monomer, with few contributions from the adjacent one (see below). Interestingly, although no information regarding interacting residues was imparted to the docking algorithm, all three residues that we have examined before (D36, D39 and E93) were found to interact with positively charged residues of the Fbw7γ* peptide. In particular, in the docking model, D36 forms a salt bridge with Fbw7γ* R47 (Figure 4c), E39 is salt-bridged to both R48 and R52 (Figure 4d), while E93 interacts with K53 (Figure 4e).

Inspection of the model showed two additional negatively charged residues that interact with Fbw7γ*. The first one is E37 that interacts with R47 (Figure 4f); the second one is E121: in this case the same residue from two different Nter-NPM1 monomers binds two different residues of the peptide, i.e., K51 and R52 (Figure 4g). Therefore, we mutated these two additional residues and measured their contribution to the binding energy (Supplementary Figure 2). When E37A and E121A single mutants were tested, we obtained dissociation constants approximately four-fold higher than wild-type, similarly to what already seen for the D36 and E39 mutants (Table 2). Furthermore, starting from the D36A-E39A-E93A triple mutant, we also prepared two quadruple mutants and a quintuple one. When the D36A-E37A-E39A-E93A mutant was tested, we obtained a $K_D = 151.5 \pm 20.5$ μM, ≈50-fold higher than wild type. Likewise, the D36A-E39A-E93A-E121A mutant showed a $K_D = 224.3 \pm 35.9$ μM, ≈70-fold higher than wild type. Finally, the quintuple mutant displayed negligible affinity for the peptide, with $K_D = 653.7 \pm 46.9$ μM (Table 2).

Overall, these data suggest that the Nter-NPM1-Fbw7γ* binding energy is dictated by multiple electrostatic contributions throughout the binding cleft. They also show a non-linear variation of the K_D upon loss of negative charges in Nter-NPM1. In fact the dissociation constant is relatively stable when only three residues are mutated, while it markedly increases upon addition of a fourth mutation. To check the effect of salt on binding we also performed titrations with wild-type protein and triple, quadruple and quintuple mutants increasing the ionic strength to 150 mM. We

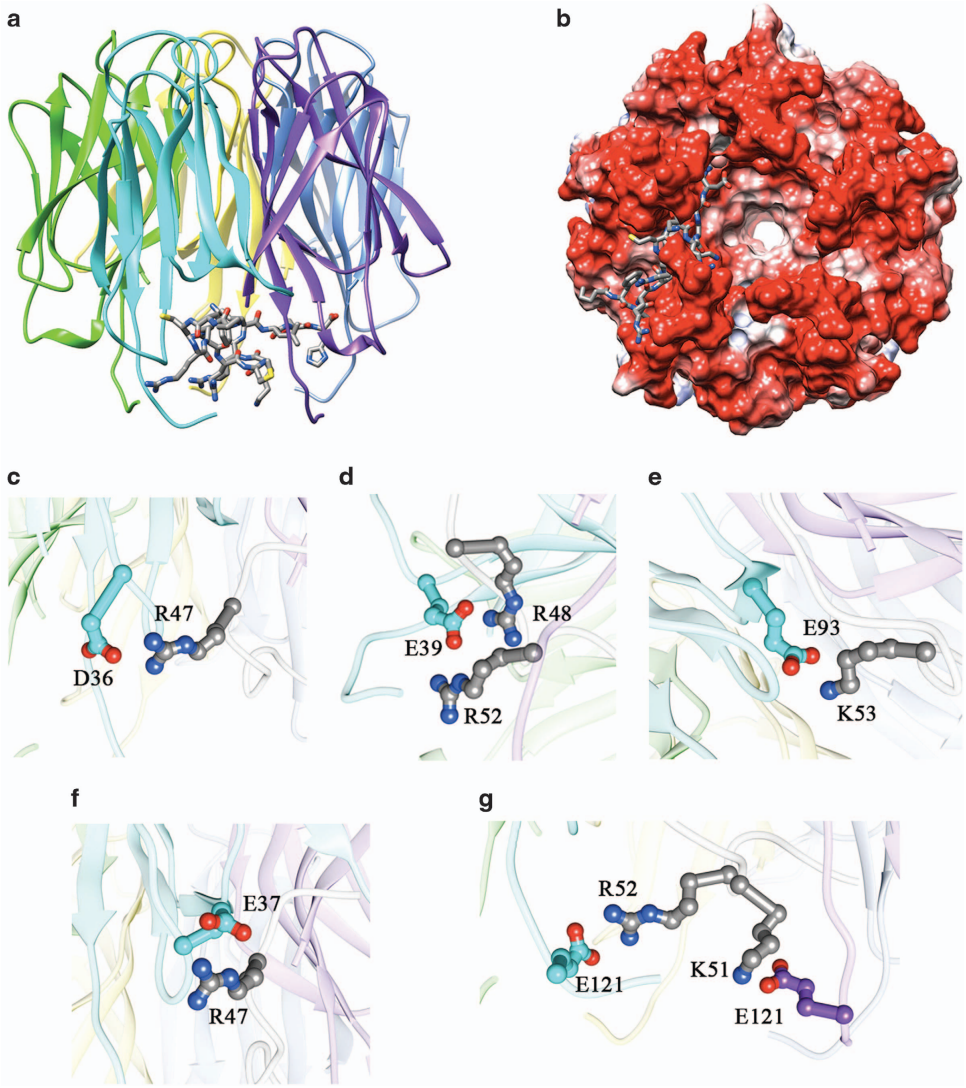

Figure 4. Molecular docking analysis of the Fbw7γ*-Nter-NPM1 interaction. The interaction between Fbw7γ* and Nter-NPM1 investigated through molecular docking analysis is shown. (**a**) Nter-NPM1 pentamer is represented in cartoon while the Fbw7γ* peptide is shown in sticks. (**b**) The Nter-NPM1 electrostatic surface is shown in a different orientation from **a**. The peptide, shown in sticks, adopts an extended conformation with its C-terminal end protruding into the central pentamer cavity. (**c**) A detail of the interaction played by Nter-NPM1 residue D36 with Fbw7γ* R47. (**d**) Nter-NPM1 residue is predicted to interact with both Fbw7γ* R48 and R52 residues. (**e**) Interaction between Nter-NPM1 E93 and Fbw7γ* K53. (**f**) Nter-NPM1 residue E37 is also predicted to interact with Fbw7γ* R47. (**g**) E121 residues from two different Nter-NPM1 monomers (the second one is shown in magenta) are predicted to interact with residue K51 and R52 residues.

obtained a general decrease in affinity, as expected, but the trend observed with the previous experiments was confirmed (Supplementary Table 1).

The same surface of Nter-NPM1 recognizes peptides from different protein partners

Since the peptides from CENP-W and Tat recognized by Nter-NPM1 share with Fbw7γ* a high positive charge (Table 1), we investigated whether the same Nter-NPM1 surface is implicated in their binding. To this end, we tested the triple, quadruple and quintuple Nter-NPM1 mutants interaction with the CENP-W* and Tat* peptides (Supplementary Figure 4). In both cases, and similarly to what already seen for Fbw7γ*, a clear trend of increasing dissociation constants is observed as a function of decreasing negative charges from the triple to the quintuple mutant (Table 3). These results indicate that the same region of Nter-NPM1 is responsible for the interaction with different nucleolar proteins that are all recognized through their predicted NoLS.

Table 3. Dissociation constants for the complexes between Nter-NPM1 selected mutants and the CENP-W* and Tat* peptides

	*CENP-W** K_D (μM)	*Tat** K_D (μM)
Nter-NPM1	6.2 ± 0.9	2.4 ± 0.5
D36A-E39A-E93A	18.4 ± 2.6	11.1 ± 2.1
D36A-E37A-E39A-E93A	71.4 ± 9.1	57.4 ± 7.0
D36A-E39A-E93A-E121A	82.8 ± 5.6	63.8 ± 5.0
D36A-E37A-E39A-E93A-E121A	734.0 ± 158.0	642.6 ± 70.8

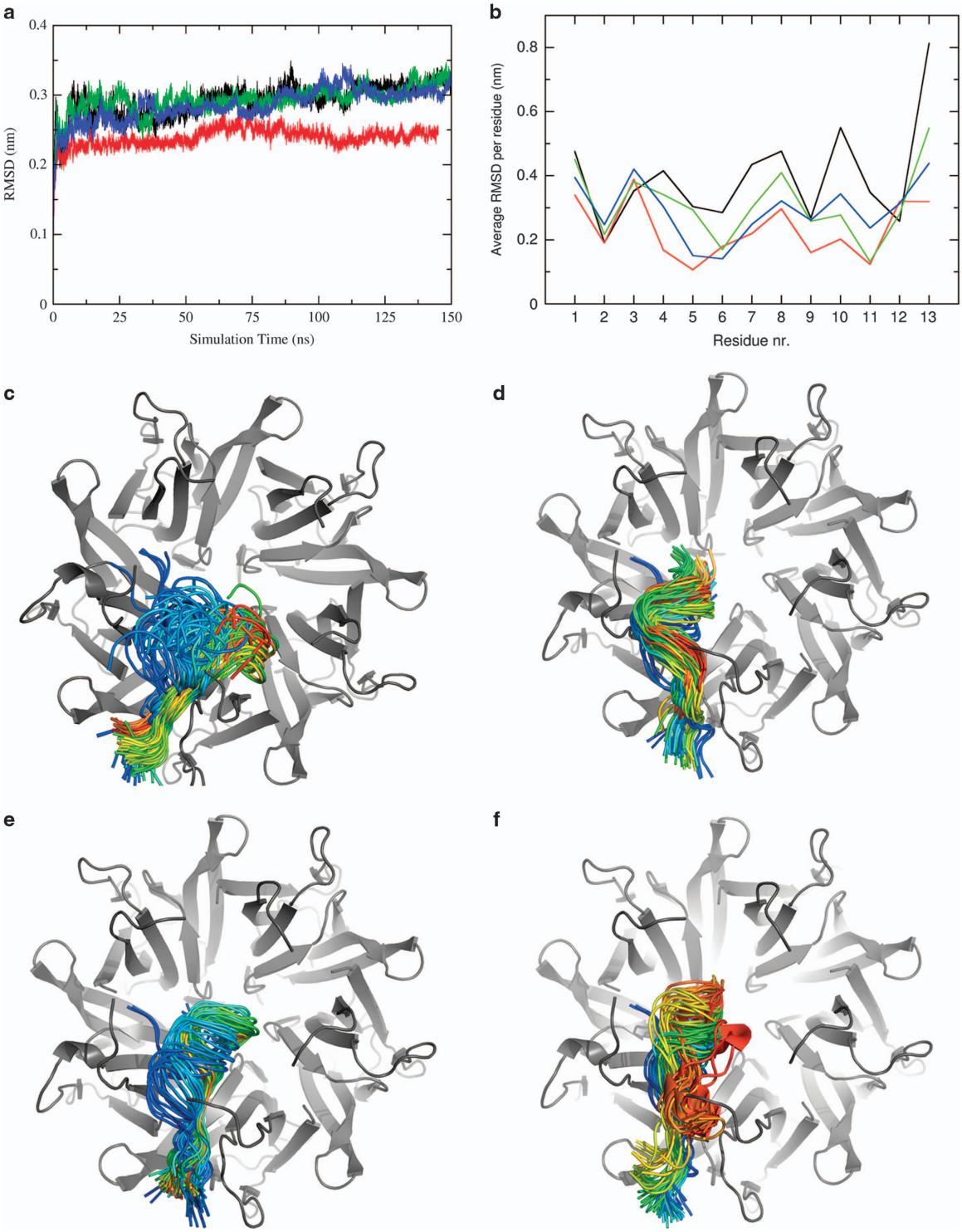

Figure 5. Molecular dynamics simulations of the interaction between Fbw7γ* and Nter-NPM1 constructs (wild-type and mutants). (**a**) Root mean square deviation (RMSD) of Nter-NPM1 and Fbw7γ* Cα atoms as a function of simulation time for WT (black line), triple (red), quadruple (green) and quintuple (blue), respectively. (**b**) Average RMSD (root mean square deviation) of Fbw7γ* as calculated for peptide Cα residues for all simulated systems. Wild-type Nter-NPM1 is shown in black, the triple D36A-E39A-E93A mutant is shown in red, the quadruple D36A-E39A-E93A-E121 mutant is shown in green, the quintuple D36A-D37A-E39A-E93A-E121 is shown in blue. (**c**) The position of the peptide at selected snapshots along the simulation time is shown. Nter-NPM1 wild-type is shown in gray cartoon, Fbw7γ* is shown as a ribbon colored from blue (simulation start time) to red (simulation end). (**d**) Same as in **c** for the interaction between the peptide and the triple D36A-E39A-E93A mutant. (**e**) Same as in **c** for the interaction between the peptide and the quadruple D36A-E39A-E93A-E121 mutant. (**f**) Same as in **c** for the interaction between the peptide and the quintuple D36A-D37A-E39A-E93A-E121 mutant.

Molecular dynamics simulations

To gain further insights into the structural requirements for binding, we performed extended molecular dynamics (MD) simulations on the model structure for the complex between Fbw7γ* and Nter-NPM1. Given their increasing effect on the dissociation constant of the complex, we simulated also the complexes formed by Fbw7γ* with the D36A-E39A-E93A triple mutant, the D36A-E39A-E93A-E121A quadruple mutant and the D36A-E37A-E39A-E93A-E121A quintuple mutant. Total simulation time for all systems was 150 ns. We firstly determined the convergence and stability of the MD trajectories, in order to ascertain the validity of conformational sampling in all simulated systems. To this end, the root mean square deviations (RMSD) of Cα coordinates of wild type and mutants Nter-NPM1 were calculated as a function of simulation time (Figure 5a). This analysis confirmed that the trajectories reached a plateau of the RMSD, a regime compatible with the conformational drift of a folded structure and that the simulation time was sufficient to equilibrate the protein dynamics.

This enabled us to investigate the nature of the Nter-NPM1-Fbw7γ* interaction by analyzing the relative conformational drift of Fbw7γ* with respect to the Nter-NPM1 and mutants structures. Figure 5b shows the averaged RMSD per peptide residue, as calculated for Cα atoms, for all simulated systems. This analysis suggests that the Fbw7γ* peptide is stabilized in the binding surface of wild-type Nter-NPM1 through interactions involving mainly its N-terminal residues 1–6, which keep a position similar to the starting structure throughout the whole simulation. This is also represented in Figure 5c, showing snapshots of the Nter-NPM1-Fbw7γ* simulation, with the conformation adopted by the peptide at different times from the beginning (blue) to the end (red) of the simulation. Differently from the N-terminal end, the C-terminal region of the peptide, which protrudes inside the central cavity of the Nter-NPM1 pentamer, populates several conformations along the simulation time (Figure 5c). Overall, the interaction of wild-type Nter-NPM1 with Fbw7γ* has the highest RMSD values as compared with the mutants (Figure 5b). It appears that no single ion pair is absolutely required for the interaction because adjacent negative residues can replace the loss of a contact. We speculate that such binding mode allows the stabilization of the peptide into the cleft without a significant entropy loss, since many energy minima can be efficiently explored by the peptide (Figure 5c). This also explains why single and double NPM1 mutants show only a weak decrease of the binding affinity for Fbw7γ*.

The analysis of the RMSD over peptide length for the triple, quadruple and quintuple mutants suggest that in all mutants the positions explored by the peptide are less variable as compared with wild type, as shown in Figure 5b which represents a measure of the average displacement of each peptide Cα atom with respect to its starting position. Indeed, snapshots of the simulations at different times indicate that the peptide docks into the binding surface of the mutants maintaining an overall less variable conformation along the simulation time (Figures 5d–f). This observation can be rationalized by taking into account the hydrophobic interactions played by the newly introduced alanine residues in the mutants, which are favored and replace many of the electrostatic interactions previously observed for the wild-type protein.

Overall we hypothesize that while Fbw7γ* is still able to bind all NPM1 mutants, the sequential loss of negative charges may be associated with a significant entropy loss upon binding and a progressively decreased affinity.

DISCUSSION

In this work we investigated the structural basis of NPM1 protein–protein associations. We started from the hypothesis that NPM1 is a 'nucleolar hub' protein because it recognizes the NoLS of different protein partners. In most proteins, the NoLS consists of a linear stretch of aminoacids, within a natively disordered region, that is rich in clustered arginines and lysines. Even though a clear sequence motif cannot be envisaged, such accumulation of positive charges within few residues is uncommon in proteins and may be searched for by specific algorithms. We employed one such algorithm to spot putative NoLS in representative proteins that are known to interact with NPM1 and to be nucleolar. Then, we showed that all these epitopes from different proteins were effectively recognized by Nter-NPM1.

A deeper analysis of the interaction highlighted additional concepts. First, we showed that at least five negatively charged residues of Nter-NPM1 contact the peptides. Then, by mutating them alone and in combinations, we could argue against the existence of specific hot-spots. Rather, we observed a substantial stability of the complex when negative charges where replaced alone or in couples, and appreciated gradually increasing dissociation constants when three to five residues where mutated at once. Molecular dynamics simulations provided a plausible mechanism to interpret these observations and suggested that the absence of hot-spots is the consequence of the fact that the loss of one interaction may be compensated by the emergence of a new one. This is possible because the peptide does not stably populate a single conformation but moves rather freely within an extended binding surface provided by Nter-NPM1. Therefore the loss of a single or a couple of ion-pairs that would destabilize one conformation may be compensated by the adoption of an alternative one by the peptide.

Such a model may also explain the amazing versatility of Nter-NPM1 in binding epitopes from a plethora of other proteins. The peptides from Fbw7γ, TAT and CENP-W that we identified here, as well as several other peptides studied by others,[3,29,30] are all positively charged but differ in residue composition, in the number of charged residues and in their position along the sequence. How can a single protein recognize them all with similar affinities? The model we propose implies that all peptides will find, within the large negatively charged surface provided by the NPM1 pentamer, a multitude of binding poses and will populate those that are more convenient to their particular distribution of positive charges. We speculate that such mechanism is at the heart of NPM1 behavior as a nucleolar hub protein.

But how NPM1 is itself enriched in nucleoli? Previous research from our and other laboratories has clarified this issue. The NPM1 nucleolar localization signal is unique and consists of W288 and W290 residues near the C terminus of the protein. These tryptophans take part to the hydrophobic core of the C-terminal three-helix bundle domain and their substitution with other residues leads to the unfolding of the C-terminal domain.[31–34] An unfolded C-terminal domain is in turn unable to interact with nucleic acids, most prominently G-quadruplex regions at ribosomal DNA, resulting in detachment from nucleoli.[35–37] Therefore, the C-terminal domain of NPM1 keeps it at nucleoli while the N-terminal domain sequesters its binding partners in the same organelle. When NPM1 moves from nucleoli, because of post-translational modifications or mutations, the NPM1 protein partners will be equally displaced, because of their interaction with the N-terminal domain.

This is ultimately what happens in AML with *NPM1* gene mutations. Mutations cause the unfolding of the C-terminal domain and consequent loss of affinity for nucleoli. Furthermore, since a new NES appears in the mutated protein, this is aberrantly translocated in the cytosol, carrying with itself protein partners like Fbw7γ and p14ARF, which will be there degraded.[21,24] Moreover, the presence in the cytosol of mutated NPM1 with a functional N-terminal domain, will result in the establishment of additional protein–protein interactions. For instance, cytosolic NPM1 binds and inhibits caspases 6 and 8[38] and the PTEN deubiquitinating

enzyme HAUSP, resulting in PTEN cytoplasmic polyubiquitinilation and degradation.[39] Thus a third important tumor suppressor is also deregulated by the presence of NPM1 in the cytosol.

AML with *NPM1* mutation is currently treated with the administration of several cycles of an anthracycline (daunorubicin, doxorubicin or others) plus cytarabine.[40] Patients carrying *NPM1* mutation without the concomitant *FLT3-ITD* alteration have good prognosis while, for the latter, chemotherapy is less effective. However, relapse is frequent and the toxicity of anthracyclines prevents many patients from its prolonged use. Importantly, NPM1 mutations are always retained at relapse and this led to the generally accepted concept that NPM1 should be specifically targeted in this kind of leukemia.[41]

Based on the experimental observations we outlined above, we have recently suggested that an effective strategy to specifically target AML with *NPM1* mutations would be that of interfering with NPM1 protein–protein interactions.[16] Here, we have characterized the extended surface of Nter-NPM1 involved in protein binding and thus provided a structural framework to search for small molecules and/or peptidomimetics targeting this surface. Future work will be directed at testing these concepts in cellular models of AML with *NPM1* mutations.

MATERIALS AND METHODS

NoLS identification

To identify NoLSs in the proteins of interests to this work we employed the method described by Scott *et al*.[20] and implemented in the NoD web server (http://www.compbio.dundee.ac.uk/www-nod/). Briefly, the NoD algorithm uses an artificial neural network trained on a large set of NoLs experimentally evaluated, to analyze a sequence in search of local enrichments of positively charged residues. The sequence of a protein is scanned in windows of 13 residues with slippage of one amino acid for each consecutive window and at each window is assigned a score, which depends on the number of positive charges. When the score is greater than 0.8, the relative sequence is identified as a predicted NoLs.

Protein constructs

The Nter-NPM1 (residues 16-123) coding sequence was obtained through gene synthesis (GeneArt, Regensburg, Germany) and cloned into the expression vector pET28+(a) (Novagen, San Diego, CA, USA) using NdeI and BamHI restriction enzymes.

Nter-NPM1 mutants were obtained by site-directed mutagenesis using the Quickchange II Lightning Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA, USA), following manufacturer's instructions. Oligonucleotides used for PCR were obtained from Primm Biotech (Milan, Italy). Forward oligonucleotides used are reported in Supplementary Table 2.

Protein expression and purification

Escherichia coli cells, BL21(DE3) (Biolabs, Ipswich, MA, USA), transformed with the expression vectors were grown to $A_{600} \sim 0.5$ in LB medium supplemented with kanamycin at 37 °C. Expression was induced by addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and cells were further grown at 20 °C for 16h. Cells were collected, resuspended in lysis buffer (Buffer A: 20 mM Hepes, pH 7.0, 20 mM imidazole), plus 5 mM $MgCl_2$, 2 μg/ml DNAse (Roche, Basel, Switzerland), Protease Inhibitor Cocktail Tablet (Roche) and sonicated. Nucleic acids were digested for 30' at 37 °C with DNAse I. Proteins were purified by affinity chromatography (HisTrap FF, GE Healthcare, USA) using a linear gradient of buffer A plus imidazole (from 20 mM to 1M). Further purification involved anion exchange chromatography (Q-Sepharose Fast Flow, GE Healthcare, USA) eluted with NaCl gradient. Fractions containing the protein, as showed by SDS–PAGE, were collected and concentrated using Amicon Ultra-15 Centricons with a 3K cut-off (Merck Millipore, Darmstadt, Germany). The protein solutions were buffer exchanged with Hepes 20 mM, pH 7.0 and stored at − 20 °C. The Cter-NPM1 protein construct (residues 225-294) was expressed and purified as previously described.[36]

Equilibrium binding experiments

All experiments were performed at 25 °C in sodium phosphate buffer 20 mM pH 7.2, or in sodium phosphate buffer 20 mM plus 100 mM NaCl pH 7.2 (which sets ionic strength to 150 mM), using a FluoroMax-4 spectrofluorometer (Jobin Yvon, Edison, NJ, USA), equipped with a water bath apparatus. Fbw7γ and CENP-W peptides, functionalized with a dansyl (5-dimetilamminonaftalen-1-sulphonyl) group at their N terminus, were purchased from JPT (Germany). Tat peptide was synthesized employing the solid phase method following standard Fmoc strategies and labeled with dansyl fluorophore at its N terminus. Crude product was purified by RP-HPLC applying a linear gradient of 0.1% TFA CH_3CN in 0.1% TFA water from 5% to 65% over 12 min using a semi-preparative 2.2×5 cm C18 column at a flow rate of 20 ml/min. Purity and identity were confirmed by LC–MS analysis.

Titration experiments were conducted with an excitation wavelength of 330 nm while the fluorescence emission spectra were collected in the range between 350 and 650 nm. Titrations were performed at constant peptide concentration (5 μM) and varying protein concentrations (from 0 to 200/400 μM). Titrations were performed in triplicate and data, analyzed with the Graphpad Prism software (https://www.graphpad.com/scientific-software/prism/), were reported as dissociation constant ± s.d.

Equilibrium binding curves were fitted to the standard quadratic equation:

$$F = \left\{ \left[\frac{([A]_0 + K_D + n)}{2} - \sqrt{\frac{([A]_0 + K_D + n)^2}{4} - [A]_0 n} \right] B + C \right\} k \qquad (1)$$

where F is the observed fluorescence signal, n and $[A]_0$ are the total concentration of non-varied and varied species, respectively, and K_D is the equilibrium dissociation constant. B and C are constants taking into account the total fluorescence change and fluorescence at $[A]_0 = 0$, respectively, and k is a term describing the slope of the curve at high protein concentration[42]. Whenever possible, under pseudo first-order conditions, the equation 1 was simplified to:

$$F = \frac{[A]_0}{[A]_0 + K_D} \qquad (2)$$

Molecular docking

In order to predict the binding mode of the peptide N_{ter}-PFCRRRMKRKLDH-C_{ter} to NPM1, tripeptides covering the whole sequence were exhaustively generated by an ad-hoc Python script.[43] The obtained peptides were then energy minimized by using the Molecular Operating Environment 2009.10 (http://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm). Steepest descents steps of energy minimization were performed until the root mean square (RMS) gradient fell below the 0.005 Å default threshold. The Amber99 force field, a distance-dependent dielectric constant and a cut-off distance of 40 Å were used during each simulation.

Molecular docking of the tripeptides was carried out by means of Molegro Virtual Docker (MVD) software (®CLCbio).[44] Flexible torsions were automatically detected by MVD, and manually checked for consistency. The structure of NPM1 (PDB: 2P1B) was prepared by automatically assigning bond orders and hybridization, and adding explicit hydrogens, charges and Tripos atom types. Missing heavy atoms were fixed by modeling them, using Modeler v.9.8[45] and PyMod.[46] A search space of 20Å radius, centered on the central cavity of the pentamer (~4385 $Å^3$) was used for docking. Cavity detection was carried out by MVD. For each tripeptide, ten runs were defined. Similar poses (RMSD < 1.2 Å) for each tripeptide were clustered, and the best scoring one was taken as representative. Other docking parameters were fixed at their default values. Thereafter, hexamer peptide sequences and their structures were generated and docked in a second round, by considering the poses of the tripeptides identified from the first round of docking runs. To this end, the most energetically favorable poses of the tripeptides were taken as template groups for template-based dockings. Finally, the 100 top scoring hexamer peptides poses were taken as template groups for template-based docking of the whole peptide fragment, using the same above-described approach. The obtained top scoring complex was subjected to a final energy minimization, using conjugated gradients until the maximum derivative was less than 0.0004 kJ $mol^{-1}Å^{-1}$.

Molecular dynamics simulations

MD simulations of complexes were performed starting from the final refined model obtained in docking calculations. Mutations were introduced with the Pymol software (www.pymol.org).

Simulation setup. Calculations were performed using GROMACS 5.0.x (www.gromacs.org) suite with the Amber99 force field. Initial structures were immersed in a triclinic simulation box, solvated with SPC water molecules.[47] Ionic strength was adjusted to set the total charge of simulation box to 0. All simulations were performed in the NVT ensemble at constant volume and constant temperature (300 K), periodic boundary conditions were applied. Initial velocities were taken from the Maxwell-Boltzmann distribution at 300 K. Long-range electrostatic interactions were calculated using the particle mesh Ewald method[48] with a 1.2 nm cut-off for the real space calculation. A 1.2 nm cut-off was used to estimate Van der Waals interactions. Pair list was updated every 10 steps. The LINCS algorithm[49] was used to constrain bond lengths; the time step for integration was 2 fs.

Simulation protocol. Intial structures for all simulations were subjected to a steepest descent minimizzation cycle to reduce steric hindrance. Then a restrained MD step-wise procedure was applied to gradually release the restraints and allow the system to equilibrate at the simulated temperature of 300 K: applied restrained were 1000, 500 and 250 kJ/mol. Total simulation time was typically 150 ns. Coordinates were saved every 5 ps.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This paper was supported by AIRC Associazione Italiana Ricerca sul Cancro IG-Grant 2014-15197 to LF.

PUBLISHER'S NOTE

REFERENCES

1 Grisendi S, Mecucci C, Falini B, Pandolfi PP. Nucleophosmin and cancer. *Nat Rev Cancer* 2006; **6**: 493–505.
2 Emmott E, Hiscox JA. Nucleolar targeting: the hub of the matter. *EMBO Rep* 2009; **10**: 231–238.
3 Mitrea DM, Cika JA, Guy CS, Ban D, Banerjee PR, Stanley CB *et al.* Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *Elife* 2016; **5**: e13571.
4 Lindstrom MS. NPM1/B23: a multifunctional chaperone in ribosome biogenesis and chromatin remodeling. *Biochem Res Int* 2011; **2011**: 195209.
5 Colombo E, Alcalay M, Pelicci PG. Nucleophosmin and its complex network: a possible therapeutic target in hematological diseases. *Oncogene* 2011; **30**: 2595–2609.
6 Federici L, Falini B. Nucleophosmin mutations in acute myeloid leukemia: a tale of protein unfolding and mislocalization. *Protein Sci* 2013; **22**: 545–556.
7 Herrera JE, Savkur R, Olson MO. The ribonuclease activity of nucleolar protein B23. *Nucleic Acids Res* 1995; **23**: 3974–3979.
8 Murano K, Okuwaki M, Hisaoka M, Nagata K. Transcription regulation of the rRNA gene by a multifunctional nucleolar protein, B23/nucleophosmin, through its histone chaperone activity. *Mol Cell Biol* 2008; **28**: 3114–3126.
9 Scott DD, Oeffinger M. Nucleolin and nucleophosmin: nucleolar proteins with multiple functions in DNA repair. *Biochem Cell Biol* 2016; **94**: 419–432.
10 Ziv O, Zeisel A, Mirlas-Neisberg N, Swain U, Nevo R, Ben-Chetrit N *et al.* Identification of novel DNA-damage tolerance genes reveals regulation of translesion DNA synthesis by nucleophosmin. *Nat Commun* 2014; **5**: 5437.
11 Okuwaki M, Matsumoto K, Tsujimoto M, Nagata K. Function of nucleophosmin/B23, a nucleolar acidic protein, as a histone chaperone. *FEBS Lett* 2001; **506**: 272–276.
12 Szebeni A, Olson MO. Nucleolar protein B23 has molecular chaperone activities. *Protein Sci* 1999; **8**: 905–912.
13 Okuda M. The role of nucleophosmin in centrosome duplication. *Oncogene* 2002; **21**: 6170–6174.
14 Wang W, Budhu A, Forgues M, Wang XW. Temporal and spatial control of nucleophosmin by the Ran-Crm1 complex in centrosome duplication. *Nat Cell Biol* 2005; **7**: 823–830.
15 Yang K, Wang M, Zhao Y, Sun X, Yang Y, Li X *et al.* A redox mechanism underlying nucleolar stress sensing by nucleophosmin. *Nat Commun* 2016; **7**: 13599.
16 Di Matteo A, Franceschini M, Chiarella S, Rocchio S, Travaglini-Allocatelli C, Federici L. Molecules that target nucleophosmin for cancer treatment: an update. *Oncotarget* 2016; **7**: 44821–44840.
17 Yung BY. Oncogenic role of nucleophosmin/B23. *Chang Gung Med J* 2007; **30**: 285–293.
18 Holmberg Olausson K, Elsir T, Moazemi Goudarzi K, Nistér M, Lindström MS. NPM1 histone chaperone is upregulated in glioblastoma to promote cell survival and maintain nucleolar shape. *Sci Rep* 2015; **5**: 16495.
19 Falini B, Mecucci C, Tiacci E, Alcalay M, Rosati R, Pasqualucci L *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med* 2005; **352**: 254–266.
20 Scott MS, Troshin PV, Barton GJ. NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. *BMC Bioinformatics* 2011; **12**: 317.
21 Bonetti P, Davoli T, Sironi C, Amati B, Pelicci PG, Colombo E. Nucleophosmin and its AML-associated mutant regulate c-Myc turnover through Fbw7 gamma. *J Cell Biol* 2008; **182**: 19–26.
22 Davis RJ, Welcker M, Clurman BE. Tumor suppression by the Fbw7 ubiquitin ligase: mechanisms and opportunities. *Cancer Cell* 2014; **26**: 455–464.
23 Welcker M, Orian A, Grim JE, Eisenman RN, Clurman BE. A nucleolar isoform of the Fbw7 ubiquitin ligase regulates c-Myc and cell size. *Curr Biol* 2004; **14**: 1852–1857.
24 Colombo E, Martinelli P, Zamponi R, Shing DC, Bonetti P, Luzi L *et al.* Delocalization and destabilization of the Arf tumour suppressor by the leukemia-associated NPM mutant. *Cancer Res* 2006; **66**: 3044–3050.
25 Bolli N, De Marco MF, Martelli MP, Bigerna B, Pucciarini A, Rossi R *et al.* A dose-dependent tug of war involving the NPM1 leukaemic mutant, nucleophosmin, and ARF. *Leukemia* 2009; **23**: 501–509.
26 Li YP. Protein B23 is an important human factor for the nucleolar localization of the human immunodeficiency virus protein Tat. *J Virol* 1997; **71**: 4098–4102.
27 Chun Y, Park B, Koh W, Lee S, Cheon Y, Kim R *et al.* New centromeric component CENP-W is an RNA-associated nuclear matrix protein that interacts with nucleophosmin/B23 protein. *J Biol Chem* 2011; **286**: 42758–42769.
28 Foltz DR, Jansen LE, Black BE, Bailey AO, Yates JR 3rd, Cleveland DW. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* 2006; **8**: 458–469.
29 Mitrea DM, Grace CR, Buljan M, Yun MK, Pytel NJ, Satumba J *et al.* Structural polymorphism in the N-terminal oligomerization domain of NPM1. *Proc Natl Acad Sci USA* 2014; **111**: 4466–4471.
30 Fantini D, Vascotto C, Marasco D, D'Ambrosio C, Romanello M, Vitagliano L *et al.* Critical lysine residues within the overlooked N-terminal domain of human APE1 regulate its biological functions. *Nucleic Acids Res* 2010; **38**: 8239–8256.
31 Scaloni F, Gianni S, Federici L, Falini B, Brunori M. Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. *FASEB J* 2009; **23**: 2360–2365.
32 Scaloni F, Federici L, Brunori M, Gianni S. Deciphering the folding transition state structure and denatured state properties of Nucleophosmin C-terminal domain. *Proc Natl Acad Sci USA* 2010; **107**: 5447–5452.
33 Scognamiglio PL, Di Natale C, Leone M, Cascella R, Cecchi C, Lirussi L *et al.* Destabilisation, aggregation, toxicity and cytosolic mislocalisation of nucleophosmin regions associated with acute myeloid leukemia. *Oncotarget* 2016; **7**: 59129–59143.
34 Di Natale C, Scognamiglio PL, Cascella R, Cecchi C, Russo A, Leone M *et al.* Nucleophosmin contains amyloidogenic regions that are able to form toxic aggregates under physiological conditions. *FASEB J* 2015; **29**: 3689–3701.
35 Chiarella S, De Cola A, Scaglione GL, Carletti E, Graziano V, Barcaroli D *et al.* Nucleophosmin mutations alter its nucleolar localization by impairing G-quadruplex binding at ribosomal DNA. *Nucleic Acids Res* 2013; **41**: 3228–3239.
36 Federici L, Arcovito A, Scaglione GL, Scaloni F, Lo Sterzo C, Di Matteo A *et al.* Nucleophosmin C-terminal leukemia-associated domain interacts with G-rich quadruplex forming DNA. *J Biol Chem* 2010; **285**: 37138–37149.
37 Gallo A, Lo Sterzo C, Mori M, Di Matteo A, Bertini I, Banci L *et al.* Structure of nucleophosmin DNA-binding domain and analysis of its complex with a G-quadruplex sequence from the *c-MYC* promoter. *J Biol Chem* 2012; **287**: 26539–26548.
38 Leong SM, Tan BX, Bte Ahmad B, Yan T, Chee LY, Ang ST *et al.* Mutant nucleophosmin deregulates cell death and myeloid differentiation through excessive caspase-6 and -8 inhibition. *Blood* 2010; **116**: 3286–3296.
39 Noguera NI, Song MS, Divona M, Catalano G, Calvo KL, García F *et al.* Nucleophosmin/B26 regulates PTEN through interaction with HAUSP in acute myeloid leukemia. *Leukemia* 2013; **27**: 1037–1043.

40 Roboz GJ. Current treatment of acute myeloid leukemia. *Curr Opin Oncol* 2012; **24**: 711–719.
41 Falini B, Gionfriddo I, Cecchetti F, Ballanti S, Pettirossi V, Martelli MP. Acute Myeloid Leukemia with mutated nucleophosmin (NPM1): any hope for a targeted therapy? *Blood Rev* 2011; **25**: 247–254.
42 Di Silvio E, Di Matteo A, Malatesta F, Travaglini-Allocatelli C. Recognition and binding of apocytochrome c to *P. aeruginosa* Ccml, a component of cytochrome c maturation machinery. *Biochim Biophys Acta* 2013; **1834**: 1554–1561.
43 Tien MZ, Sydykova DK, Meyer AG, Wilke CO. PeptideBuilder: a simple Python library to generate model peptides. *PeerJ* 2013; **1**: e80.
44 Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. *J Med Chem* 2006; **49**: 3315–3321.
45 Webb B, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol* 2014; **1137**: 1–15.
46 Bramucci E, Paiardini A, Bossa F, Pascarella S. PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. *BMC Bioinformatics* 2012; **13**: S2.
47 Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In Pullman P. (ed). *Intermolecular Forces*. Reidel: Dordrecht, The Netherlands, 1981, pp 331–342.
48 Darden T, York D, Pedersen L. Particle mesh Ewald - an N·log(N) method for Ewald sums in large systems. *J Chem Phys* 1993; **98**: 10089–10092.
49 Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 1997; **18**: 1463–1472.

Supplementary Information accompanies this paper on the Oncogenesis website (http://www.nature.com/oncsis)

Acknowledgement

I would like to express my gratitude to my advisors Prof. Carlo Travaglini-Allocatelli and Dr. Adele di Matteo for their valuable guidance, mentoring and continuous support.

My sincere thanks also go to Prof. Luca Federici and to my labmates - Sara, Mimma and Daniele.

I am grateful to the scientists, researchers and students, to all the amazing people that I met along my way, and in particular to those from “the first floor” of the Department of Biochemical Science “A. Rossi Fanelli”. It was great sharing this experience with all of you. You made it happen.

Finally, my deepest gratitude goes to my family and friends. This dissertation would not have been possible without your warm love, continued patience, and endless support. Words cannot express how grateful I am.