A manually curated network of the PML nuclear body interactome reveals an important role for PML-NBs in SUMOylation dynamics

Ellen Van Damme¹, Kris Laukens², Thanh Hai Dang², Xaveer Van Ostade¹*

¹. Laboratory of Protein Chemistry, Proteomics and Signal Transduction, Department of Biomedical Sciences, University of Antwerp (Campus Drie Eiken), Universiteitsplein 1 – Building T, 2610 Wilrijk, Belgium.
². Intelligent Systems Laboratory (ISLab) Department of Mathematics and Computer Science, University of Antwerp (Campus Middelheim), Middelheimlaan 1 – Building G, 2020 Antwerp, Belgium

Correspondence to: Prof. Dr. Xaveer Van Ostade, xaveer.vanostade@ua.ac.be, tel: 0032 (0)3 365 23 19

Received: 2009.10.20; Accepted: 2010.01.09; Published: 2010.01.12

Abstract

Promyelocytic Leukaemia Protein nuclear bodies (PML-NBs) are dynamic nuclear protein aggregates. To gain insight in PML-NB function, reductionist and high throughput techniques have been employed to identify PML-NB proteins. Here we present a manually curated network of the PML-NB interactome based on extensive literature review including database information. By compiling ‘the PML-ome’, we highlighted the presence of interactors in the Small Ubiquitin Like Modifier (SUMO) conjugation pathway. Additionally, we show an enrichment of SUMOylatable proteins in the PML-NBs through an in-house prediction algorithm. Therefore, based on the PML network, we hypothesize that PML-NBs may function as a nuclear SUMOylation hotspot.

Key words: PML-NB, SUMOylation, Cytoscape, protein-protein interaction, network

Introduction

Nuclear domain 10 (also called Kremer bodies or PODs) was first observed almost 50 years ago by electron microscopy as nuclear dense granular bodies[1, 2]. These punctate structures are now dubbed “PML nuclear bodies” (PML-NBs) after their scaffold protein Promyelocytic Leukaemia Protein (PML, also known as TRIM19 or MYL). The PML gene consists of nine exons and alternative splicing yields six nuclear PML isoforms (PML I-VI) and one smaller cytoplasmatic isofrom (PMLVII). However, of these seven isoforms there are several alternative spliced variants yielding at least 18 different forms of PML[3]. At the conserved N-terminus, PML contains a motif that defines the TRIM (Tripartite Motif) family of proteins. It is characterized by an RBCC motif composed of a conserved RING domain, one or two B-boxes, and an α-helical coiled-coil domain [4, 5]. The RING domain is a cysteine rich domain with zinc bound in a cross-brace conformation. This structure is ideal for the formation of larger protein aggregates and therefore crucial in processes such as cell growth, oncogenesis, apoptosis, and RNA trafficking during viral infections [5, 6]. The B-box provides the correct orientation and alignment of the coiled-coil domain [7]. When correctly positioned, the coiled-coil domain is responsible for homo and hetero-dimer interactions between TRIM proteins[3]. The RBCC motif of the PML protein makes it possible to sustain a highly dynamic turnover of partner proteins during the cell cycle [8, 9] which could explain the functionally promiscuous nature of the PML-NBs. PML is not only responsible for maintaining the integrity of the complex but also for the recruitment and correct localization of all PML-NB proteins, e.g. p53, Sp100, Daxx and CBP [4]. SUMOylation plays an important role in the scaffold function of PML. SUMO-3 attachment has
been reported to be responsible for the correct nuclear localization of PML [10], and without SUMO1 modification several protein partners are not recruited to the PML-NB [10-12]. However, unSUMOylatable mutants of PML (3KR) still show the typical nuclear speckle pattern [11] indicating that also other factors play a role in PML-NB formation. Recently, PML was identified as the first protein degraded by SUMO-dependent polyubiquitination after treatment with arsenic [12].

SUMOylation is a post-translational modification where one of the four isoforms of the Small Ubiquitin-like Modifier (SUMO) protein is conjugated to a target protein as a single SUMO molecule (SUMO1) or in polymeric chains (SUMO2/3)[13]. The enzymatic cascade is similar to the ubiquitin pathway[14] (Figure 1). However, unlike ubiquitination, which targets proteins for degradation, SUMOylation modifies proteins to interact in nuclear transport[10, 15], transcriptional regulation[14, 16], apoptosis[17, 18] and protein trafficking[19].

SUMO is synthesized as a precursor and its conjugation requires exposure of two C-terminal glycine residues. The processing of immature SUMO is one of the three functions of SUMO specific proteases (SENP). In addition, SENPs are also capable of dismantling SUMO chains and reversing SUMOylation by cleaving the isopeptide bonds between SUMO and its protein targets or between the several SUMOs[20]. Thus far, there are six known SENP isoforms (SENP1-3 and SENP5-7)[20] with different assignments in the cascade (Table 1).

### Table 1. The SENP isoforms show different affinity for the several SUMO isoforms in maturation processing and de-conjugation.

<table>
<thead>
<tr>
<th>SENP isoform</th>
<th>Maturation</th>
<th>De-conjugation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SENP1</td>
<td>SUMO1&gt;SUMO2&gt;&gt;SUMO3[21]</td>
<td>SUMO1, SUMO2[22, 23]</td>
</tr>
<tr>
<td>SENP2</td>
<td>SUMO2&gt;SUMO1, poor activity towards SUMO3[24, 25]</td>
<td>SUMO2&gt;SUMO1, poor activity towards SUMO3[24, 25]</td>
</tr>
<tr>
<td>SENP3</td>
<td>No processing activity</td>
<td>SUMO2, SUMO3[26, 27]</td>
</tr>
<tr>
<td>SENP5</td>
<td>SUMO3[26, 27]</td>
<td>SUMO1[24, 25]</td>
</tr>
<tr>
<td>SENP6</td>
<td>No processing activity[20, 26]</td>
<td>SUMO2, SUMO3[29]</td>
</tr>
<tr>
<td>SENP7</td>
<td>No processing activity[20, 26]</td>
<td>SUMO2, SUMO3&gt;&gt;&gt;SUMO1[28]</td>
</tr>
</tbody>
</table>
After maturation, SUMO is activated in an ATP dependent manner by thio-ester bond formation with the E1 activation proteins UBA2/AOS1[14]. The activating E1 enzyme Uba2p was first identified in yeast where it is, together with Aos1, necessary and sufficient for Smt3p (yeast homologue of SUMO) activation in vitro [30]. The human homologue hUBA2 was shown to act in the same way[31]. Subsequently SUMO is transferred from UBA2 to the SUMO conjugating enzyme UBC9 [14, 32]. UBC9 recognizes the consensus sequence ΨKXE where Ψ is a large hydrophobic amino acid and K is the lysine for SUMO conjugation[33]. UBC9 is believed to be the only SUMO conjugating enzyme[32, 34] and SUMOylation of its binding partners is enhanced by interaction with p14ARF[35]. Finally, an isopropyl bond between the terminal glycine of SUMO and the ε-aminogroup of lysine of the target protein is formed by an E3 ligase. Several unrelated proteins have been attributed E3 SUMO ligase activity: nucleoporin protein RANBP2[36], TOPORS[37, 38], several PIAS proteins[39-41], polycomb group protein Pc2[42], and RNF4[43]. Recently an anti-SUMO ubiquitin-protein isopeptide ligase (E3) has been described: mel-18 interacts both with HSF2 and the SUMO E2 UBC9 and inhibits UBC9’s activity which decreases HSF2 SUMOylation[44].

Because of the many associated proteins, PML-NBs are involved in a wide variety of cellular processes such as DNA damage and repair[45, 46], apoptosis[47-49], senescence[50, 51], cell cycle[52], transcription[53], carcinogenesis[54] and virus infection[55, 56]. Despite the involvement in such vital processes, PML deficient mice are normal, fertile and do not have a higher occurrence of tumors [57]. Although the actual function of the bodies still has to be determined, three non-mutually exclusive hypotheses about the function of PML-NBs have been proposed in literature. The first hypothesis gives PML-NBs a role as a nuclear depot thereby providing a mechanism for regulating nucleoplasmic levels of proteins in which PML bodies release proteins as they are needed[58]. Since nascent RNA has been found in the proximity of the PML-NBs[59] and PML-NBs seem to associate in regions with high transcriptional activity[60, 61], the second hypothesis proposes that PML-NBs are associated with nuclear activity such as gene transcription and DNA replication [62]. Indeed, most well-described partners of PML-NBs such as p53, Daxx, Mdm2 and Sp100, all have transcription activation or repression activity [4, 49, 63-66]. Furthermore, PML-NB structure is dependent on the integrity of the surrounding DNA. In fact, many factors which accumulate in PML-NBs are known histone methyltransferases, histone deacetylases or DNA methyltransferases, suggesting that PML-NBs may play an important role as epigenetic regulator of transcription and replication (for a review see[67] and references therein). The third hypothesis is that PML-NBs provide a nuclear platform for post-translational modification (PTM) such as phosphorylation [68, 69], acetylation[70] and SUMOylation[71].

These three hypotheses support the concept of dynamic interactions with PML and a high turnover of protein associates, instead of a static complex with a fixed number of partners. The involvement of PML-NBs in such a variety of processes has led to an enormous research interest in this enigmatic protein which has culminated in an overwhelming number of publications. In recent years, several techniques such as affinity capture, yeast two hybrid, co-localization and mass spectrometry have been employed to identify proteins of the PML complex which could lead to a better understanding of the PML-NB function. This reductionist approach has successfully identified several of the components of the PML-NBs but added more pieces to the puzzle instead of narrowing the spectrum of possible functions. Therefore an integrative “Systems Biology” approach could be essential to grasp the true complexity of the PML-NBs and to allow us to generate hypotheses based upon the vast amount of information available in distributed sources.

Not all PML partners described in literature are well represented in protein databases such as IntAct[72], Molecular Interactions Database (MINT)[73], BIOGRID[74], Human Protein Reference Database (HPRD)[75] and Nuclear Protein Database (NPD)[76]. This paper attempts to bridge the gap between (the lack of) database content and literature information. An integrated view of all the existing information is indispensable to gain true understanding of the functionality of large protein complexes such as PML-NBs.

We present a curated network model of PML-NBs based on information extracted from protein interaction databases (MINT, IntAct, BIOGRID, HPRD and NPD) and fragmented information obtained through extensive literature search. For the compilation of the network we used the open source network analysis software Cytoscape. It features high flexibility and the possibility to analyze protein networks with various (downloadable) plug-in software extensions allowing us to analyze and interpret the collected PML network. Based on this model we offer a system-wide perspective on the pathways and proteins that gather at the nuclear bodies. The network is
publicly available (Proteomics division at http://www.ua.ac.be/ppse) to make it a central information resource for “PML-omics” researchers.

Compiling a manually curated PML-NB network

A manually curated PML-NB interaction network was established in Cytoscape[77] by integration of PML interaction data retrieved from databases and literature (figure 2).

First we assembled a primary PML-NB network established with data retrieved from protein databases i.e. MINT[73], BIOGRID[74], HPRD[75], NPD[76] and IntAct[72]. Second, we expanded the primary network by adding PML-NB components which were exclusively found through extensive literature (only peer-reviewed journals) search. Finally, to create a biologically relevant network, interactions between PML-NB associated proteins were retrieved with the aid of Cytoscape import tools. Via the UniProtID we searched the IntAct and cPATH database[78] to further complete the network. After importing all available data from these databases the information was confirmed or completed by searching the BIOGRID, HPRD and MINT databases also using the UniProtID. Interactions found by cPATH and IntAct exploration which could not be verified in other databases were manually checked in literature.

To compile the network, PML was the only protein used as PML-NB determinant. Often Sp100 or Daxx are used as PML-NB determinants but these proteins also dynamically move in and out of the PML-NBs[59]. This makes PML the only protein which always defines the PML-NBs.

Interesting is the presence of a sub-network displaying Alt-associated-PML nuclear bodies (APBs) (figure 2, boxed). These special PML bodies are only seen in a subset of interfase nuclei in telomerase-negative immortalized human cells. These cells maintain their telomeres by a mechanism known as alternative lengthening of telomeres (ALT). Thus far, APBs have not been observed in mortal cells or telomerase positive immortal cells [79, 80]. In the APB, PML co-localizes with telomere binding proteins TRF1 and TRF2, Replication Protein A, RAD50, RAD51, RAD52, RAP1, Mre11, TIN2 and NBS1/p95[79, 81-84]. For RAD51 and RAD52 there is no evidence yet but all other components, including PML are required for APB formation [85]. The function of these bodies remains unclear. It is suggested that PML is required for the transcriptional activation of p53 in ALT cells and thus induces apoptosis via the p53 pathway[86].

After integrating the PML-NB components from database search and literature the network contains 166 nodes (proteins) and 781 edges (interactions). To provide an overview of the characteristics of the PML-NB components and interactions which connect these we provide several edge and node attributes (Table 2). The criteria used to annotate these attributes are standardized and the order in which several sources are used is strictly maintained.

Whereas in most publications only the most common PML-NB partners are mentioned, this is the first time PML-NB components are brought together in a comprehensive network. Depending on the researcher’s interest, the network can be analyzed from different angles, eventually using additional tools such as cytoscape plugins.

Analyzing the PML-NB interactome

We analyzed the general properties of the PML-NB components using the attributes described above in table 2. The network contains 166 protein associates of PML of which 70 are described in literature but are not available in any of the standard protein interaction databases.

The majority (57%) of the PML-NB proteins are nuclear, 33% of the PML-NB interactome shuttles actively between nucleus and cytoplasm, while 9% of the components are predominantly cytoplasmatic.

To provide an overview of the processes in which PML-NBs are involved, we categorized the UniProt function keywords in groups and distributed the 166 PML-NB partners in these groups. It is possible that one protein simultaneously belongs to several groups. Figure 3 shows the functional distribution. The representation of proteins involved in transcription regulation is remarkably high. As expected, most of these proteins appear to be nuclear or nucleus/cytoplasm shuttles. Another interesting and well represented group involves proteins which are part of the virus host interaction. Proteins belonging to this group are mostly shuttles between cytoplasm and nucleus. These shuttles emphasize once more the role of PML-NBs in viral infection[93]. In addition, PML nuclear bodies have been implicated in processes as apoptosis and cell cycle it is not surprising that many of the protein partners are directly involved in these processes. An interesting group is the one of the small protein conjugation, e.g. ubiquitin and SUMO, which is found under “PTM (small protein)” in the figure 3.
**Figure 2.** A manually curated PML-NB network. The central node is the scaffold protein PML. PML-NB components are connected to PML. Black edges represent protein components found in any of the databases (MINT, BIOGRID, IntAct, NPD, HRPD), pink edges were exclusively found in literature. We also represent the interactions between PML-NB components which are depicted as blue edges. For each node we provided the canonical name. Boxed components are those exclusively found in ALT cells.

**Table 2.** Overview of the different node and edge attributes in the PML-NB interaction network. In the ‘field’ column we present the attribute as it is assigned in the network. In the ‘description’ column we provide a brief explanation of what is embodied by the attribute. Finally, in the ‘criteria’ column the different sources of information are mentioned.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Criteria (in order)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Node Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>UniProtID [87]</td>
<td>Retrieved from UniProt</td>
</tr>
<tr>
<td>Canonical Name</td>
<td>Common name</td>
<td>Most used in literature &gt; First designation in UniProt</td>
</tr>
<tr>
<td>Cellular Localization</td>
<td>Subcellular localization differentiates amongst the several organelles in the nucleus and cytoplasm.</td>
<td>GO (cellular component ontology) &gt; Locate [88] (marked with **) &gt; Literature &gt; predicted localization by Locate</td>
</tr>
<tr>
<td>Simplified Cellular Localization</td>
<td>Only differentiates between nucleus, cytoplasm and shuttles between these two compartments.</td>
<td>‘nucleus’ (all nuclear organelles); ‘cytoplasm’ (all cytoplasmic organelles) and ‘shuttle’ between cytoplasm and nucleus.</td>
</tr>
<tr>
<td>Function</td>
<td>For this attribute we used the UniProt ontology as a simplified version of the GO terms to provide a comprehensible overview. To annotate the network with the complete GO annotations we refer to Cytoscape plugins such as BINGO[89].</td>
<td>UniProt Ontologies &gt; Panther database [90]</td>
</tr>
<tr>
<td>Connection with disease</td>
<td>This field is filled in when mutations, knockout or other modifications of the protein resulting in disease. If a MIM number was available it is mentioned, otherwise there is reference to specific publications. This attribute is additional information and does not mean that the PML-NBs are explicitly involved in the disease process.</td>
<td>UniProt (General Annotation) and literature</td>
</tr>
<tr>
<td>Induction</td>
<td>Information from UniProt and literature about regulation by several (bio)chemical stimuli.</td>
<td>UniProt (General Annotation) and literature</td>
</tr>
<tr>
<td>SUMO Consensus</td>
<td>We manually searched if any PML-NB components contained a predicted negatively charged amino acid-dependent sumoylation motif (NDSM)[91] or phosphorylation-dependent sumoylation motif (PDSM)[92] and present these in this attribute. Proteins predicted by us are labeled as ‘IS Lab’.</td>
<td>Literature</td>
</tr>
<tr>
<td>SUMO Isoform</td>
<td>Here we provide an overview of the conjugated isoform described in literature. Proteins marked in this attribute with ‘pathway’ are enzymes or actors of the SUMO conjugation machinery.</td>
<td>Literature</td>
</tr>
<tr>
<td>Remarks</td>
<td>In the ‘Remarks’ attribute we provide potentially important details about the protein</td>
<td>Literature and/or UniProt</td>
</tr>
<tr>
<td><strong>Edge Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author(s)</td>
<td>Reference to the first author, year of publication and PubMedID.</td>
<td>PubMed</td>
</tr>
<tr>
<td>Source</td>
<td>Indication of source where a certain interaction was found.</td>
<td>Literature and/or database (IntAct, HPRD, BIOGRID, MINT and NPD) search</td>
</tr>
<tr>
<td>Methods</td>
<td>Gives an overview of the methods used to detect the interaction. As a designation for these methods we used the same evidence code as used on BIOGRID.</td>
<td>Literature and/or database (IntAct, HPRD, BIOGRID, MINT and NPD) search</td>
</tr>
<tr>
<td>Fluorescence Microscopy</td>
<td>Gives details about the conditions for co-localization studies e.g. endogenous levels, overexpression data and or partial overlap (if mentioned in the publication)</td>
<td>Literature</td>
</tr>
<tr>
<td>Cell type</td>
<td>Cell type used for PML interaction studies.</td>
<td>Literature</td>
</tr>
<tr>
<td>PML isoform</td>
<td>For PML interactions it is important to make reference of the PML isoform used if this data is available.</td>
<td>Literature</td>
</tr>
<tr>
<td>Remarks</td>
<td>In the ‘Remarks’ attribute we provide potentially important details interaction</td>
<td>Literature</td>
</tr>
</tbody>
</table>
Figure 3. Functional distribution of the PML-NB components. For each functional category the number of nuclear, cytoplasmatic and shuttle proteins are plotted. The categories were based on the keywords in UniProt Ontologies. Several related categories were merged to facilitate interpretation, such as 'intracellular signaling' (merge of Notch signaling, Wnt signaling, signal transduction and intracellular signaling ontologies), 'differentiation and growth' (merge of growth, differentiation, muscle development, spermatogenesis and neurogenesis), 'PTM (chemical)' (acylation and phosphorylation), 'RNA processes' (RNA splicing, catabolism and processing), nuclear structure (chromatin and centrosome) and miscellaneous (lipid metabolism, oncogene and phagocytosis).

The information embedded in the edges illustrates the diversity of techniques underlying the integrated data that have been used to detect interactions with PML. Affinity capture, in vitro reconstitution of protein complexes and co-localization by immunofluorescence are the most commonly used methods to detect interactors of the PML-NBs. Further research will demonstrate if components found with these methods interact binary with PML. Less frequently used are the two-hybrid methodology or biochemical assays, two techniques which do in fact prove binary interactions with PML.

Of the total number of PML-NB components (166), 59% of the interactions were proven with only one technique. One must therefore keep in mind that false positives may be present in the network.

Experiments were performed in diverse cell types which range from commonly used immortal lines such as HeLa, MCF-7 and HEK293 to primary cell cultures. Using several cell lines gives an overall idea of the composition of the PML-NB which is relevant for integration in systems biology and contributes to the knowledge of PML–NB composition in different cell types. However, this generalization has to be treated with caution. The possibility exists that we adopt certain proteins as part of the complex while this architecture is unique for a certain cell type (e.g. ALT cells). Equally important to create a uniform view of the PML protein complex is the necessity to clearly state –if possible- which isoform was used. Since PML has 6 different splice variants with different kinetics [9], it is expected that these isoforms may (partially) bind other partner proteins. In only one third of the cases the PML isoform was clearly stated. Thus far the most used isoform is PMLIV. Therefore, to determine isoform-dependent differences in PML-NBs it is vital that researchers clearly state which PML variant is used in their experiments whenever possible.

Apart from the direct interactions with PML, we imported the links between the several PML partners from the IntAct database to show distinct biological pathways that meet at the PML-NBs. Hence, next to the direct edges (166) with PML, these 166 partners are interconnected by 615 edges (figure 2).

Upon analyses of the functions of the PML-NB components, a significant number of mostly nuclear proteins seems to be involved in ubiquitin-like modification processes such as SUMOylation and ubiquitination (figure 3, PTM (small protein)). Due to the
close relationship between PML SUMOylation and NB formation[10, 11, 94, 95] and because of PML-NBs may act as a platform for post-translational modifications; we explored the possibility if the PML-NBs could be SUMOylation hotspot in the nucleus.

Are PML-NBs SUMOylation hotspots in the nucleus?

PML-NBs are a hub for numerous proteins and are involved in a range of functions. Moreover, the complex is highly dynamic and proteins such as CBP and Sp100 show a dynamic exchange between nuclear bodies and nucleoplasm[96].

We wondered if the dynamics of the complex would be related to SUMOylation of NB proteins. Thus, the PML-NBs would be a (de)SUMOylation platform where SUMO is coupled or detached to/from its target protein. Possible evidence confirming this hypothesis lies in the observation that a large number of confirmed SUMOylated proteins localizes at the NBs. In fact, apart from SUMO4, all SUMO isoforms are found in the PML-NBs. Furthermore, not only several SENPs were co-localized with PML but also activating enzyme UBA2[97], the only SUMO conjugating enzyme UBC9[98, 99] and several SUMO ligases such as RANBP2 and PIASy [100-103] (Figure 4). One of these, RANBP2, promotes PML SUMOylation [104]. In addition, another PML-NB component, HDAC7, has been described to exert “SUMO E3 ligase-like” activity and also enhances SUMO modification of PML [105]. In fact, the most interesting SUMO E3 ligase at the PML-NBs could be PML itself. Indeed, in yeast, the RING domain of PML is sufficient for PML auto-SUMOylation and SUMOylation of other proteins [106]. Moreover, in transfected mammalian cells, PML RING mutants show a dramatical decrease in PML auto-SUMOylation [107]. In vitro, when purified PML is incubated with HeLa lysate, UBC9 and SUMO1, this leads to auto-SUMOylation [98]. PML was already described to mediate SUMOylation for proteins such as YAP1 [108]. However, one must keep in mind that, although all these observations suggest a SUMO E3 ligase activity for PML, indirect effects (e.g. stimulation or increased expression of enzymes from the SUMOylation machinery) may also be involved and further research is required to define PML as a novel SUMO E3 ligase.

Dynamics between SUMOylation and de-SUMOylation seems important for the architecture and function of the nuclear bodies. In fact, manipulations of SENPs show distinct effects on PML-NBs. It has been shown that IL6 cytokine induces the upregulation of mRNA expression of SENP1. Augmented SENP1 expression causes PML de-SUMOylation which abolishes the suppressor activity of PML on STAT3 [109]. On the other hand, suppression of SENP1 increases the number of SUMOylated proteins and the size of the PML-NBs[110]. SENP2 co-localizes with PML-NBs and PML enhances its effects. SENP2 hydrolyses SUMO1 from PML in vitro and changes the cellular localization of PML protein partners such as CBP, Daxx and p53[111]. Further on, truncated SENP5 co-localizes in the PML-NBs instead of in the nucleolus and both truncated- and wild type-SENP5 show the ability to remove SUMO2 or SUMO3 from PML Lys160 or Lys490[26]. Finally, SENP6 suppression results in the accumulation of SUMO2/3-modified species in PML bodies [29] but SENP6 has not been co-localized with PML yet. These observations suggest that not only PML is influenced by the presence of SENPs but that they also have an effect on other PML-NB targeted proteins.

All the necessary components to attach SUMO to a target are present at the PML-NBs but also the SENPs to remove SUMO again. This means that the complete cycle of SUMOylation and de-SUMOylation could take place at the PML-NBs (figure 4). In fact, PML contains a SUMO binding motif which forms a non-covalent interaction with SUMO attached to protein partners[107]. Possibly these proteins are recruited to make a functional switch at the PML-NBs by detachment of SUMO.

A large amount of PML partners have been described in literature as SUMOylated (38%). However, this percentage may well be an underestimation because SUMOylation of many PML partners has not been investigated yet. Therefore, to have an idea about other possible SUMOylation targets in the network we searched for PML-NB proteins with a negatively charged amino acid-dependent sumoylation motif (NDSM) motif as predicted by Yang et al.[91]. In fact, more than 56% of the PML-NB components contain an NDSM motif of which about half was already found to be SUMOylated in “wet lab” experiments.
Figure 4. Representation of the SUMOylation pathway in relation to the PML-NB. Indicated are the proteins which have been linked to PML (blue, linked with PML). SUMO is synthesized as an immature protein which needs processing by SENPs. SENP1, SENP2 and SENP5 exert processing activity and have been linked to the PML-NBs implying that maturation may take place at the NBs. After maturation, the SUMO protein is activated by two E2 enzymes of which UBA2 was co-localized with the PML-NBs. Also the only conjugating UBC9 can be found in the PML-NBs and molecules which influence UBC9 such as p14ARF and MEL-18 have been targeted to the PML-NBs. For final conjugation of SUMO to a target protein, an E3 ligase is needed and apart from PML which potentially contains E3 activity itself, four other ligases (TOPORS, RNF4, RANBP2, PIASy) have been targeted to the NBs. We propose that the target protein, the E3 ligase and UBC9-SUMO meet at the nuclear bodies for the completion of the SUMO pathway. Finally, SUMOylation can be reversed by SENPs. Only de-conjugation SENP3 and SENP7 have not been linked to the PML-NBs thus far.

In-house prediction of SUMOylation targets

Complementary to existing knowledge of SUMOylation sites, we employed an in-house developed machine learning approach to predict genome-wide SUMOylation sites.

We used the method based on Conditional Random Fields that previously proved successfully to predict kinase specific phosphorylation sites[112]. A 9 residue sequence window surrounding the annotated SUMOylation sites from UniProt was used to train a Conditional Random Field-based model as described in Dang et al. (2008)[112], with a feature function template optimized with a Genetic Algorithm as described in Dang et al. (2009) (in publication). Known SUMOylation sites were extracted from UniProt (downloaded august 18, 2009) based on the sequence feature "CROSSLINK" and a feature description containing the string "SUMO". Feature annotations containing the "probable" or "by similarity" qualifiers were removed. This data was used as training and test data.

The performance of the training step was assessed by k-fold (k=4) cross validations. In four fold cross validation, the protein set is partitioned into four subsets, of these subsets a single subset is retained as validation data for testing the model. The remaining three subsets are used as training data. This process is repeated four times with each of the four subsets used exactly once as the test set.

The resulting Receiver Operator (ROC) curve is shown in figure 5a. Its excellent performance is
demonstrated by an Area Under the Curve (AUC) value of 0.91. Also in figure 5 is a Venn-diagram of the overlap of what is described in literature and what is predicted. In general there is a large overlap between our screening method and the search for NDSM motifs[91]. We determine fewer proteins as SUMOylation target compared to the NDSM paper, which is determined by the stringency of our threshold. We chose to include only few insecurities and high confidence in our predictions. Figure 5b demonstrates that the overlap between literature and our approach and the overlap between the NDSM predictions and literature is equal, even though Yang et al. found more unconfirmed targets. Interestingly, there appears to be a portion of targets which were not predicted by our method or the method of Yang et al., suggesting that there are other motifs and surrounding amino acids which characterize a SUMOylation target.

Figure 5. A. Receiver Operator Curve (ROC) of our in-house prediction algorithm with Area Under the Curve (AUC) value of 0.91. B. Overlap between literature, our predictions and the predictions performed by Yang et al.. Due to the different approach there are some unique prediction data for both our prediction method and the predictions performed by Yang et al.. C. Predicted percentages by our in-house algorithm (FPR =1%). Genome-wide we predict about one fourth of the proteins to be SUMOylated. Membrane and extracellular proteins are less likely to be SUMOylation targets then cytoplasmatic or nuclear proteins. Within the PML-NBs we see a clear enrichment of SUMOylatable proteins.

http://www.biolsci.org
To assess a potential enrichment of SUMOylated proteins at the PML-NBs, percentages of predicted SUMOylated proteins were compared between the full human proteome and the PML-NB.

To accomplish this, the trained model was used to predict potential SUMOylation sites in the full human proteome (using the UniProt sequences), with a predefined maximum allowed false positive rate (FPR) of 1% as described in Dang et al.[112]. Figure 5c gives an overview of the predicted percentages of SUMOylatable proteins in several cellular compartments. Genome wide about 24% of the protein sequences contained at least one predicted potential SUMOylation site. Within proteins annotated with the GO Cellular component terms "cytoplasm" and "nucleus", respectively 26% and 33% were predicted to contain one or more SUMOylation sites. Within the assembled PML network, 48% contained 1 or more predicted SUMOylation motifs, demonstrating a clear overrepresentation of SUMOylation motifs in the PML network that cannot solely be attributed to its nuclear localization.

Since our predictor is trained with all possible known SUMOylation sites, its predictions are not focussing on one particular motif (such as the NDSM motif) but on all possible motifs. Together with the excellent performance demonstrated by the ROC curve and the stringent thresholds applied, we believe that this is the most reliable approach to be able to draw conclusions from the intracellular distribution of predicted SUMOylation sites. The observed discrepancies between our predictor and the NDSM results may, besides algorithmic performance differences, also reflect the diversity of the training sequences and the stringency of the performed prediction.

Conclusion

Dynamic complexes such as the PML-NBs require high-throughput methods as well as specialized reductionist approaches for characterization and identification of all the participating components. As much, the PML community should be cautious for scattered data making it difficult to maintain a general overview. We believe that our PML network helps in giving a comprehensive insight in the vast amount of PML data that has been generated.

Creating this overview focused the attention on a role of the PML-NBs in SUMOylation. The presence of the enzymes for both SUMOylation and de-SUMOylation suggests that both the attachment and detachment of SUMO can take place at the PML-NBs. In fact, the properties of the PML protein such as the RING domain and coiled coil motif are ideal to recruit unmodified proteins which are targets for SUMOylation. Amongst others, PML itself may act as an E3 SUMO ligase. In addition, PML contains a SUMO binding motif that can bind SUMOylated proteins which might be in turn de-SUMOylated.

To estimate if there was an enrichment of possibly SUMOylated proteins at the PML-NBs, independent of cellular localization, we used an in-house prediction method for a genome wide screen for SUMOylatable proteins. We show that 24% of the proteins genome wide is predicted to be SUMOylated. Although in the nucleus this number rises slightly to 33% we found that 48% of the PML-NB components contain one or more possible SUMOylation sites. This suggests that the PML-NBs are enriched sites for SUMOylated proteins.

Taken together the presence of (de)-SUMOylation enzymes combined with an enrichment in SUMOylatable proteins suggests that PML-NBs are a hotspot for nuclear SUMOylation where protein targets meet the components of the (de)-SUMOylation pathway.

The dynamic regulation of key point proteins at the PML-NBs supports this hypothesis since the regulatory function of SUMOylation on PML-NB related processes such as transcription [14, 113, 114], cell cycle[115], apoptosis[116], senescence[117] and viral infection [118-120] is slowly uncovered. As an example we present in table 3 the influence of SUMOylation on PML-NB related transcription factors.

Further, this hypothesis could be the glue between the three previously formulated hypotheses about PML-NB function. Since SUMOylation is often related to a functional switch, it is not unthinkable that by SUMOylation certain protein factors are retained in the nucleus until they are needed and are released by de-SUMOylation or vice versa. This SUMOylation dependent switching could meticulously regulate specific nuclear processes such as transcription and DNA related processes. In fact, our functional analysis reveals a very high number of transcription factors confirming the close relationship between PML-NBs and the modulation of transcription, possibly in a SUMO-dependent manner. Similarly, several protein partners are involved in DNA replication, damage sensing and repair which indicate a role in DNA related processes including epigenetic regulation [67].

The PML-NBs can thus be seen as a nuclear relay station where protein fractions involved in a variety of nuclear processes make a SUMOylation regulated functional switch. In addition, it is important to state that SUMOylation itself is regulated. Increasingly, the crosstalk between SUMOylation and other PTMs such
as acteylation and phosphorylation (e.g. via the PDSM SUMOylation motif [92]) is uncovered [121, 122]. Since at the PML-NBs also kinases, acetyltransferases and methyltransferases gather, it is likely that we are just exploring the borders of the regulatory complexity within this protein complex.

**Table 3. Effect of SUMOylation on transcription factors present in the PML-NBs.** Several transcription factors have a predicted PDSM motif but have not been shown to be SUMOylated (ARNT3A, Ataxin7, FLASH, HDAC2, HHEX, HIRA, LBP1, mTOR, MZF, NCOR1, NR4A1, PAR4, PAX5, SIRT1, SKI, SP140, THAP1 and ZNF24). About the remainder of the transcription factors associated with PML-NBs there is no available data regarding SUMOylation.

<table>
<thead>
<tr>
<th>UniProt ID</th>
<th>Name</th>
<th>Motif</th>
<th>Residue</th>
<th>Effect of SUMOylation</th>
<th>Positive/negative Effect (+/-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P27540</td>
<td>ARNT</td>
<td>NDSM</td>
<td>K245</td>
<td>SUMOylation affects the ability of ARNT to interact with cooperative molecules such as PML[123]</td>
<td>-</td>
</tr>
<tr>
<td>Q9BYV9</td>
<td>BACH2</td>
<td>NDSM</td>
<td></td>
<td>SUMOylation is an important regulatory system for the mobility of the nuclear domains formed by Bach2[124]</td>
<td>+</td>
</tr>
<tr>
<td>Q92793</td>
<td>CBP</td>
<td>NDSM</td>
<td></td>
<td>SUMO modification negatively modulates the transcriptional activity of CREB-binding protein[125]</td>
<td>-</td>
</tr>
<tr>
<td>P01100</td>
<td>C-FOS</td>
<td>NDSM</td>
<td>K299, K257</td>
<td>SUMOylation down-regulates c-Fos/c-Jun AP-1 dimer activity [126]</td>
<td>-</td>
</tr>
<tr>
<td>P05412</td>
<td>C-JUN</td>
<td>NDSM</td>
<td>K259, K257</td>
<td>SUMOylation down-regulates c-Fos/c-Jun AP-1 dimer activity [126]</td>
<td>-</td>
</tr>
<tr>
<td>P10242</td>
<td>C-MYB</td>
<td>N/PDSM</td>
<td>K499, K523</td>
<td>SUMO1-ylation regulates the transactivation function of c-Myb[127]</td>
<td>-</td>
</tr>
<tr>
<td>Q9UE72</td>
<td>Daxx</td>
<td>NDSM</td>
<td>K630, K631</td>
<td>Role of Daxx SUMOylation unknown[125, 128]</td>
<td>-</td>
</tr>
<tr>
<td>Q13547</td>
<td>HDAC1</td>
<td>NDSM</td>
<td>K444, K476</td>
<td>SUMOylation reduces HDAC1-mediated transcriptional repression[131]</td>
<td>-</td>
</tr>
<tr>
<td>Q86Z02</td>
<td>HIPK1</td>
<td>NDSM</td>
<td></td>
<td>SUMOylation induces a cytoplasmic translocation leading to ASK1-JNK activation[132]</td>
<td>+</td>
</tr>
<tr>
<td>Q9HZX6</td>
<td>HIPK2</td>
<td>NDSM</td>
<td>K25</td>
<td>SUMOylation inhibits HIPK2-induced JNK activation and p53-independent antiproliferative function[133]</td>
<td>-</td>
</tr>
<tr>
<td>Q03933</td>
<td>HSF2</td>
<td>NDSM</td>
<td>K82</td>
<td>SUMOylation of HSF2 results in conversion to the active DNA binding form[134]</td>
<td>+</td>
</tr>
<tr>
<td>Q9UJ22</td>
<td>LEF1</td>
<td>NDSM</td>
<td></td>
<td>SUMOylation induces potent repression of LEF1 activity[135]</td>
<td>-</td>
</tr>
<tr>
<td>Q99607</td>
<td>MEF/ELF4</td>
<td>NDSM</td>
<td>K657</td>
<td>SUMOylation down-regulates ELF4 activity[136]</td>
<td>-</td>
</tr>
<tr>
<td>Q15596</td>
<td>NCOA2</td>
<td>NDSM</td>
<td>K239, K731, K788</td>
<td>SUMOylation of GRIP1 (mouse equivalent of NCOA2) enhances binding to the androgen receptor[137]</td>
<td>+</td>
</tr>
<tr>
<td>Q9Y618</td>
<td>NCOR2</td>
<td>N/PDSM</td>
<td></td>
<td>SUMOylation of K684 is required for NFAT1 transcriptional activity, SUMOylation of K897 is only required for nuclear anchorage[136]</td>
<td>+</td>
</tr>
<tr>
<td>O95644</td>
<td>NFAT</td>
<td>NDSM</td>
<td>K684, K897</td>
<td>SUMOylation is associated with transcription repression[139]</td>
<td>-</td>
</tr>
<tr>
<td>Q00482</td>
<td>NR5A2</td>
<td>NDSM</td>
<td>K173, K289</td>
<td>SUMOylation of K684 is required for NFAT1 transcriptional activity, SUMOylation of K897 is only required for nuclear anchorage[136]</td>
<td>+</td>
</tr>
<tr>
<td>Q8N726</td>
<td>p14ARF</td>
<td>N/A</td>
<td></td>
<td>Not SUMOylated. Promotes SUMOylation by enhancing UBC9[35]</td>
<td>+</td>
</tr>
<tr>
<td>P04637</td>
<td>p53</td>
<td>NDSM</td>
<td>K386</td>
<td>SUMOylation induces premature senescence and stress response[116]</td>
<td>+</td>
</tr>
<tr>
<td>Q9H3D4</td>
<td>p63</td>
<td>NDSM</td>
<td>K637</td>
<td>SUMOylation has a negative effect on p63 driven transcription[140]</td>
<td>-</td>
</tr>
<tr>
<td>O15350</td>
<td>p73</td>
<td>NDSM</td>
<td>K627</td>
<td>SUMO1-ylation of p73 is involved its proteasome-dependent degradation, subnuclear localization and modulation of interactions with other SUMO1 substrates[116]</td>
<td>-</td>
</tr>
<tr>
<td>Q8N2W9</td>
<td>PIASy</td>
<td>NDSM</td>
<td>K35</td>
<td>SUMO1-ylation of PIASy is necessary for PIASy-dependent activation of Tcf-4[41]</td>
<td>+</td>
</tr>
<tr>
<td>Q05516</td>
<td>PLZF</td>
<td>NDSM</td>
<td>K242, K387, K396</td>
<td>SUMOylation enhances the transcriptional repression activity, correlating with a loss of PLZF-mediated growth suppression[141]</td>
<td>+</td>
</tr>
<tr>
<td>P29590</td>
<td>PML</td>
<td>NDSM</td>
<td>K65, K160, K490</td>
<td>SUMOylation is important for nuclear localization of PML and nuclear body formation[4]</td>
<td>+</td>
</tr>
<tr>
<td>P06400</td>
<td>pRB</td>
<td>NDSM</td>
<td></td>
<td>SUMOylation induces premature senescence and stress response[142]</td>
<td>+</td>
</tr>
<tr>
<td>P14373</td>
<td>RFP/TRIM27</td>
<td>NDSM</td>
<td></td>
<td>SUMOylation strengthens the transcriptional repressive activity of RFP[143]</td>
<td>+</td>
</tr>
<tr>
<td>P78317</td>
<td>RNF4</td>
<td>N/A</td>
<td></td>
<td>SUMOylation of RNF4 not shown, RNF4 is a SUMO-dependent Ubiquitin ligase for poly-SUMOylated proteins[12]</td>
<td>-</td>
</tr>
<tr>
<td>Q01826</td>
<td>SATB1</td>
<td>NDSM</td>
<td>K744</td>
<td>SUMOylation enhances caspase cleavage of SATB1[144]</td>
<td>-</td>
</tr>
<tr>
<td>Q965T3</td>
<td>SIN3A</td>
<td>NDSM</td>
<td>K744</td>
<td>SUMOylation regulates tumorsupression of SIN3a[38]</td>
<td>-</td>
</tr>
<tr>
<td>Q6FEW1</td>
<td>SIZN</td>
<td>NDSM</td>
<td></td>
<td>SUMOylation mediates localization in PML-NBs[145]</td>
<td>-</td>
</tr>
<tr>
<td>P84022</td>
<td>SMAD3</td>
<td>NDSM</td>
<td></td>
<td>SUMOylation affects the DNA-binding activity of Smad3 and induces nuclear export[146]</td>
<td>-</td>
</tr>
<tr>
<td>Q08407</td>
<td>SP1</td>
<td>NDSM</td>
<td>K16</td>
<td>SUMOylation inhibits cleavage of Sp1 N-terminal negative regulatory domain and inhibits Sp1-dependent transcription[147]</td>
<td>-</td>
</tr>
<tr>
<td>P11831</td>
<td>SRF</td>
<td>NDSM</td>
<td>K147</td>
<td>Activated SRF is suppressed by its SUMOylation[148]</td>
<td>-</td>
</tr>
<tr>
<td>Q15022</td>
<td>SUZ12</td>
<td>NDSM</td>
<td>K72, K73</td>
<td>SUMOylation modulates PRC2 repressive activity[129]</td>
<td>+/-</td>
</tr>
</tbody>
</table>

http://www.biolsci.org
Q9NQB0 TCF4 NDSM K75 SUMOylation is involved in beta-catenin-dependent and Tcf-4-mediated gene expression in the Wnt signaling pathway[41] +
O15164 TIFI/TRIM24 NDSM K690, K708 unknown
Q92993 TIP60 NDSM K430, K451 SUMOylation of TIP60 augments its acetyltransferase activity in vitro and in vivo[149] +
P13056 TR2 NDSM SUMOylated TR2 recruits corepressor RIP140 to act as a repressor for its target Oct4[150] -
O95551 TTRAP Unknown
Q9UBW7 ZNF198 NDSM K963 SUMOylation of ZNF198 is important for PML body formation[151] +
Q9Y4E5 ZNF451 NDSM K297 SUMOylation is involved in beta-catenin-dependent and transcription regulators between promyelocytic leukemia bodies and nucleoplasm[152] +

Acknowledgements

Research by EVD is funded by a Ph.D grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). KL and DTH are supported by the Flemish Institute supporting Scientific—Technological Research in industry (IWT).

We apologize to the authors whose work has been overseen in this publication. We welcome comments, corrections and additions to the network by any reader or PML-enthusiast. Please email your contributions to ellen.vandamme@ua.ac.be.

Conflict of Interest

The authors have declared that no conflict of interest exists.

References


53. Lai HK, Borden KL. The promyelocytic leukemia (PML) protein suppresses cyclin D1 protein production by altering the nuclear cytoplasmic distribution of cyclin D1 mRNA. Oncogene 2000 Mar 23;19(13):1623-34.


141. Chao TT, Chang CC, Shih HM. SUMO modification modulates the transrepression activity of PLZF. Biochem Biophys Res Commun 2007 Jun 29;358(2):475-82.


