

A compound C-terminal nuclear localization signal of human SA2 stromalin

Leszek J. Tarnowski^a, Michal Milewski^b, Jan Fronk^c, Anna Kurlandzka^a

^aInstitute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

^bDepartment of Medical Genetics, Institute of Mother and Child, Warsaw, Poland

^cInstitute of Biochemistry, Faculty of Biology, University of Warsaw, Warsaw,
Poland

Corresponding author: **Anna Kurlandzka**

Address: Institute of Biochemistry and Biophysics, Polish Academy of Sciences,
Pawińskiego 5A, 02-106 Warsaw, Poland

e-mail: ania218@ibb.waw.pl

phone: (48) 22 592 1318

fax: (48) 22 592 2190

Running title: Nuclear localization signal of SA2 stromalin

Keywords: bipartite NLS, nuclear import, SA2 stromalin

Abbreviations: DAPI, 4',6-diamidino-2-phenylindole dihydrochloride; GFP, green fluorescent protein; LMB, leptomycin B; NLS, nuclear localization signal; NES, nuclear export signal

Financial support: This work was supported by the Ministry of Science and Higher Education [grant N N303 456838 to A.K.].

ABSTRACT

Stromalins are evolutionarily conserved multifunctional proteins with the best known function in sister chromatid cohesion. Human SA2 stromalin, likely involved in the establishment of cohesion, contains numerous potential nuclear localization (NLS) and nuclear export signals (NES). Previously we have found that the C-terminus of SA2 contains NLS(s) functional in human cells. However, the identity of this signal remained unclear since three NLS-like sequences are present in that region. Here we analyzed the functionality of these putative signals by expressing GFP-tagged C-terminal part of SA2 or its fragments in a human cell line and in the yeast *Saccharomyces cerevisiae*. We found that in human cells the nuclear import is dependent on a unique compound di- or tripartite signal containing unusually long linkers between clusters of basic amino acids. Upon expression of the same SA2 fragment in yeast this signal is also functional and can be easily studied in more detail.

INTRODUCTION

The SA2 protein belongs to the family of stromalins, multifunctional proteins with the best known function in sister chromatid cohesion. The cohesion is brought about by ring-shaped complexes composed of three subunits: Smc1, Smc3, and a kleisin (Rad21/Scc1 in mitosis and Rec8 in meiosis). The SA/Scc3 protein binds to the central domain of Scc1 and initially was included in the complex (Haering *et al.* 2002, Nasmyth & Haering, 2009). However, it has been demonstrated recently that the SA ortholog of *Schizosaccharomyces pombe* (Psc3^{Scc3}) is involved in the initial steps of the establishment of cohesion (Murayama & Uhlmann, 2014, Marston, 2014). In yeast and human cells Scc3/SA binds two proteins that are not considered part of the cohesin core complex: Wpl1/Rad61 and Pds5. The Scc3-Wpl1-Pds5 sub-complex acts as a regulator of cohesin activity (Ghandi *et al.*, 2006; Sutani *et al.*, 2009).

In human somatic cells the cohesin complexes associate with either of the two SA proteins, SA1 or SA2 (Losada *et al.*, 2000; Sumara *et al.*, 2000). SA1 and SA2 are clearly homologous, but their individual roles and functional domains have not been fully characterized.

We have recently shown that the both human SA proteins contain numerous potential nuclear localization (NLS) and nuclear export signals (NES), and that some of them are functional. One of the orthologs, SA2, but not SA1, can be exported from the nucleus using a Crm1-dependent route. We have also shown that in HeLa cells the functional NLSs of SA1 and SA2 are localized to different regions of the protein, at the N-terminus in SA1 and at the C-terminus in SA2 (Tarnowski *et al.*, 2012).

Here we present further studies on the functionality of three putative NLSs localized in the C-terminal part of SA2. Notably, interference with any of those NLSs destabilized the SA2 protein and only its C-terminal part could be studied. In human cells two out of the three potential NLSs functioned autonomously, but were significantly weaker than all three combined. They are separated by long stretches of 44 and 63 amino acids and we propose that in fact they together constitute a compound, bipartite or tripartite NLS. The third signal alone had hardly any nuclear localization potential. We also established conditions in which these signals could be studied in the full-length protein expressed in *Saccharomyces cerevisiae*. This required removal of an N-terminal NLS and inactivation of Crm1-dependent nuclear export. Upon heterologous expression in *S. cerevisiae* the compound C-terminal NLS of SA2 was functional as well.

MATERIALS AND METHODS

Strains and growth conditions. *Escherichia coli* XL1-Blue MRF' (Stratagene) was used for molecular manipulations. *S. cerevisiae* strains MNY8 (*CRM1-T539C*), with an amino acid

substitution producing a leptomycin B (LMB)-sensitive version of Crm1, and the control MNY7 (*CRM1*) [Neville & Rosbash, 1999] were provided by Dr. Michael Rosbash, Brandeis University, Waltham, MA, USA. Yeast growth and transformation followed standard procedures (Sherman, 1991). LMB treatment was exactly as described before (Tarnowski *et al.*, 2012).

HeLa cell culture. HeLa cells (European Cell Culture Collection, catalogue no. 93021013) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). Cells were incubated in polystyrene flasks (Sarstedt) in 5% CO₂-balanced air at 37°C. All cell culture reagents were from Gibco/Life Technologies.

Construction of plasmids encoding GFP fusion proteins.

cDNA serving as template for amplification of SA2 variant 4 (EMBL/GenBank Accession Number NM_006603.3 or NM_006603.4) was purchased from OriGene. Nucleotide sequences encoding SA2 fragments (C-terminal 161 amino acids or shorter sub-fragments, as detailed further) were PCR-amplified and fused by their 3' terminus to sequences encoding 2xGFP, through a sequence encoding a dipeptide (EF) linker, and were expressed in HeLa or yeast cells. For expression in HeLa cells the PCR-generated DNA fragments encoding parts of SA2 were subcloned into the plasmid pCl-neo. Expression in yeast was from the centromeric vector pUG35, as described in (Tarnowski *et al.*, 2012). For expression of full-length SA2-GFP in yeast the protein was modified by deleting the N-terminal NLS K32-K47 as described in (Tarnowski *et al.*, 2012); the resulting protein was named SA2ΔK32-K47-GFP. DNA sequences of all primers are available on request.

Fluorescence microscopy. Yeast cells were observed and images were taken using a Nikon Eclipse E800 fluorescence microscope with a 100x objective. GFP-fusion proteins were visualized in liquid-grown cells fixed with 4% formaldehyde for 20 min. DAPI was used to stain DNA. To estimate the percentage of yeast cells with a given SA2-GFP localization, at least 100 cells were analyzed. For the analysis of the distribution of GFP-fused proteins in HeLa cells an IX71 Olympus fluorescence microscope was used. For each protein variant tested, at least 100 GFP-expressing cells were analyzed from two independent transfection experiments. For quantitation of GFP signal distribution between nucleus (N) and cytoplasm (C), average pixel intensity was calculated using CellF (Olympus) program for 5- μ m diameter circles representing a randomly picked region of either compartment in 6 cells per experimental variant. Background pixel intensity was subtracted and the resulting values were divided (N/C) to give the signal strength ratio. Unpaired Student's t test was used to evaluate significance of differences versus the value calculated for GFP protein alone or for the C-terminus-GFP fusion. $p < 0.01$ was considered significant and $p < 0.001$ highly significant.

Western blotting. To visualize GFP-tagged proteins on Western blots, yeast cells were grown o/n, total protein was extracted and protein samples (100 μ g/lane) were subjected to 8% SDS-PAGE and electrotransferred onto Hybond-C extra membrane, the membrane was probed with an anti-GFP antibody (A.v. Peptide antibody Living Colors AB, Becton Dickinson) followed by anti-rabbit alkaline phosphatase-conjugated secondary antibody (Promega), and developed with Western Blue Stabilized Substrate for Alkaline Phosphatase (Promega).

RESULTS

Characterization of C-terminal nuclear localization signals (NLSs) of SA2

The C-terminal fragment of human stromalin SA2 (variant 4, 1231 amino acids, EMBL/GenBank Accession Number NM_006603.3 or NM_006603.4) contains three putative nuclear localization signals at positions R1071-V1084, P1129-S1135 and P1199-E1206 (Tarnowski *et al.*, 2012). They are shown in Fig. 1 and for convenience we will refer to them in this paper as NLS1, NLS2 and NLS3. Our earlier studies showed that a lack of this region precludes the nuclear localization of SA2-GFP in HeLa cells.

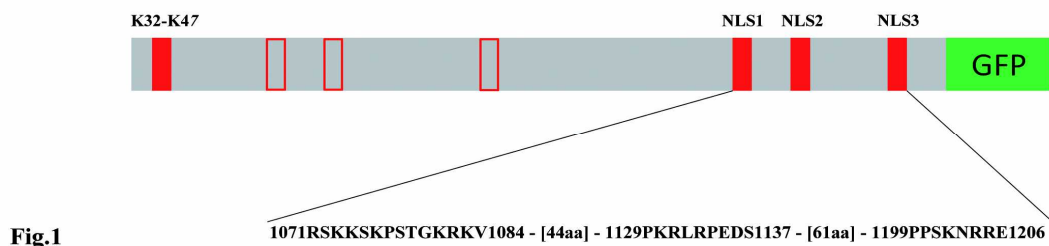


Figure 1. C-terminus of SA2 contains three putative NLSs. Scheme of the whole SA2 protein containing 7 putative NLS. Open red boxes: NLS not discussed in detail in the present paper, solid red boxes: NLSs studied here. Amino acid sequences of the C-terminal NLSs and lengths of linkers between them are shown. The NLSs were identified using The Eukaryotic Linear Motif Resource (ELM) (<http://elm.eu.org>) (Gould *et al.*, 2009) and PSORTII (<http://psort.ims.u-tokyo.ac.jp/>) (Horton & Nakai, 1997) servers. The drawing is not to scale.

To characterize the role of each potential NLS individually or in combination we attempted to construct a set of full-length SA2-GFP fusion proteins with deletions of or substitutions in individual NLS sequences. However, an intact C-terminal region turned out to be critical for the SA2 stability because deletions of individual NLS or alanine substitutions in their sequences often produced no GFP fluorescence upon transient expression of the constructs in HeLa cells (not shown). We therefore had to trim the SA2 protein from its N-

terminus so that only the C-terminus or individual signals fused to 2xGFP (further referred to as “GFP”) were expressed.

The whole C-terminal part of SA2 comprising 161 amino acids fused with GFP localized exclusively to the nucleus (Fig. 2). However, short fragments containing only one NLS fused to GFP behaved differently: NLS1 and NLS2 accumulated in both the cytoplasm and nucleus, with a slight but significant preference for the nucleus, whereas the fluorescent signal coming from NLS3-GFP had a similar distribution to the signal produced by GFP alone. This indicates that NLS1 and NLS2 individually are weak nuclear localization signals and NLS3 does not function autonomously. To test directly whether NLS1 and NLS2 constitute a bipartite signal we fused a fragment spanning only these signals separated by the native 44-amino acid linker with GFP. Unfortunately, that construct did not produce fluorescence in HeLa cells (not shown). A fragment comprising NLS2 and NLS3 separated by the native 63-amino acid linker behaved exactly as NLS2 alone, i.e., directed GFP to both the nucleus and the cytoplasm (Fig. 2).

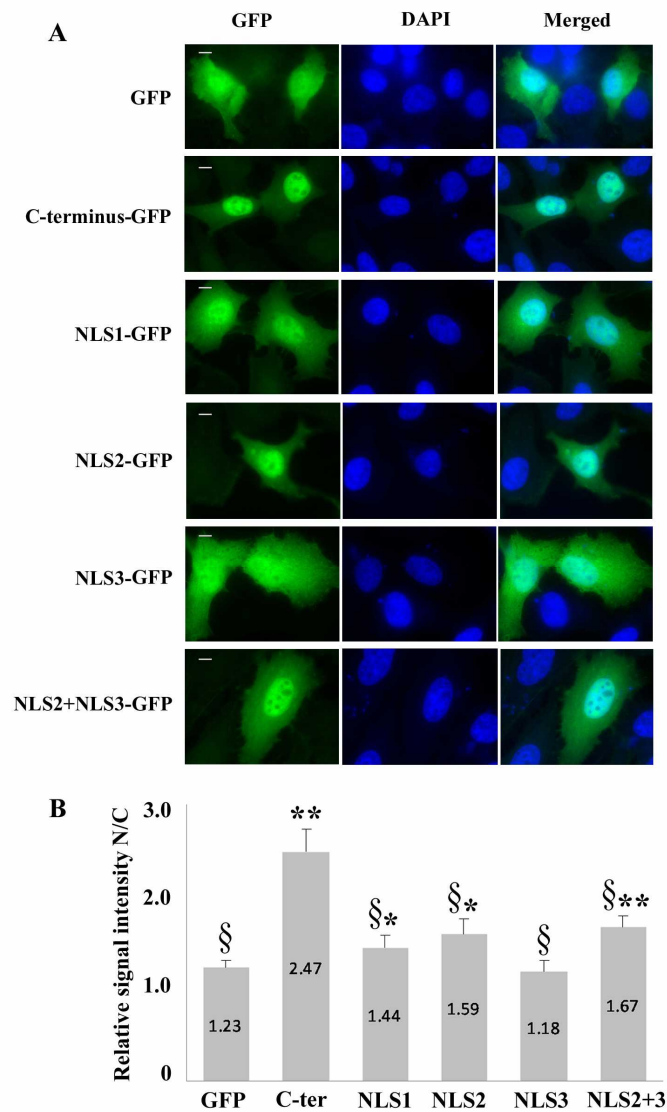


Figure 2. C-terminus of SA2 targets GFP to the nucleus in HeLa cells. **A.** Consecutive rows show HeLa cells expressing, from the top: 2xGFP reporter protein alone as a control; and 2xGFP preceded by: C-terminal 161 amino acids of SA2, NLS1, NLS2, NLS3 alone, and NLS2 + NLS3. Cells were analyzed 24 hours post transfection. Transfection efficiency was 50-80% for the control plasmid encoding GFP, and for other plasmids - ca. 5%. Bar, 10 μ m. **B.** GFP fluorescence distribution between nucleus and cytoplasm. Signal intensity ratio was determined as described in Materials and Methods and is shown as average \pm SD for 6 cells per experimental variant. *, significantly different from GFP control ($p < 0.01$); **, highly significantly different from GFP control ($p < 0.001$); §, highly significantly different from C-terminus ($p < 0.001$).

Since the NLS1-GFP and NLS2-GFP fusion proteins were only partially nuclear and NLS3-GFP localized the same as the control GFP, one can safely infer that none NLS alone was as efficient as the whole C-terminal region in targeting to the nucleus. This suggests that these NLSs act in synergy or may constitute a bipartite (NLS1 + NLS2) or tripartite NLS.

C-terminal nuclear localization signals of SA2 can be studied in *S. cerevisiae*

Since full-length SA2 protein with its C-terminal NLSs compromised could not be expressed in HeLa cells we tried to use *S. cerevisiae* yeast to overcome this drawback. We turned to this simple model organism since we had previously used it successfully to study the nucleocytoplasmic transport of human SA2 (Tarnowski *et al.*, 2012). In *S. cerevisiae* full-length SA2 was imported to the nucleus using the N-terminal NLS K32-K47 (Fig. 1 shows position of this NLS in SA2). When that NLS was missing, SA2 was still imported, but that import was masked by the simultaneous Crm1-dependent export from the nucleus. Thus, the C-terminal NLSs can only be investigated when NLS K32-K47 is removed and the Crm1-dependent nuclear export blocked. This can be achieved either by inactivating the nuclear export signal (NES) L1022-L1033 or by inhibiting the Crm1 exportin with leptomycin B (LMB) in an LMB-sensitive strain. The shift of SA2 Δ K32-K47-GFP to the nucleus upon LMB treatment (Fig. 3, upper panel) confirmed that NLSs other than K32-K47 are functional in *S. cerevisiae*.

To confirm that the secondary NLSs functioning in *S. cerevisiae* are indeed localized to the C-terminus, here we subjected to LMB treatment a strain expressing GFP-tagged SA2 Δ K32-K47 Δ C161 (devoid of the C-terminus). As expected, the fusion protein did not relocate to the nucleus (Fig. 3, middle panel), which proved that there are no functional NLS in the SA2 protein except for the K32-K47 one and those in the C-terminal region.

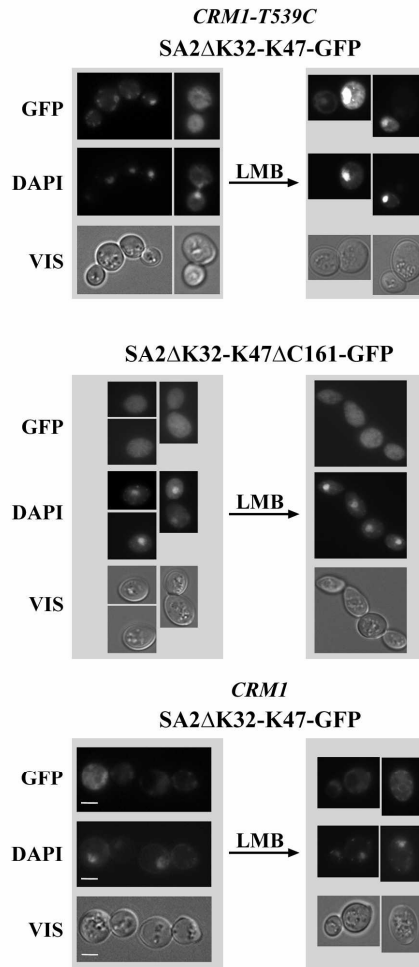


Fig. 3

Figure 3. Functionality of C-terminal NLSs of SA2 in *S. cerevisiae*. Subcellular localization of SA2-GFP variants was analyzed after addition of LMB (Crm1 inhibitor) to 40 ng/ml to cells in logarithmic phase of growth. Strain *CRMI-T539C* bears an LMB-sensitive version of Crm1. Upper panel: cells of *CRMI-T539C* strain expressing SA2ΔK32-K47-GFP protein. Upon LMB treatment the fusion protein is retained in the nucleus. Middle panel: cells of the same strain expressing SA2ΔK32-K47ΔC161-GFP. LMB treatment does not affect the localization of the fusion protein throughout the cell. Lower panel: cells of control *CRMI* strain expressing SA2ΔK32-K47-GFP. Since wild-type Crm1 is insensitive to LMB, no nuclear accumulation of the fusion protein is observed upon LMB addition. Ca. 100 cells were observed for each experimental condition. SA2ΔK32-K47-GFP localized predominantly to the nucleus in ca. 80% of cells of *CRMI-T539C* strain following LMB treatment whereas

no such localization was observed for SA2 Δ K32-K47 Δ C161-GFP. In the control *CRMI* strain treated with LMB SA2 Δ K32-K47-GFP localized to the cytoplasm in 100% of cells. GFP represents fluorescence of fusion proteins, DNA was stained with DAPI, VIS – transmitted light. Cells viewed at 1000 x magnification, bar, 3.5 μ m.

To study these signals individually, we tested in yeast exactly the same fragments of SA2 fused with GFP as we did in HeLa cells. Results presented in Fig. 4 A show that the C-terminus of SA2 containing all three NLSs localizes exclusively to the nucleus, as it did in HeLa cells. However, except for a very weak activity of NLS1, individual NLSs were insufficient to mediate the nuclear import of the GFP reporter. All fusion proteins were expressed intact (Fig. 4 B), although NLS3-GFP was present in two forms, likely indicating partial proteolysis. We next constructed fusion proteins with NLS pairs separated by their native linkers: NLS1 and NLS2, and NLS2 and NLS3 fused with GFP, but the GFP fluorescence was not enriched in the nucleus and was similar to that produced by NLS2-GFP (not shown). These observations demonstrate that in yeast cells the C-terminal fragment is effectively imported only when all three NLSs are present.

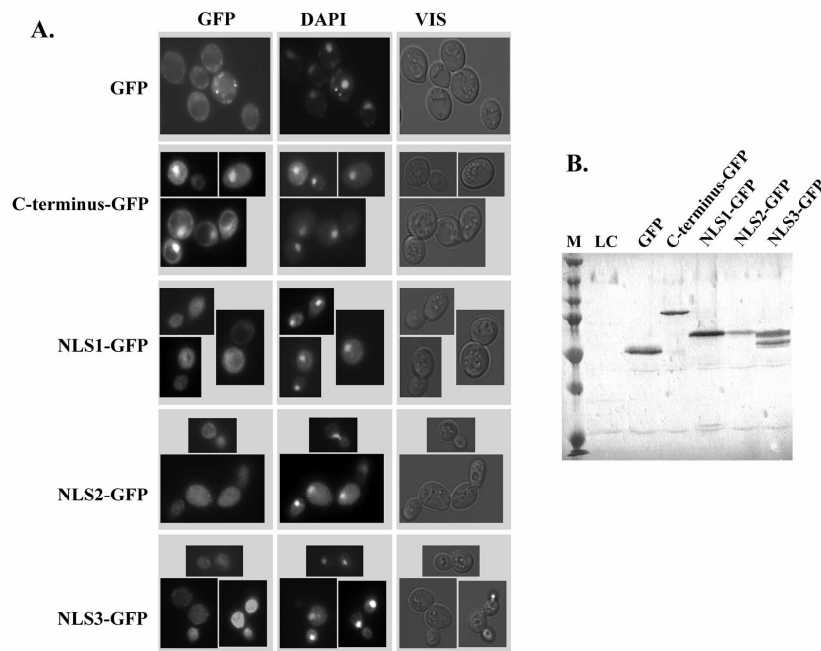


Figure 4. C-terminus of SA2 targets GFP to the nucleus in *S. cerevisiae* cells. **A.** Consecutive rows show yeast cells expressing, from the top: 2xGFP reporter protein alone as a control; and 2xGFP preceded by: C-terminal 161 amino acids, NLS1, NLS2, and NLS3. In yeast cells individual NLS do not function autonomously. Ca. 100 cells were observed for each experimental condition. GFP represents fluorescence of fusion proteins, DNA was stained with DAPI, VIS – transmitted light. **B.** All fusion proteins comprised of fragments of SA2 and 2xGFP expressed in yeast have predicted molecular weights. Yeast strain MNY8 was transformed with centromeric plasmid pUG35 bearing hybrid genes encoding fragments of SA2 protein as indicated. Cells were grown to middle-exponential phase in selective medium. Immunoblots of whole-cell extracts were probed with anti-GFP antibodies. Aliquots of 100 μ g of protein/lane were resolved by SDS/8% PAGE. Lane M – marker, LC: loading control, yeast strain without a plasmid, lane 2xGFP: same strain bearing pUG35-2xGFP, lane C-terminus-2xGFP: same strain with plasmid bearing 161 C-terminal amino acids fused to double GFP, lanes NLS1-, NLS2-, NLS3-2xGFP, as before, but plasmids bearing sequences encoding respective NLSs.

Altogether, here we have shown that the compound C-terminal NLS of human SA2 is functional in yeast, although the strength of individual signals differs slightly between these two systems. Thus, full-length SA2 protein with individual C-terminal NLS compromised can

be studied in *S. cerevisiae* to decipher details of the SA2 nuclear import. This heterologous system is therefore a valuable complement to the human cell model, in which interference with the individual C-terminal NLS in an otherwise intact SA2 lead to its instability.

DISCUSSION

A classical NLS is a single stretch of four amino acids of which three are basic or a single such amino acid separated by a linker from a cluster of three to five basic ones. The length of the linker was initially defined as 10-12 amino acids (Kalderon *et al.*, 1984 a, b). However, some linkers longer than those have been found in some functional NLS (e.g. Lange *et al.* 2010). The linker function is dependent on its sequence and possibly also its flexibility. It is speculated that longer linkers could provide greater opportunity for subtle regulation through post-translational modifications and by allowing easy access of small molecules modulating the NLS activity (Lange *et al.*, 2010; Fontes *et al.*, 2000; Marfori *et al.*, 2011). Data indicating serine phosphorylation of SA2 in a region close to NLS2 preceding SA2 dissociation from chromatin (Hauf *et al.*, 2005) support this concept.

In humans SA2 is linked to the development and progression of several types of cancer such as bladder cancer and melanoma (for review: Losada, 2014). Thus, identifying details and functional significance of its nuclear import seems important. Both our present and others' (Kong *et al.* 2014) studies concentrated on the import of the C-terminal fragment of SA2. This restriction results from the fact that detailed studies of this region using full-length SA2 are impossible since most plasmids encoding proteins with deletions of or mutations in individual NLS do not produce fluorescence of GFP reporter, likely due to the rapid degradation of the fusion proteins, while NLS2 fused to GFP and full-length SA2 manipulated in this region are toxic to cells. The isolated C-terminal region does not show these drawbacks. Recent data of Kong *et al.* (2014), who found that mutations in NLS1 or 2 abolish

the nuclear localization of the C-terminus of SA2 in HeLa cells were obtained independently using almost exactly the same fragment of SA2 that was used in our study. Our findings are complementary to their report and suggest that nuclear import of SA2 is dependent on a compound signal comprising short clusters of basic amino acids separated by long linkers. While individually two of those basic clusters (NLS1 and NLS2) do show some nuclear targeting capacity and the third (NLS3) very little if any, we propose that in fact all three are required for fully efficient nuclear import. In yeast, the bipartite signal comprising NLS1, linker of 44 amino acids and NLS2 is insufficient for nuclear import and all three NLSs are required; in mammalian cells the situation is probably the same, but one cannot show directly that NLS1 + NLS2 are ineffective since such a construct is unstable.

Using HeLa and yeast cells we have found that nuclear import of human SA2 requires the compound NLS located in its C-terminus. However, using short peptides for the analysis can be misleading, since they might only function in a properly folded (ideally, full-length) protein. Here, we have demonstrated that the C-terminal NLS can be investigated in a subtly modified full-length SA2 upon heterologous expression in *S. cerevisiae* and this system can be used for further, more detailed analyses of the complex nuclear import of SA2.

Acknowledgments

We thank Mr. Maciej Zaczek for assistance in some *S. cerevisiae* experiments.

References

Fontes MR, Teh T, Kobe B (2000) Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. *J Mol Biol* **297**: 1183–1194.

Gandhi R, Gillespie PJ, Hirano T (2006) Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr Biol* **16**: 2406-2417.

Gould CM, Diella F, Via A, Puntervoll P, Gemund C, *et al.* (2009) ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res* **38**: (Database issue) D167–D180.

Horton P, Nakai K (1997) Better prediction of protein cellular localization sites with the nearest neighbors classifier. *Proc Int Conf Intell Syst Mol Biol* **5**: 147–152.

Haering CH, Lowe J, Hochwagen A, Nasmyth K (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**: 773–788.

Hauf S, Roitinger E, Koch B, Dittrich CM, Mechtler K, Peters JM (2005) Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol* **3**: e69.

Kalderon D, Richardson WD, Markham AF, Smith AE (1984a) Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311**: 33–38.

Kalderon D, Roberts BL, Richardson WD, Smith AE (1984b) A short amino acid sequence able to specify nuclear location. *Cell* **39**: 499–509.

Kong X, Ball AR Jr, Pham HX, Zeng W, Chen HY, Schmiesing JA, Kim JS, Berns M, Yokomori K (2014) Distinct functions of human cohesin-SA1 and cohesin-SA2 in double-strand break repair. *Mol Cell Biol* **34**: 685-698.

Lange A, McLane LM, Mills RE, Devine SE, Corbett AH (2010) Expanding the definition of the classical bipartite nuclear localization signal. *Traffic* **11**: 311-323.

Losada A, Yokochi T, Kobayashi R, Hirano T (2000) Identification and characterization of SA/Scp3p subunits in the *Xenopus* and human cohesin complexes. *J Cell Biol* **150**: 405–416.

- Losada A (2014) Cohesin in cancer: chromosome segregation and beyond. *Nat Rev Cancer* **14**: 389- 393.
- Marfori M, Mynott A, Ellis JJ, Mehdi AM, Saunders NF, Curmi PM, Forwood JK, Bodén M, Kobe B (2011) Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim Biophys Acta* **1813**: 1562-1577.
- Marston AL (2014) Chromosome segregation in budding yeast: sister chromatid cohesion and related mechanisms. *Genetics* **196**: 31-63.
- Murayama Y, Uhlmann F (2014) Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**: 367-371.
- Nasmyth K, Haering CH (2009) Cohesin: its roles and mechanisms. *Annu Rev Genet* **43**: 525–558.
- Neville M, Rosbash M (1999) The NES–Crm1 export pathway is not a major mRNA export route in *Saccharomyces cerevisiae*. *EMBO J* **18**: 3746–3756.
- Sherman F (1991) Getting started with yeast. *Methods Enzymol* **194**: 3–21.
- Sumara I, Vorlaufer E, Gieffers C, Peters BH, Peters J-M (2000) Characterization of vertebrate cohesin complexes and their regulation in prophase. *J Cell Biol* **151**:749–761.
- Sutani T, Kawaguchi T, Kanno R, Itoh T, Shirahige K (2009) Budding yeast Wpl1(Rad61)-Pds5 complex counteracts sister chromatid cohesion-establishing reaction. *Curr Biol* **19**: 492-7.
- Tarnowski LJ, Kowalec P, Milewski M, Jurek M, Plochocka D, Fronk J, Kurlandzka A (2012) Nuclear import and export signals of human cohesins SA1/STAG1 and SA2/STAG2 expressed in *Saccharomyces cerevisiae*. *PLoS One* **7**: e38740.