SHORT COMMUNICATION

A toolbox of endogenous and heterologous nuclear localization sequences for the methylotrophic yeast Pichia pastoris

Astrid Weninger¹, Anton Glieder¹ and Thomas Vogl¹,²,*

¹Institute for Molecular Biotechnology, Graz University of Technology, Petersgasse 14/2, 8010 Graz, Austria and ²Queensland University of Technology, 2 George St., Brisbane QLD 4000, Australia

*Corresponding author: Queensland University of Technology, 2 George St., R Block Level 4, Brisbane, QLD 4000, Australia. Tel: +433168734072; E-mail: thomas.vogl@aon.at

One sentence summary: A set of NLSs is provided that can be applied to optimize the nuclear import of heterologous proteins in Pichia pastoris, establishing a toolbox for molecular and synthetic biology applications.

Editor: Hyun Ah Kang

ABSTRACT

Nuclear localization sequences (NLSs) are required for the import of proteins in the nucleus of eukaryotes. However many proteins from bacteria or bacteriophages are used for basic studies in molecular biology, to generate synthetic genetic circuits or for genome editing applications. Prokaryotic recombinases, CRISPR-associated proteins such as Cas9 or bacterial and viral polymerases require efficient NLSs to function in eukaryotes. The yeast Pichia pastoris is a widely used expression platform for heterologous protein production, but molecular tools such as NLSs are limited. Here we have characterized a set of 10 NLSs for P. pastoris, including the first endogenous NLSs (derived from P. pastoris proteins) and commonly used heterologous NLSs. The NLSs were evaluated by fusing them in N- and C-terminal position to an enhanced green fluorescent protein showing pronounced differences in fluorescence levels and nuclear targeting. Thereby we provide a set of different NLSs that can be applied to optimize the nuclear import of heterologous proteins in P. pastoris, paving the way for the establishment of intricate synthetic biology applications.

Keywords: nuclear localization sequence; nuclear targeting; Pichia pastoris; heterologous gene expression; parts for synthetic biology
Solely the well characterized Simian Virus 40 (SV40) large T antigen NLS has been applied in P. pastoris by purpose for the nuclear import of a single heterologous protein, namely the prokaryotic bacteriophage T7 RNA polymerase (Hobl et al. 2013). Additional NLSs, which are functional in P. pastoris, have been identified haphazardly by different groups, when trying to improve heterologous protein production. Gradovoa and Padkina (2010) described that the NLS of a bovine gamma interferon directed the translocation into the P. pastoris nucleus and removal of this sequence led to the cytoplasmic accumulation of the recombinant protein. Yang et al. (2004) obtained 8-fold higher levels of secreted human topoisomerase I after removal of a NLS. Yet none of these NLSs have been further characterized for the targeted nuclear import of heterologous proteins. Here we aimed to bridge this gap by identifying a set of efficient NLSs for P. pastoris.

We tested five heterologous (Table S1, Supporting Information) and five endogenous (Table S2, Supporting Information) NLSs for the nuclear import of an enhanced green fluorescent reporter protein (eGFP) in P. pastoris and investigated whether C- or N-terminal positioning has an influence on the expression level and import efficiency. We used only heterologous sequences, which have previously been characterized in various eukaryotes and several yeasts for the nuclear import of proteins. The NLS of the large SV40 T antigen (Lanford and Butel 1984), derived from the Simian Virus 40, has been successfully used in Saccharomyces cerevisiae (Nelson and Silver 1989), Schizosaccharomyces pombe (Fleig et al. 2000) and P. pastoris (Hobl et al. 2013). The translocation signal of the human myc protein (HsMyc) contains a monopartite consensus motif similar to SV40-NLS and has been characterized for the import of heterologous proteins in mammalian cell lines (Dang and Lee 1988; Makker, Dingwall and Laskey 1996). Crystallographic analysis revealed that HsMyc interacted with a key factor of the S. cerevisiae nuclear import pathway (Conti and Kuriyan 2000). The NLS of the Xenopus leavis derived nucleoplasm protein (XINuc) is the most prominent bipartite classical NLS and has been successfully used in S. cerevisiae (Conti and Kuriyan 2000) and Sc. pombe (Azad et al. 1997). The non-classical NLS of the S. cerevisiae transcription repressor ScMat2 has been successfully used for the import of an Escherichia coli beta-galactosidase to the S. cerevisiae nucleus (Hall, Hereford and Herskowitz 1984). The functionality of import sequences, for instance the NLS of the S. cerevisiae transcription factor Swi5p (ScSwi5), is regulated by a phosphorylation-dependent mechanism (Jana 1995), leaving the question if such an NLS is also functional in a non-natural host such as P. pastoris.

We also tried to identify endogenous NLSs, which might have beneficial import characteristics compared to the heterologous import sequences. Five NLSs of putative nuclear proteins in the genome sequence of P. pastoris (De Schutter et al. 2009) were selected according to the published consensus motifs (Chook and Blobel 2001). The P. pastoris (Pp) protein sequences were used as queries for a BLAST search in S. cerevisiae, confirming that they are homologs of the S. cerevisiae nuclear proteins Nob1p, Sda1p, Set7p and Uba1p (Cherry et al. 2012) (Table S2, Supporting Information). The putative NLSs of these proteins contain the bipartite consensus motif of NLSs for the classical import pathway and are located on the C-terminus. Additionally, we also included an NLS from the P. pastoris homolog of ScSwi5.

The heterologous and endogenous NLSs were directly fused to the eGFP coding sequence (S3a, Supporting Information). All DNA and protein sequences of the NLSs used in this study are listed in Tables S1 and S2 (Supporting Information). The eGFP-NLS fusion proteins were expressed under the control of the exceptionally strong, methanol inducible promoter of the alcohol oxidase 1 gene (P<sub>AOX1</sub>) (Vogl and Glieder 2013). P<sub>AOX1</sub> is repressed by glucose and can be strongly induced with methanol, when glucose is depleted. The transformants were cultivated for 60 hours in glucose-containing media and methanol induction was performed for 24 hours (S3b, Supporting Information). The fluorescence of the cells under the microscope was very intense and it was not possible to localize subcellular compartments, because of the high levels of intracellular eGFP (S3b, Supporting Information). Thus, we modified the cultivation protocol in order to reduce eGFP expression. The cells were induced for four hours with methanol-containing media, subsequently glucose was added to repress P<sub>AOX1</sub> and cultivation was prolonged for additional four hours (S3b, Supporting Information). Thereby the production of eGFP ceased and the reporter protein could accumulate in the nucleus due to additional time for sorting. We have also ruled out negative effects of methanol induction on nuclear targeting by cloning a subset of the NLSs and controls with constitutive promoters (S4, Supporting Information).

The localization of the eGFP-NLS fusions was observed with a Leica DM LB fluorescence microscope (Fig. 1a and b) and fluorescence levels were measured with a fluorescence spectrometer (Fig. 1c). The nuclear localization of eGFP was confirmed by Hoechst 33 258 staining (S3c, Supporting Information) and the results are summarized in Table S5 (Supporting Information). The microscopy images (Fig. 1a) show that the NLS of the SV40 large T antigen, the XINuc, HsMyc and ScSwi5 proteins were able to direct the reporter protein to the nucleus, when they were fused either N- or C-terminally to eGFP. Only the C-terminal application of the ScMat2 NLS did not promote eGFP translocation. The ScMat2 NLS was also ineffective, when fused N-terminally to a recombinant protein in mammalian cells (Chelsky, Ralph and Jonak 1989). In its natural role the ScMat2 NLS is not located close to the termini but rather in the middle of the protein sequence and functionality of this sequence may depend on the adjacent protein context. The fact that heterologous NLSs are recognized very efficiently in P. pastoris underlines the finding that the import pathway is well conserved among different eukaryotic species (DeGrasse et al. 2009).

N-terminal SV40 NLS-eGFP fusions were transported to the nucleus, but the fusion protein showed low fluorescence levels (as determined by fluorescence spectroscopy measurements, Fig. 1c). We assume that the N-terminal fusion of this NLS affected the stability or folding of eGFP (Fig. 1c). This finding is remarkable, since the SV40 NLS is the most commonly used sequence for the nuclear import of heterologous proteins. Also for the endogenous PpSet7 and PpUba1-fusion constructs relatively low fluorescence levels were measured. Here, bright nuclei and little amount of cytoplasmic eGFP were observed under the microscope. The microscopy images (Fig. 1a) show that the NLS of the PpSet7 and PpUba1 appear to be very effective for the nuclear import of eGFP. The NLSs PpNob1, PpSda1 and PpSwi5 did not promote translocation, although these non-functional targeting sequences are similarly organized as the functional sequences of PpSet7 and PpUba1 (a cluster of basic aa followed by a spacer and a second cluster of basic aa). This finding implies that additional residues might be involved in the translocation in the context of the native proteins.

Quantitative fluorescence spectroscopy measurements after 24 hours methanol induction show that all fusion constructs affect the expression levels of the reporter gene compared to the control (eGFP without an NLS fusion, produced in the cytosol, Fig. 1c). The addition of the NLSs might trigger protein degradation or have an effect on the conformation, which may reduce reporter protein fluorescence. The reduction of the signal...
Figure 1. Fusions of heterologous and endogenous NLSs to an eGFP reporter protein show pronounced differences in nuclear targeting and fluorescence levels in P. pastoris. (a) Heterologous NLSs of S. cerevisiae (left side of the panel) and from higher eukaryotes and a eukaryotic virus (right side of the panel) were fused N- and C-terminally (indicated by prefixes N- and C-) to an eGFP reporter gene and transformed in P. pastoris. The nuclei were stained with Hoechst 33258. Translocation of the fusion proteins was observed by fluorescence microscopy. Bright field images (BF) and fluorescence images using different filters are shown (Hoechst, eGFP). As a localization control the eGFP gene was expressed without an NLS (w/o NLS). The CBS 7435 wild-type strain not expressing eGFP was used as a negative control. All cells were treated identically as described in S3b (Supporting Information). (b) Endogenous NLSs from P. pastoris nuclear proteins were fused C-terminally to the eGFP reporter gene and analyzed as described in panel (a). (c) Quantitative fluorescence spectroscopy measurements of eGFP-NLS fusion constructs. The relative fluorescence levels of the eGFP-NLS fusion constructs to the control (GFP without an NLS) are shown. The strains were measured according to the protocol outlined in S3b (Supporting Information). Mean values and standard deviations of biological 6-fold replicates are shown.
strength can also be caused by the translocation of reporter protein to the nucleus, where eGFP is surrounded by an additional layer, the nuclear envelop, which might impede the strength of the fluorescence signal. The relative fluorescence levels after four hours methanol induction and additional four hours growth on glucose are provided in S3b (Supporting Information). We also aimed to test the NLSes fused to red fluorescent protein variants mCherry and dTomato (Shaner et al. 2004) but noticed mistargeting of the control construct without an NLS (S6, Supporting Information). Hence, it appears that mCherry and dTomato contain a cryptic sequence recognized as targeting signal in P. pastoris thereby complicating its use as reporter for evaluating intracellular localization of fusion proteins/peptides in P. pastoris.

The efficiency of an NLS depends on the size and sequence of the protein to which it has been linked (Nelson and Silver 1989). Therefore, it is favorable to test more than one NLS for the import of a heterologous protein of interest and the efficiency of a particular NLS may vary depending on the protein to which it is fused. The NLSes reported here constitute on the one hand a useful resource specifically for the P. pastoris community. On the other hand, the findings appear also relevant for other yeast systems: NLSs can show drastic differences in functionality ranging from efficient nuclear targeting even to detrimental effects on the protein of interest. Therefore, our work may provide a useful resource for researchers, who want to establish similar toolboxes of NLSs in other host organisms.

**SUPPLEMENTARY DATA**

Supplementary data are available at FEMSYR online.

**ACKNOWLEDGEMENTS**

The images in the graphical abstract showing heterologous proteins were kindly provided by the RCSB PDB (taken from the Molecule of the Month features by David S. Goodsell, doi: 10.2210/rcsb_pdb/mom_2003_6; 10.2210/rcsb_pdb/mom_2014_12; 10.2210/rcsb_pdb/mom_2015_1). The schematic illustration of the P. pastoris cell was taken from Vogl, Hartner and Glieder (2013).

**FUNDING**

The research leading to these results has received funding from the Innovative Medicines Initiative Joint Undertaking project CHEM21 under grant agreement no. 11 5360, resources of which are composed of financial contribution from the European Union’s Seventh Framework Programme (FP7/2007-2013) and EFPIA companies’ in kind contribution. We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 09-N14-B12 and PIA companies’ in kind contribution. We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 09-N14-B12 and PIA companies’ in kind contribution. We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 09-N14-B12 and PIA companies’ in kind contribution. We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 09-N14-B12 and PIA companies’ in kind contribution. We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 09-N14-B12 and PIA companies’ in kind contribution. We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 09-N14-B12 and PIA companies’ in kind contribution. We gratefully acknowledge the Austrian Science Fund (FWF) project number W901 09-N14-B12 and PIA companies’ in kind contribution.

**Conflict of interest.** None declared.

**REFERENCES**


