Search for the Nuclear Localization Signal of Ime4

Christian Monroy Hernandez

University of New Orleans

Follow this and additional works at: https://scholarworks.uno.edu/honors_theses

Part of the Biology Commons

Recommended Citation
https://scholarworks.uno.edu/honors_theses/115

This Honors Thesis-Unrestricted is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Honors Thesis-Unrestricted in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Honors Thesis-Unrestricted has been accepted for inclusion in Senior Honors Theses by an authorized administrator of ScholarWorks@UNO. For more information, please contact scholarworks@uno.edu.
Search for the nuclear localization signal of Ime4

An Honors Thesis

Presented to

The Department of Biology

of the University of New Orleans

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science, with University High Honors

and Honors in Biology

By

Christian Monroy Hernandez

May 2018
Acknowledgements

I would like to express my sincerest gratitude to my mentor and advisor throughout my undergraduate career, Dr. Mary Clancy. This entire research endeavor would not have been possible without your advice, expertise and your boundless patience. Thank you for letting me be part of your lab throughout these years; I owe this project, and my research career, to you.

I would also like to thank Dr. Zhengchang Liu and Dr. Wendy Schluchter for agreeing to be my second reader and third reader, respectively. Your input throughout the different stages of this project is greatly appreciated.

Finally, I would like to thank all of my friends and family who have supported me throughout these four years. Special thanks go to my parents: you believed in me from the very start and encouraged me throughout the entire way; none of this would have been possible without you. It has been quite a journey, and I am blessed to have had you all by my side.
Table of Contents

List of Figures ..............................................................................................................iv
Abstract.......................................................................................................................v

Background.................................................................................................................1

Ime4 as a Conserved Methyltransferase.................................................................1
MTases as part of a nuclear complex......................................................................3
General overview of nuclear transport.................................................................6
Classical pathway of nuclear import.................................................................7
Non-classical nuclear localization.................................................................8
Nuclear localization of METTL3 and Ime4.........................................................9

Results......................................................................................................................11

Discussion...............................................................................................................20

Materials and Methods..........................................................................................24

Yeast strains..........................................................................................................24
Media......................................................................................................................24
Primer Design.........................................................................................................25
Template Acquisition and Restriction Digest.....................................................26
PCR Amplification.................................................................................................27
PCR Product Purification.......................................................................................28
Plasmid Assembly.................................................................................................28
E. coli Transformation...........................................................................................29
Plasmid Isolation.................................................................................................30
Microscopy............................................................................................................30
List of Figures

Figure 1: Protein sequence alignment of Ime4 and METTL3……………………………………1
Figure 2: Crystal structure visualization of METTL3-METTL14 heterodimer…………………4
Figure 3: C-terminal sequence of Ime4 highlighting basic residues…………………………11
Figure 4.1: C-terminal sequence of Ime4 showing amplicons of inserts used…………………..12
Figure 4.2: Agarose gel visualization of Ime4 inserts and vector used…………………………12
Figure 5: Expression of GFP-pC128 in BY4741 cells ..................................................14
Figure 6: Expression of GFP-Far10 in BY4741 cells......................................................15
Figure 7: Expression of GFP-pC105 in BY4741 cells......................................................16
Figure 8: Expression of GFP-pC073 in BY4741 cells......................................................17
Figure 9: Expression of GFP-pC059 in BY4741 cells......................................................18
Figure 10.1: Expression of GFP-pC128 in kar4Δ/Δ cells.................................................19
Figure 10.2: Expression of GFP-pC128 in slz1Δ/Δ cells.................................................20
Figure 11: Summary of localization patterns displayed by inserts……………………………20
Figure 12: Secondary structure of the C-terminus of Ime4 in relation to insert amplicons……21
Abstract

Ime4 is the catalytic subunit of a conserved methyltransferase (MTase) complex found in yeast, *S. cerevisiae*. This complex is responsible for creating the RNA modification $N^6$-methyladenosine ($m^6$A), the most common post-transcriptional modification in higher eukaryotes. There is evidence to suggest that $m^6$A is an important mediator of gene expression control within the cell and has been associated with a diverse array of phenotypic effects, notably as a conserved determinant of cell fate. The MTase complex is known to be a nuclear protein, the compartment where it is believed to carry out most of its methylation activity. Recently, the nuclear localization signals (NLS) of the subunits of the human MTase complex were experimentally identified, whereas the NLSs of the yeast MTase complex remain unknown. Here, we have experimentally identified the amino acid sequence $^{517}$RKYQEFMKSKTGSHTGKK$^{540}$, located within the C-terminal region, as a putative bipartite NLS for Ime4.

Keywords: Ime4, $m^6$A, methyltransferase, nuclear, localization, NLS
Background

Ime4 as a Conserved Methyltransferase

Ime4 is a methyltransferase (MTase) found in yeast, *S. cerevisiae*, which acts as the catalytic subunit of the multi-protein complex known as the MTase complex. The MTase complex is responsible for creating $N^6$-methyladenosine ($m^6$A), a chemical modification of RNA transcripts that involves the addition of a methyl group to the 6th position of adenosine bases (Clancy et al. 2002; Agarwala et al. 2012). $m^6$A is the most abundant post-transcriptional modification in higher eukaryotes, and is found in most mammals, plants, insects and viruses (Clancy et al. 2002; Lee et al. 2014). In humans, $m^6$A is formed by Ime4’s orthologous protein METTL3; both are members of the MT-A70 protein superfamily and share a remarkable level of sequence similarity, particularly at the catalytic region (Bokar et al. 1997).

Figure 1: Protein alignment shows highly conserved amino acid residues between METTL3 and Ime4 (alignment done with software developed by Stothard 2000). Within METTL3, conserved amino acids R465, R471, H474 and H478 are believed to serve as RNA-binding residues; likewise, conserved amino acids D377, D395, N539 and E532 are critical S-adenosylmethionine (METTL3’s methyl donor) binding residues, necessary for enzymatic activity (Wang et al. 2016).
Regulation by m\(^6\)A takes place in association with several other interacting proteins. Similarly to how MTase complex “writes” m\(^6\)A, a set of RNA-binding proteins belonging to the YTH domain family “read” it, and are believed to facilitate the degradation of the methylated transcript by targeting them towards P-bodies (Dominissini et al. 2012; Wang et al. 2014). Finally, the fat mass and obesity-associated protein (FTO) and ALKBH5 act as demethylases which can “erase” m\(^6\)A; so far, this mechanism has not been found to be present in yeast (Jia et al. 2011; Fu et al. 2014).

The range of effector functions of MTase-mediated methylation is incredibly varied and poorly understood, but, at its core, m\(^6\)A appears to play a role in gene expression control under certain circumstances. In mammalian cells, the direction in which m\(^6\)A drives this process is very context-dependent, as there is evidence linking methylated transcripts with increased levels of translation, as well as increased levels of degradation (Wang et al. 2014; Meyer et al. 2012; Meyer et al. 2015). In yeast, methylated transcripts are linked with higher levels of ribosomal association at the start of meiosis, suggesting that Ime4-mediated methylation may induce a higher rate of transcript translation (Bodi et al. 2015). Yeast does contain a YTH domain protein (Pho92) which targets \(PHO4\) transcripts for degradation, suggesting there may be a link between m\(^6\)A and transcript degradation; but until now, the \(PHO4\) transcript has not been identified as a methylated transcript, and no explicit association of m\(^6\)A-associated transcript degradation is known (Bodi et al. 2015; Wang et al. 2014).

However, identifying the explicit phenotypic effects of this m\(^6\)A-mediated gene expression is particularly challenging in humans, due to the presence of ~7,000 methylated mRNAs and ~300 methylated ncRNA (Fu et al. 2014). Certain generalized phenotypic observations following perturbation of the m\(^6\)A machinery have been observed in higher eukaryotes, though. For example,
in human and mouse cell lines, elimination of certain members of the MTase complex, including METTL3, causes an inability of embryonic stem cells to terminate their naive state, leading to embryonic lethality (Batista et al. 2014; Shwartz 2016). In Drosophila, m^6A has been shown to have a strong interaction with the Notch signaling pathway, a conserved pathway involved in metazoan development (Hongay et al. 2011). Likewise, in Arabidopsis, m^6A is essential for embryonic development, as well as for the normal growth of mature individuals (Bodi et al. 2012).

Conversely, in yeast, Ime4 is not an essential protein but it is considered a key regulator for the entry into meiosis following nutritional starvation. Most strains have noticeable difficulties undergoing meiosis when lacking a functional copy of Ime4, believed to be a result of the absence of methylase activity within these cells (Clancy et al. 2002; Agarwala et al. 2012). Additionally, MTase-driven methylation also appears to be a negative regulator for entry into foraging pseudohyphal (PH) growth, an alternative growth program that occurs in response to nutrient starvation (Agarwala et al. 2012). Thus, Ime4’s methylation effect in yeast following nutritional starvation is consistent with the apparent conserved function of m^6A in cell-fate decision and normal organismic development.

**MTases as part of a nuclear complex**

As mentioned before, both Ime4 and METTL3 carry out their MTase activity as part of a conserved protein complex (Ke et al. 2017; Agarwala et al. 2012; Wang et al. 2014). The initial isolation of the methyltransferase complex from HeLa cells yielded two fractions of nuclear extract of sizes ~200 kDa and ~800 kDa. The first identified subunit was the 70 kDa METTL3, then known MT-A70; subsequently, METTL14, another member of MT-A70 protein superfamily, was discovered to also be part of the MTase complex, forming a tight heterodimer with METTL3
coordinated by an extensive network of hydrogen bonding (see Figure 2). Their heterodimeric complex interaction is necessary for full levels of RNA methylation to take place, as isolated METTL3 and METTL14 do not exhibit significant levels of m\(^6\)A. However, unlike METTL3, METTL14 does not contain a SAM binding site and is not believed to directly possess methyltransferase activity; instead it is thought to serve as a stable scaffold to facilitate RNA binding (Wang et al. 2014; Wang et al. 2016; Schöller et al. 2018). Subsequently, WTAP was also identified to be part of the MTase complex; like METTL14, WTAP mutants show a significant decrease in m\(^6\)A levels, suggesting it plays an important role in coordinating the methylation activity (Wang et al. 2014). WTAP appears to interact with the METTL3-METTL14 heterocomplex and drive its subnuclear localization towards nuclear speckles, where they associate with pre-mRNA processing factors (Ping et al. 2014).

Figure 2: A crystal structure of METTL3 (orange) and METTL14 (yellow) in the SAM-bound form. The residues labelled in green are a series of basic amino acids, which comprise the putative
Ime4 was the first identified member of the MTase complex after its methyltransferase activity was characterized (Clancy et al. 2002). Subsequently, the meiotic proteins Mum2 and Slz1 were also discovered to coordinate methylation activity the yeast MTase complex; together, the three proteins are known as the MIS complex (Agarwala et al. 2012). Sequence analysis revealed homology between Mum2 and its WTAP, while Slz1 was not identified to have a mammalian homologue (Fu et al. 2014). Ime4 is the only subunit with known SAM-binding and MTase domains, although Mum2 appears to play a critical role in the methylation activity, as mum2Δ/Δ mutant cells lack significant levels of m6A. It is likely that Mum2 aids Ime4 by either activating its catalytic region or by helping it bind to its mRNA substrates (Agarwala et al. 2012). Slz1, on the other hand, plays a role in the nucleolar localization of the MIS complex; full levels of methylation cannot be achieved in sporulating diploid cells until Slz1-mediated nucleolar localization takes place (Schwartz et al. 2013; Agarwala et al. 2012).

It is possible that the MTase complex possesses additional uncharacterized subunits. It is known that WTAP is associated with many other proteins and could potentially recruit them to the MTase complex (Fu et al. 2014). Similarly, according to the *Saccharomyces Genome Database*, apart from Slz1 and Mum2, Ime4 has many other putative protein interactions; notably, yeast two-hybrid assays identified Kar4, the METTL14’s ortholog, among them (Ito et al. 2001; Růžička et al. 2017). Although Kar4 is a member of the MT-A70 protein superfamily, it lacks the SAM binding residues and is unlikely to have MTase activity; thus far, it has not been shown to be part of the MTase complex in yeast (Lahav et al. 2007).
An interesting common factor between the MTase complexes in yeast and human is the nuclear localization of their members. In HeLa cells, fractionation experiments have determined that most of the methylation activity occurs within the nucleoplasm on nascent pre-mRNA (Ke et al. 2017). As mentioned before, the MTase complex also undergoes an additional level of subnuclear localization, but the two organisms differ in their destination; in yeast the MTase complex localizes to the nucleolus, while in human it localizes to nuclear speckles. It is still unclear how the subnuclear localization plays a role in the methylation (i.e., whether the methylase activity is carried out within the nucleoplasm or in these subnuclear locations). However, in both organisms, knockdown of the protein believed to drive the subnuclear localization of the complex (Slz1 in yeast and WTAP in humans) leads to reduced levels of m^6A (Agarwala et al. 2012; Liu et al. 2013).

Thus, nuclear localization appears to be a crucial step for the normal enzymatic activity of the MTase complex. Nuclear transport is a tightly regulated process within the cell and little is known about how Ime4 carries out nuclear localization. Given the intricate nature of this transport pathway, a quick overview of the current knowledge of nuclear import is given below. Subsequently, our current understanding regarding the specifics of nuclear import of the MTase complex in yeast and humans will be reviewed.

**General overview of nuclear transport**

A common mechanism, by which many proteins localize to their corresponding compartments is the recognition of a specific amino acid “signal sequences” by transport proteins, either co-translationally or post-translationally (Nielsen, et al. 1999). Each compartment has a specific set of sequences, and this sequence alone is sufficient to direct the transport of the cargo
protein to its intended destination (Hurt et al. 1984). In the case of nuclear localization, transport is mediated by short series of amino acid sequences termed nuclear localization signals (NLS) and nuclear export signal (NES), for protein import and export, respectively. These signals are recognized by a family of proteins known as karyopherins, which include “importins”, mediating import of proteins by binding to NLSs, and exportins, mediating export of proteins by binding to NESs (Lange et al. 2007). Once bound to their protein cargo, karyopherins interact with the nuclear pore protein complex on the nuclear envelope, which provide selective entry into the nucleus (Marfori et al. 2007).

**Classical pathway of nuclear import**

The most well-known NLS is the classical NLS (cNLS), which was first identified in the simian virus 40 (SV40) large-T gene. This prototypical monopartite cNLS, with sequence \( ^{126} \text{PKKKRKV}^{132} \), is notable for being composed mainly of basic amino acid residues, here shown in bold (Kalderon et al. 1984). An additional sequence type was further characterized through the nucleoplasmin chaperonin of *Xenopus laevis*, which was shown to have a bipartite cNLS (\( ^{155} \text{KRPAATKKAGQAKKKK}^{170} \)) showing two patches of basic amino acids separated by a linker sequence (Robbins et al. 1991).

Many additional monopartite and bipartite cNLSs of varying lengths and sequences have been identified, but notably all of them have crucial basic amino acid residues that are necessary for full nuclear localization of the protein. In humans, all cNLS-driven nuclear localization is carried out by a heterodimer karyopherin (Kap) complex composed of an adaptor protein, Kap\(\alpha\), that binds directly to the cargo and subsequently binds itself to Kap\(\beta1\) (Conti et al. 1998). In yeast, a similar cNLS-driven pathway exists, mediated by Srp1 (Kap\(\alpha\) homolog) and Kap95 (Kap\(\beta1\).
homolog) (Tab et al. 2000). However, multiple non-cNLS nuclear localization pathways exist which initiate the transport of their cargo by the direct recognition of the NLS by a Kapβ-like protein, without employing a Kapα adaptor (Fries et al. 2007; Kobayashi et al. 2013).

Non-classical nuclear localization

The Kapβ protein family is responsible for the majority of nuclear transport within the cell, having 20 known members in humans and 14 in yeast (Kobayashi et al. 2013; Lee et al. 2006). Only a few members of the Kapβ family have well characterized pathways with a corresponding NLS, notably Kapβ1 which mediates the cNLS import pathway. Likewise, Kapβ2, also known as transportin, is another well characterized karyopherin responsible for the recognition of another well defined NLS known as the PY-NLS, with consensus sequence R/H/K-X(2-5)-P-Y (Lee et al. 2006).

Another well studied member of the Kapβ family, Kap121p, provides an added level of complexity to the Kap-NLS interactions, as it is able to recognize multiple NLS sequences. Specifically, it recognizes the NLS with consensus sequence KV/IxKx1-2K/H/R while also being able to recognize the rg-NLS, a peptide sequence containing multiple arginine and glycine residues. The non-specificity of NLS recognition can also be reversed, as NLSs can be recognized by multiple Kap proteins, exemplified by the rg-NLS which is also recognized by the Kap104p (Kobayashi et al. 2013; Leslie et al. 2004). This non-specific Kap-NLS interaction can reach an impressive level of degeneracy, shown by the conserved consensus NLS (R/KxxL(x)nV/YxxV/IxK/RxxK/R) found in H2A and Asr1p, which is recognized by 5 different yeast importins (Kap114p, Kap95p, Kap123p, Pse1p, and Kap104p) (Fries et al. 2007).

Although nuclear import was first characterized via the cNLS, the complexity of the regulation of nucleo-cytoplasmic transport has become apparent as more non-cNLS-mediated
methods of localization are discovered. It is now known that it is also possible for proteins to undergo nuclear localization without having any form of NLS; instead, they may undergo localization as a heterodimer, “piggy backing” on other protein(s) that do possess an NLS (Steidl et al. 2004). Similarly, other NLS-independent mechanisms have also been discovered, instead using post-translational modifications at specific residue, such as phosphorylation, to carry out localization (Wu et al. 2008). Ultimately, the extensive network of possible pathways, signals, and mechanisms that can mediate nuclear transport makes the task of creating a comprehensive characterization a distant notion.

**Nuclear localization of METTL3 and Ime4**

To date, no studies have experimentally identified NLS on members of the MIS complex. *In silico* predictions of putative NLS only yielded an obvious cNLS for Slz1, with sequence \text{12KKYEVKLKPDRQVKKNK}\text{28} (Lange et al. 2007). Similar predictions for Ime4 identified only a weak bipartite NLS within the nonconserved region, with sequence \text{290RETASENKRI RNVSIPFYTLGNCSAHCKKALP}\text{323}. Likewise, an extremely weak bipartite NLS for Mum2 with sequence \text{228EHKEYKENCHIPQTFLPASLEVIFRKLSS}\text{257} was also identified (Kosugi et al., 2009).

In humans, however, a recent study was successful in experimentally identifying the presence of NLSs within members of the MTase complex. Within METTL3, a monopartite variant of the basic NLS was identified, with sequence \text{209AKKS R K}\text{215}; mutation of this sequence resulted in the cytoplasmic localization of METTL3, suggesting that it indeed mediates METTL3’s nuclear transport. Likewise, the same study also identified a putative WTAP NLS, with sequence \text{5PLPKKVR L}\text{13}; mutation of this sequence also resulted in cytosolic localization of WTAP.
METTL14, though, was found to not have its own NLS, as mutation of a predicted putative NLS still resulted in METTL14’s nuclear localization. Instead, it was found that METTL14 is dependent on the METTL3’s NLS, as METTL14 remained cytosolic in METTL3-NLS mutants. This suggests METTL14 utilizes the piggy-back mechanism, colocalizing into the nucleus with METTL3 in their heterodimeric state (Schöller et al. 2018). Presumably, once the three subunits have localized in the nucleus, WTAP interacts with the METTL3-METTL14 heterodimer and directs its subnuclear localization towards nuclear speckles (Ping et al. 2014).

Unfortunately, METTL3’s recently discovered NLS is not located within the conserved region of Ime4, and thus is of little use as a reference for identification of any possible NLS in Ime4 (see alignment on Figure 1). Previous work in the laboratory has been carried out in an attempt to experimentally explore this possibility through mutational analysis. Two Ime4 mutants (lacking residues 2-215 and 546-599, respectively) were created but both still localized within the nucleus, suggesting Ime4’s putative NLS would be located elsewhere within the protein. Furthermore, this study also concluded that it is unlikely for Ime4 to piggyback off any possible NLS Mum2 might possess, as one of the mutants (Δ2-215) removed the putative Mum2 binding site of Ime4 (Dehon, 2012).

*In silico* predictions of nucleolar localization signals within Ime4 yielded a putative nucleolar localization signal (NoLS) at sequence $^{528}$TSHTGTKKDQPSKLQQHQYQWN$^{556}$ (using software developed by Scott et al. 2010). As it has been previously shown that NLS and NoLS can overlap (Liu et al. 2006), this region is of particular interest for searching for potential NLSs within Ime4. Granted, Slz1 has been suggested to mediate nucleolar localization of the MIS complex, as Mum2 and Ime4 did not localize within the nucleolus in ime1Δ/Δ mutants, which lack Slz1 expression (Agarwala et al.
However, the precise dynamics of the subnuclear localization are not well understood, and some form of redundancy with a putative NoLS could be possible. Regardless, the presence of several basic amino acids patches in this unexplored general region still makes it a promising region to search for possible NLSs (see Figure 3).

![Figure 3: Within the C-terminal region of Ime4 (~aa 471-550) there is a series of basic amino acids (shown in red box) that could potentially be part of a putative NLS within Ime4.](image)

Thus, through this study we attempted to use molecular cloning techniques to explore the C-terminal region of Ime4 for the presence of a potential NLS. We created several Ime4 mutants, each missing a different region of the C-terminus, and successfully determined a key amino acid sequence that is sufficient to drive nuclear localization of mutant. This could lead to the identification of the NLS within Ime4.

**Results**

It has been observed that many NLSs are located in the N- or C-termini of proteins, potentially due to the disordered nature of these region; this provides a higher level of structural flexibility and allows for longer linker sequences in bipartite NLS (Kosugi et al. 2009). Previous mutational analysis done in the laboratory suggests that Ime4’s NLS is not located within the N-
terminus or the extreme C-terminus (last 50 residues) (Dehon, 2012). Due to the presence of several basic amino acid residues within the rest of the C-terminal area [aa ~471-550], we decided to explore the possibility that these residues could be part of a putative NLS.

**Figure 4.1**: Shown above are the regions of the four Ime4 fragments that we expressed, in an attempt to pinpoint the NLS. After the pilot fragment protein, pC128 (128 residues; aa 473-600), was identified to undergo nuclear localization, the additional constructs were created to further pinpoint the amino acid sequence coding for the NLS. The respective start sites of pC128, pC105 (105 residues; aa 496-600) and pC073 (73 residues; aa 528-600) are shown with an arrow; residues underlined in green compose the insert pC059 (59 residues; aa 468-525, 600). The basic amino acid residues that could play a role in nuclear localization are boxed in red. Insert design and cloning procedures are outlined in the Materials and Methods section.
To further test this hypothesis, we created a series of constructs that express different sections of the C-terminal region of Ime4; a summary of all clones created, and their respective amplicons is shown in Figure 4.1. The first fragment amplified encodes the last 128 amino acid residues of IME4; thus, the IME4 fragment (pC128) expressed residues 473-600. We cloned this fragment in the plasmid vector pRS416-ADH1p-GFP, where it was fused to the 3’ end of the open reading frame of the normally cytosolic biomarker, green fluorescent protein (GFP). Following transformation of the construct into the S. cerevisiae strain BY4741, the GFP-pC128 fusion was expressed from the ADH1 promoter while the cells were grown in SC minimal liquid media. These cells were subsequently imaged using fluorescence microscopy as described in the Materials and Methods. The GFP-pC128 fusion protein was observed to locate in the nucleus. Nuclear localization was confirmed by co-staining with DAPI, a DNA binding fluorophore (Figure 5).
Figure 5: *BY4741* cells expressing the GFP-pC128 (residues 473-600 of Ime4) fusion protein were grown in SC minimal media and imaged using fluorescence microscopy (procedure detailed in Materials and Method). GFP-pC128 was observed to localize to the nucleus. Nuclear localization was confirmed by co-staining with DAPI.

As a negative control, the *pRS416-ADH1p-GFP* vector was also used to express a GFP-Far10 fusion. Far10 has been shown to interact with other Far proteins involved in cell cycle arrest following pheromone treatment; as it has not been shown to be nuclear, it is an ideal control for imaging purposes (Kemp et al. 2003). Compared to cells expressing GFP-Far10 (Figure 6), transformants expressing the GFP-pC128 fusion (Figure 5) showed a greater degree of nuclear localization of GFP. This indicates that the presence of GFP within the nucleus of these cells is due to amino acids 473-600 of the Ime4 protein.
Figure 6: *BY4741* cells transformed with the GFP-Far10 fusion were grown in SC media and imaged under similar conditions as GFP-pC128 transformants. GFP-Far10 showed cytosolic localization, with no sign of nuclear localization; this cytosolic localization makes it a useful negative control for imaging purposes.

Once GFP-pC128 was experimentally determined to localize in the nucleus, two smaller fragments were PCR amplified to further isolate the NLS: *pC105* [aa 496-stop] and *pC073* [aa 528-stop]. These inserts were both cloned into the *pRS416-ADH1p-GFP* vector and transformed into *BY4741* in a similar fashion to pC128. Upon expression of GFP-pC105, we observed distinct localization and enrichment within the nucleus, similar to GFP-pC128 expression.
**Figure 7:** *BY4741* cells transformed with the GFP-pC105 fusion were grown in SC media and imaged under similar conditions as the previous constructs. The GFP-pC105 fusion showed clear localization and enrichment within the nucleus, similarly to GFP-pC128.

In contrast, imaging of GFP-*pC073* transformants showed a cytosolic and weakly nuclear localization pattern of GFP (Figure 8). This localization pattern may be due to the disruption of a potential bipartite signal, which could have an upstream element (\textsuperscript{514}DVERKYQEFMKS\textsuperscript{527}) immediately before *pC073*'s amplicon start site. If that is the case, the putative downstream element (\textsuperscript{534}TKKIDKKQ\textsuperscript{541}) encoded within *pC073* could be enough to elicit the weak nuclear localization that was observed, but not full localization observed by GFP-pC128 and GFP-pC105 fusions.
BY4741 cells transformed with GFP-pC073 [aa 528-600] were grown in SC media and showed cytosolic and weakly nuclear localization patterns. The split localization pattern may be due to a potential disturbance of a putative bipartite NLS. Its downstream element (\textsuperscript{514}TKKIDKKQ\textsuperscript{541}) is within pC073’s coding region, and this segment could cause the weak nuclear localization observed in transformants; its putative upstream element (\textsuperscript{514}DVERKYGEMSKT\textsuperscript{527}) might be necessary for full nuclear localization to take place.

Additionally, we also created a bipartite construct fusing the coding region for residues 468-525 with the extreme 3' end, which included amino acid 600 and part of the 3' UTR [nt 2798-3033], included for its regulatory functions. The resulting fragment, pC059 (aa 468-525, 600), was subsequently cloned into the pRS416-ADH1p-GFP vector. Upon expression of GFP-pC059 (Figure 9), it underwent mainly cytosolic localization with slight nuclear localization, similarly to GFP-pC073. Since the upstream element of the putative bipartite NLS (\textsuperscript{514}RKYQEMKS\textsuperscript{525}) is encoded by pC059, this segment may be enough to cause weak nuclear localization but may not be enough for full nuclear localization to take place.
Figure 9: BY4741 cells transformed with GFP-pC059 [aa 468-525; 600] were grown in SC media and showed cytosolic and weakly nuclear localization patterns. Like pC073, its split localization pattern may be due to a potential disturbance of a putative bipartite NLS. Part of its upstream element (514RKYQEFMKS) is located within pC073’s coding region; this sequence may be enough to elicit the weak nuclear localization that is observed.

Finally, as mentioned previously, nuclear localization can take place through NLS-independent pathway, such as the “piggyback” mechanism. Since in humans METTL14 has been shown to piggyback off METTL3’s nuclear localization signal (Schöller et al. 2018), we decided to explore whether Ime4 utilizes this mechanism to undergo nuclear localization. To test this possibility, we decided to transform some of our constructs into mutant yeast strains deficient in proteins believed to interact with Ime4.

We tested the possibility of Ime4 undergoing nuclear localization through the piggyback mechanism, using the NLS of the known transcription factor Kar4 or the MIS complex member, Slz1 (Lahav et al. 2007; Agarwala et al. 2012). Kar4 shows strong homology to METTL14 in humans and is thought to be its orthologous protein in yeast (Růžička et al. 2017), making it a
protein that could possibly drive Ime4’s nuclear localization. We transformed constructs coding for GFP-pC128 into kar4Δ/Δ and slz1Δ/Δ mutants, respectively, and examined the transformants using fluorescence microscopy. Both strains showed localization patterns similar to those observed in BY4741 transformants, exhibiting clear enrichment and localization of GFP within the nucleus (Figure 10.1; 10.2). These results suggest that Ime4 can undergo nuclear localization in a Kar4 and Slz1 independent manner, and could encode for its own NLS; however, it is still possible that Ime4 could use the piggyback mechanism using the NLS of a different protein.

Figure 10.1
kar4Δ/Δ (Figure 10.1) and slz1Δ/Δ (Figure 10.2) cells were transformed with GFP-pC128 and grown in SC media. Fluorescence microscopy revealed localization and enrichment of GFP within the nucleus, suggesting that Ime4’s nuclear import is independent of Kar4 and Slz1.

Discussion

Figure 11: Summary of localization patterns displayed by all the constructs that were tested.
Nuclear import is a tightly regulated process, and most proteins attempting to enter this compartment need to possess some sort of NLS to successfully enter the nucleus, or to interact with a protein that does. The MTase complexes, both in humans and in yeast, are believed to carry out most of their catalytic activity within the nucleus and therefore must have some mechanism that mediates their nuclear import. In humans, the core member of the MTase complex, METTL3, was recently discovered to undergo nuclear localization using what appears to be a variant of the monopartite cNLS (Schöller et al. 2018). In yeast, however, no known NLS has been experimentally identified within Ime4; however, previous research conducted in the laboratory suggested it might be located within the C-terminus (Dehon 2012). To further explore this, we created several fusion constructs expressing different sections of the C-terminus of Ime4 and observed their localization pattern (See Figure 11 for summary). Based on our results, we identified the amino acid sequence $^{517}\text{RKYQEFMSKKTGTSHTGTKKIDKK}^{540}$ as a strong candidate for a putative bipartite NLS.

![Figure 12](image)

**Figure 12:** Ime4 fragments shown in relation to the secondary structure of Ime4; residues 524-597 are predicted to compose a large disordered region. Linker sequence of the putative bipartite NLS
is located within the predicted disordered region, allowing for greater flexibility and a greater linker length (Prediction of secondary structure done with software developed by Rost et al. 1994).

This putative NLS appears to be a variant of the bipartite cNLS, exemplified by *Xenopus laevis*’ nucleoplasmin protein \(155^{\text{KRPAATKKAGQAKKKK}}_{170}\), albeit with a longer linker sequence and a spread out downstream element (Robbins et al. 1991). Secondary structure predictions revealed that the linker sequence [aa 519-534] is located within a large a disordered protein region of Ime4, spanning residues 524-597 (Figure 12). Disordered regions are a common location for bipartite NLSs, as they allow proteins to have a greater degree of structural flexibility (Kosugi et al. 2009). Furthermore, although slightly divergent from the consensus sequence of bipartite cNLSs, it is not an unconventional when compared to all known bipartite NLS. Much longer bipartite sequences have been identified, such as the yeast Ty1 integrase NLS, which has a 29-amino acid linker sequence (Lange et al. 2010). Likewise, the human adenovirus E1A was shown to have a bipartite NLS \(K-X_4-R-X_3-RR-X_21-KRPRP\), with a highly disordered upstream element, composed of 11 total amino acid with four basic amino acids spread throughout (Cohen et al. 2014).

We also determined that the nuclear localization of Ime4 appears to be Slz1 and Kar4 independent, as the GFP-pC128 insert was still underwent nuclear import in slz1Δ/Δ and kar4Δ/Δ mutants. Since the nuclear import of Ime4 appears to also be Mum2 independent (Dehon 2012), it is likely that Ime4 does indeed possess an NLS, rather than using the piggy-back mechanism for nuclear import. However, it is likely that Ime4 employs the piggy-back mechanism to undergo subnuclear localization, through Slz1. It is worth noting that the downstream element of our putative NLS overlaps with the *in silico* prediction of a putative NoLS within Ime4 \(528^{\text{TSHTGTKIDKKQPSKLOQHQQYWNN}}_{556}\; (\text{Scott et al. 2010}). Whether this
downstream element in fact mediates nucleolar localization is still to be determined. It would be interesting to explore if the putative NoLS interacts in any way with Slz1 during nucleolar localization (Schwartz et al. 2013). This could lead to a greater understanding of the subnuclear localization of the MIS complex, and subsequently identifying the dynamics of the MIS complex members in relation to their subnuclear location.

Moreover, although residues 517-540 appear to code for a putative bipartite NLS, our experimental approach does not allow us to identify which residues are critical for nuclear import to occur. The GFP-pC059 insert only underwent weak nuclear localization, suggesting that the putative downstream element is needed for full localization; but because that construct does not include lysine 526, we cannot exclude the possibility that Ime4 may employ a much shorter bipartite signal, with sequence \text{RKYQEFMKS}^{526}, with residues 524-526 as a functional downstream element. However, comparison to the nucleoplasmin bipartite NLS suggests that, although possible, it would not be likely, as the nucleoplasmin NLS also contains two lysine residues within the linker sequence which are nonessential for nuclear localization to take place (Makkerh et al. 1996).

Thus, in future experiments, we would carry out site directed mutagenesis of residues within this general region particularly the basic residues to precisely identify the sequence of Ime4’s NLS. Furthermore, to confirm the sequence as a bipartite NLS, we would isolate residues 517-540 of Ime4 and insert it within the C-terminus of GFP; if GFP is observed to undergo nuclear localization, the sequence can conclusively be identified as an NLS. Moreover, we would do a negative control to analyze whether the rest of the protein [aa 1-527] is able to undergo nuclear localization when expressed with the \text{pRS416-ADH1p-GFP}; this would allow us to test for the possibility of an additional NLS being located in another region of Ime4. Finally, we could also
explore the weak nuclear import we observed after expression of GFP-pC073 and GFP-pC059 and determine whether, in fact, it was driven by the isolated upstream and downstream elements of the putative NLS, respectively.

To conclude, we have successfully identified residues 517-540 as critical for the truncated C-terminus of Ime4 to undergo nuclear localization. We also determined that the C-terminus of Ime4 is able to undergo nuclear localization in an Slz1 and Kar4 independent pathway, suggesting the necessity of an NLS within Ime4 to mediate nuclear import. Upon careful observation of the 517-540 sequence, and comparison to known NLSs, it appears that this sequence contains a novel variant of the bipartite cNLS.

**Materials & Methods**

**Yeast Strains**

All experiments were performed using *S. cerevisiae* strain *BY4741* unless explicitly stated. *kar4Δ/Δ* and *slzΔ/Δ* were also used to test for interaction of Kar4 and Slz1 with Ime4, respectively; these mutants were of *BY4741* background.

**Media**

Luria broth (LB) media was used for *E. coli* culture. Plates were made using 1.5% agar (0.015 g/ml), 1% bacto-tryptone (0.01 g/ml), 0.5% yeast extract (0.005 g/ml), 0.5% NaCl (0.005 g/ml), adjusted to a 7.5 pH. Ampicillin was used as a selectable marker for successful transformants (100 μg/ml). Culture media was made using the same ingredients excluding agar.

Yeast culture employed two different media: yeast extract peptone dextrose (YEPD) was used as a rich media, while synthetic complete (SC) media was used as selective media. YEPD liquid media included 1% yeast extract (0.01 g/ml), 2% peptone (0.02 g/ml) and 2% dextrose (0.02
g/ml); 2% agar (0.02 g/ml) was added for plate preparation. SC media was created with 10X Bacto-yeast nitrogen base (YNB) without amino acids (0.067 g/ml), Bacto Casamino acids (0.01 g/ml), 5 ml adenine (4 mg/ml), 2% dextrose (0.02 g/ml).

**Primer Design**

Desired primers for polymerase chain reaction (PCR) amplification of gene segments of IME4 were designed in the lab and were purchased from Integrated DNA Technologies (IDT) in a dry powder form; they were subsequently dissolved in deionized water to create a 100 μM solution. With the exception of pC059, each insert was PCR amplified using one set of forward and reverse primers. The same reverse primer sequence was used in all inserts, designed to anneal to the reverse complement strand of the 3’ UTR of IME4 and to have a Xho1 restriction site digest the oligonucleotide’s 5’ end; the forward primer sequence was chosen depending on what section of the C-terminus of Ime4 was desired to be expressed and was designed to have a BamH1 site at its respective 5’ end. Additionally, a linker sequence was included at the 5’ end of the forward primer, upstream from the BamH1 site; this allowed for the insert to have the correct open reading frame, in relation to GFP and to allow efficient digestion of the PCR products for cloning purposes.

The pC059 insert was a bipartite constructed created by fusing the coding region of interest (aa 468-525; NT 2401-2575) with the regulatory region of the extreme 3’ end (aa 600; nt 2798-3033). This was done by using two sets of primers: forward 1/reverse 1 and forward 2/reverse 2. Forward 1 is a standard primer, being the 5’ end of the sense strand of our coding region of interest; Reverse 1, though, was designed by joining the reverse complement sequence of nucleotides 2798-2815 (the stop codon region) to the 3’ end of the reverse complement of sequence 2558-2275 (the end of the desired coding region, aa 468-525).
For the second primer pair, Reverse 2 is a standard reverse primer, being the 5’ end of the antisense strand of the 3’ UTR of Ime4; forward 2, though, was designed to be the reverse complement of the 5’ end of reverse 1. This allowed the joint amplification of the two segments of DNA in a series of PCR reactions outlined below, in the “PCR amplification” subsection. A complete list of the primer sequences, as well as their amplicon nucleotide and amino acid coordinates is detailed below.

### Primers used for cloning Ime4 fragments

<table>
<thead>
<tr>
<th>Insert</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Amplicon Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>pC128</td>
<td>5’-TAC AAG GAT CCA TAG CAG AAC GGT TGG CAG G-3’</td>
<td>5’-AG CTT TCT CCA GGC GAC AAC-3’</td>
<td>NT: 2417-3033 AA: 473-600</td>
</tr>
<tr>
<td>pC105</td>
<td>5’-AG CTT TCT CCA GGC GAC AAC-3’</td>
<td>5’-AG CTT TCT CCA GGC GAC AAC-3’</td>
<td>NT: 2486-3033 AA: 496-600</td>
</tr>
<tr>
<td>pC073</td>
<td>5’-GGA ACC AGC CAC ACT GGT ACT-3’</td>
<td>5’-AG CTT TCT CCA GGC GAC AAC-3’</td>
<td>NT: 2582-3033 AA: 528-600</td>
</tr>
<tr>
<td>pC059</td>
<td>(Forward 1) 5’-ATC AGG ATC CGA TGA ACT GTA TGG TAT A-3’</td>
<td>(Reverse 1) 5’-TTA TTT AAC GCT TTA CTG GCT CTT CAT AAA CTC CTG -3’</td>
<td>NT: 2401-2575 AA: 468-525</td>
</tr>
<tr>
<td></td>
<td>(Forward 2) 5’-CAG TAA AGC GTT AAA TAA -3’</td>
<td>(Reverse 2) 5’-TTT TCT CGA GAG CTT TCT CCA GGC GAC AAC-3’</td>
<td>NT: 2798-3033 AA: 600</td>
</tr>
</tbody>
</table>

### Template Acquisition and Restriction Digest

The full-length IME4 template used was a vector bound insert found in a plasmid made by former graduate student in the Clancy lab, Jenisha Ghimire. The plasmid was linearized through a
20 μl restriction digest mixture in order to be usable in subsequent PCR amplification reactions. A typical restriction digest reaction included 3 μl of template (<1000 ng of DNA), 1 μl of EcoR1 restriction enzyme purchased from New England Biolabs, 2 μl of supplier provided restriction enzyme buffer and 14 μl of deionized water. The reaction was incubated at 37°C for approximately 30 minutes. Successful completion of the reaction was verified through TBE agarose gel electrophoresis.

**PCR Amplification**

Each insert was isolated through PCR amplification using High Fidelity Herculase II fusion Pfu polymerase, purchased from Agilent/Stratagene Technologies, and a linearized vector-bound full-length *IME4* template. Inserts pC128, pC105, pC073 were all amplified through a single standard PCR reaction. For each sample, a 50 μl reaction mixture was prepared using 0.5 μl of Pfu polymerase, 20 μl of associated polymerase buffer, 0.5 μl forward and reverse primer, respectively, 1 μl of deoxyribonucleotide triphosphates (10 mM; Promega) and 5 μl of linearized template DNA. After an initial denaturation step of 5 minutes at 95°C, the reaction was cycled from 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 60 seconds. This was repeated for 30 cycles.

Insert pC059 was prepared through a two-phase PCR amplification. In the first phase, two distinct PCR reactions were carried out: 1) uses primer pair forward 1/reverse 1 which amplified the the coding region of interest of *IME4* (nt 2401-2575, aa 468-525); 2) uses primer pair forward 2/reverse 2 which amplifies the 3’ end regulatory region of *IME4* (nt 2798-3033, aa 600). The PCR products were subsequently purified using TAE agarose gel purification (details in “PCR Purification” subsection).
In the second phase, 10 μl of each of the two PCR products from phase one were placed in a standard 50 μl PCR reaction, without the addition of primers. This allows for the fusion of the two separate PCR products, due to the reverse complementarity of primer reverse 1 and forward 2 (detailed in “Primer Design” subsection), creating a double-stranded DNA molecule. This reaction was allowed to run for 15 cycles. Subsequently, the double stranded DNA molecule was subjected to a normal PCR amplification, using primers Forward 1 and Reverse 2.

**PCR Product Purification**

Following PCR amplification, primers were removed and the PCR product was purified using TAE agarose gel purification or column PCR purification. The gel purification was done using the procedure and materials provided in the StrataPrep DNA Gel Extraction Kit (Catalog #400766). For column purification, the procedure and materials used for were those provided in the StrataPrep PCR Purification Kit (Catalog #400771).

**Plasmid Assembly**

The vector plasmid used, *pRS416-ADH1p-GFP*, was kindly provided by Zhengchang Liu. The vector included the *ADH1* promoter to drive transcription of an open reading frame (ORF) of the fluorescent marker, green fluorescent protein (GFP), followed by a pre-existing insert flanked by BamH1 and XhoI restriction sites. Additionally, the vector included the yeast selectable marker *URA3*, which codes for an essential enzyme in the *de novo* biosynthesis of pyrimidines; likewise, the AmpR gene was also included within the vector, encoding for a β-lactamase granting ampicillin resistance to *E. coli*. Finally, the yeast *CEN6/ARS4* cassette was also included, allowing replication of the plasmid within yeast, while the pMB1 origin was used to allow replication within *E. coli*.
The first step to include our desire insert was the removal of the pre-existing insert through a restriction digest using BamH1 and Xho1, on the 5’ and 3’ of the insert, respectively. The same restriction digest procedure outlined above was used, employing 2 μl of each restriction enzyme, BamH1 and Xho1, and the appropriate buffer. Following confirmation of the restriction digest with TBE gel electrophoresis, the vector was isolated through TAE agarose gel purification.

In a separate reaction, the desired purified PCR product underwent a parallel restriction enzyme digest with BamH1 and Xho1 to create DNA overhangs that would allow annealing with the vector. The digested insert was subsequently purified using agarose gel purification. Following TBE agarose gel confirmation, the digested vector and insert are ligated together. The 40 μl reaction mixture was composed of 10 μl of vector, 10 μl of insert, 4 μl of buffer, 0.5 μl of T4 DNA ligase (400 units/μl), and 15.5 deionized water. The reaction was allowed to incubate at room temperature for one hour and was subsequently incubated at 4°C overnight to allow for the enzyme to be active at multiple temperatures for optimal efficiency.

**E. coli Transformation**

Following the ligation reaction between the vector and insert, the plasmid DNA was transformed into XL1-Blue competent *E. coli* cells (Stratagene). 50 μl of competent cells were added to a chilled 12 ml plastic tube. 5 μl of plasmid DNA were added directly into the cells and the mixture was incubated in ice for 20 minutes. The tube was then heat shocked for 45 seconds in a 42°C water bath. The tube was then incubated for 2 minutes in ice, 1 ml of preheated SOC liquid media (5 ml LB, 50 ul 20% glucose, 50 ul 1 M MgSO₄) was added. The mixture was incubated for 30 minutes while shaking; it was finally plated on LB plates (150 μl/plate) and incubated at 37°C.
**Plasmid Isolation**

Plasmids were isolated using the mini-prep procedure and materials provided by the StrataPrep Plasmid Miniprep Kit (Catalog #400761). To confirm the integrity of the plasmid sequence, we submitted the plasmids for sequencing; to date, we are still pending to receive and process the results of the sequencing.

**Yeast Transformations**

The desired plasmid was transformed into cells at log phase, growing in liquid YPED media. The transformation was performed according to the Linda Hoskins/Hahn Lab Transformation method ([http://labs.fhrc.org/hahn](http://labs.fhrc.org/hahn)).

**Microscopy**

Transformants were observed using standard fluorescence microscopy using the Nikon® NIS-Elements Basic Research software. An excitation wavelength range of 465-495 nm and an emission wavelength range of 515-555 nm was used for GFP imaging. DAPI imaging was carried out using an excitation wavelength of 330-380 nm and a visualization wavelength of 435-485 nm.
References


31


Dehon, P. M. (2012). Localization and Mutational Analysis of the Nuclear and Aggregation-Prone Ime4 Protein in Saccharomyces cerevisiae.


