**DISTRIBUTION AGREEMENT**


In presenting this dissertation as a partial fulfillment of the requirements for an advanced

degree from Emory University, I hereby grant to Emory University and its agents the

non-exclusive license to archive, make accessible, and display my dissertation in whole

or in part in all forms of media, now or hereafter known, including display on the world

wide web.  I understand that I may select some access restrictions as part of the online

submission of this dissertation.  I also retain the right to use in future works (such as

articles or books) all or part of this dissertation.



_____          _____
Allison Lange                                                                    Date

Mechanisms of Nucleocytoplasmic Transport

By
Allison Lange
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Graduate Program in Biochemistry, Cell and Developmental Biology

_____
Anita H. Corbett, Ph.D.
Advisor

_____
Keith M. Berland, Ph.D.
Committee Member

_____
Ichiro Matsumura, Ph.D.
Committee Member

_____
Maureen A. Powers, Ph.D.
Committee Member

_____
Keith D. Wilkinson, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

_____
Date

**Mechanisms of Nucleocytoplasmic Transport**


By

Allison Lange
B.A., Dartmouth College, 2001

Advisor:  Anita H. Corbett, Ph.D.


An abstract of a dissertation submitted to the Faculty of the
Graduate School of Emory University in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Graduate Program in Biochemistry, Cell and Developmental Biology

2008

**ABSTRACT**

The genome of the eukaryotic cell is separated from the contents of the cytoplasm by the nuclear membrane. This division allows for precise regulation of cellular activity by providing controlled access to the nuclear interior; however, it also requires the existence of a highly specific mechanism of nuclear import to facilitate and manage this process. Accordingly, nuclear import factors recognize and bind protein cargo in the cytoplasm, traverse the nuclear membrane, release cargo into the nucleus, and return to the cytoplasm for further rounds of import. This dissertation focuses on the cargo recognition and release steps, while also delving into the state of import receptors when unbound by cargo.

Proteins destined for the nucleus contain nuclear localization signals (NLSs), which are recognized by soluble receptors called importins or karyopherins. The classical bipartite NLS has traditionally been defined as consisting of two stretches of basic amino acids separated by a linker of 10 amino acids. In Chapter 2, we find that this canonical consensus is artificially limiting, and that classical bipartite NLSs with linker regions of between eight and twenty residues can mediate nuclear import and interaction with importin $\alpha$. In Chapter 3, we investigate the recently proposed PY-NLS, which is recognized by Kap$\beta$2 and essentially consists of a hydrophobic or basic region upstream of an $R/H/KX_{2-5}PY$ motif, and show for the first time that this new NLS is functional *in vivo* and conserved across species. In Chapter 4, we transition into the nuclear interior and examine the cargo release step by validating *in vivo* the structures of importin $\alpha$ with two factors proposed to be involved in the delivery of classical NLS-containing protein

cargo: the export factor for importin α, Cse1/RanGTP, and the nucleoporin, Nup2.

Finally, in Chapter 5, we contemplate the state of the import receptor once it has been

recycled to the cytoplasm as it prepares for its next round of import.  We solve the crystal

structure of unbound importin β and show that the receptor assumes a compact, ring-like

conformation when not bound to cargo, similar to the free states of other characterized

transport receptors.

**Mechanisms of Nucleocytoplasmic Transport**

By

Allison Lange
B.A., Dartmouth College, 2001

Advisor: Anita H. Corbett, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of
Emory University in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Graduate Program in Biochemistry, Cell and Developmental Biology

2008

# CONTENTS

**FIGURES**

**TABLES**

## ABBREVIATIONS

| | |
|---|---|
| 5-FOA | 5-fluoroorotic acid |
| ARM | Armadillo |
| BSA | Bovine serum albumin |
| cNLS | Classical nuclear localization signal |
| DAPI | 4',6-diamidino-2-phenylindole |
| DIC | Differential interference microscopy |
| ELISA | Enzyme-linked immunosorbent assay |
| FG | Phenylalanine-glycine |
| GFP | Green fluorescent protein |
| GST | Glutathione S-transferase |
| HEAT | Huntingtin, elongation factor 3, A subunit of PP2A, TOR1 |
| IBB | Importin β binding |
| IPTG | Isopropyl-thio-galactopyranoside |
| KaRF | Karyopherin release factor |
| LB | Luria broth |
| NLS | Nuclear localization signal |
| NPC | Nuclear pore complex |
| PLAC | Pepstatin A, leupeptin, aprotinin, chymostatin |
| PMSF | Phenylmethylsulfonyl fluoride |
| RanGEF | Ran guanine exchange factor |
| RanGAP | Ran GTPase activating factor |
| RGG | Arginine-glycine-glycine |
| SAXS | Small-angle X-ray scattering |
| SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SV40 | Simian virus 40 |

# CHAPTER 1

## Introduction and Background

Portions of this chapter adapted from:

**Classical nuclear localization signals: definition, function, and interaction with importin α.**

Allison Lange[1], Ryan E. Mills[1], Christopher J. Lange[1], Murray Stewart[2], Scott E. Devine[1], and Anita H. Corbett[1]

[1]Department of Biochemistry, Emory University, School of Medicine, Atlanta, GA 30322
[2]MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

**Overview of Nucleocytoplasmic Transport**

In eukaryotic cells, the genetic material and transcriptional machinery of the nucleus are separated from the translational machinery and metabolic systems of the cytoplasm by the nuclear envelope. This segregation facilitates the precise regulation of cellular processes such as gene expression [1], signaling [2], and cell cycle progression [3] through selective regulation of bi-directional transport between the nucleus and the cytoplasm. However, this physical separation also necessitates the existence of molecular machinery that specifically recognizes cargo in one compartment, translocates it through the nuclear pore, and releases it in the other compartment. Nuclear transport systems of this kind were first proposed when a nuclear targeting signal in the Simian virus 40 (SV40) large-T antigen was characterized over twenty years ago [4, 5]. Since then, several pathways for nucleocytoplasmic transport have been described, of which, the classical nuclear protein import pathway is the best characterized.

Transport of macromolecules into and out of the nucleus occurs through large, proteinaceous structures called nuclear pore complexes (NPCs) [6-9]. NPCs allow passive diffusion of ions and small proteins (<40kDa), but restrict passage of larger molecules to those containing an appropriate targeting signal [10, 11]. NPCs perforate the nuclear membrane and have eight-fold radial symmetry, filaments that extend into the cytoplasm, and a basket structure that extends into the nucleus [7, 8]. The pores are constructed from a class of proteins called "nucleoporins", a subset of which contain a tandem series of phenylalanine-glycine (FG) repeats, that line the central transport channel of the pore [8, 12-14]. Transport through the pore is rapid and prolific. For

example, the nuclei of a dividing human cell hosts 3000-5000 NPCs, which mediate about 1000 transport events per second [15, 16].

The active transport of macromolecular cargo between the cytoplasmic and nuclear compartments is facilitated by specific soluble carrier proteins. These carriers are collectively referred to as "karyopherins" [17], with those involved in import and export termed "importins" [18] and "exportins" [19], respectively. This distinction may be artificial though, since in an ideal system, a transport factor would import one molecule, switch to export mode, and return to the cytoplasm carrying a different cargo. For example, both import and export cargos for the *S. cerevisiae* Msn5 receptor have been identified [20-24]. Many transport receptors are members of the importin β superfamily and are called collectively "β-karyopherins" [25]. Cargo proteins can bind directly to β-karyopherins; however, in classical nuclear import, the interaction between the β-karyopherin and the cargo is mediated by the adaptor molecule, importin α. Recent modeling studies have shown that addition of an adaptor to the system results in lower transport efficiency, but this feature also may allow for greater control over cargo accumulation levels [26, 27].

The energy for nuclear transport is provided by the small Ras-family GTPase, Ran [28]. Like other GTPases [29], Ran cycles between a GTP- and a GDP-bound state. The nucleotide state of Ran is modulated by regulatory proteins, primarily the Ran guanine nucleotide exchange factor (RanGEF) in the nucleus and the Ran GTPase activating protein (RanGAP) in the cytoplasm [30-33] (Figure 1.1). Because these key regulatory factors are compartmentalized, the different forms of Ran are asymmetrically distributed in the cell, with RanGTP enriched in the nucleus and RanGDP enriched in the

3
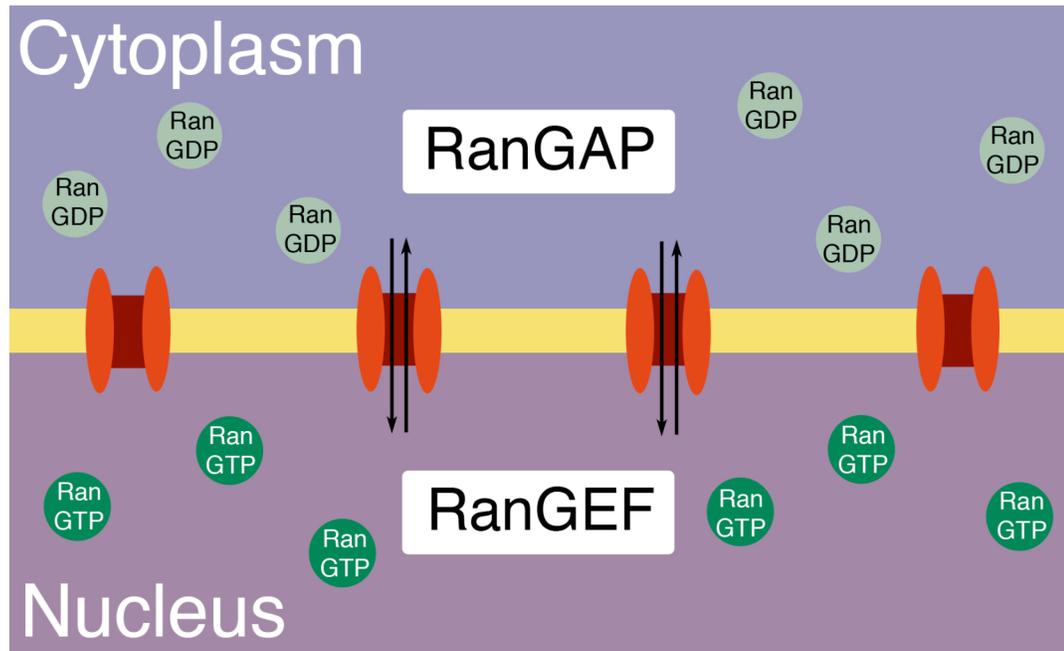
FIGURE 1.1 **The Ran gradient.** The proteins that regulate the Ran cycle are asymmetrically distributed in the cell, with the Ran <u>G</u>TPase <u>a</u>ctivating <u>p</u>rotein (RanGAP) in the cytoplasm and the Ran guanine nucleotide <u>e</u>xchange <u>f</u>actor (RanGEF) in the nucleus [30-33]. This distribution results in a predominantly cytoplasmic localization for RanGDP and a predominantly nuclear localization for RanGTP [34, 35].

cytoplasm [34, 35]. This compartmentalization allows Ran to impart directionality to nuclear transport by acting as a molecular switch that controls the binding and release of cargo. Therefore, import receptors bind cargo in the cytoplasm in the absence of RanGTP and release cargo in the nucleus upon RanGTP binding to the complex. In contrast, export receptors bind cargo in the nucleus in complex with RanGTP with hydrolysis to GDP in the cytoplasm triggering dissociation.

**The Classical Nuclear Import Cycle**

The best understood pathway of nucleocytoplasmic transport is the classical nuclear import pathway (Figure 1.2). Here, importin α recognizes and binds cargo in the cytoplasm, linking it to the β-karyopherin, importin β [36]. Importin β then mediates interaction of the trimeric complex with the nuclear pore as it translocates into the nucleus. Once the import complex reaches the nucleus, it is dissociated by RanGTP. Binding of RanGTP to importin β causes a conformational change that results in the release of the importin α/cargo complex [37, 38]. An auto-inhibitory region on the importin β-binding (IBB) domain of importin α [39, 40]; the nucleoporin Nup2 (Nup50 or Npap60 in vertebrates) [41-43]; and the export receptor for importin α, Cse1/RanGTP (CAS/RanGTP in vertebrates) [44], then work together to deliver the cargo into the nucleus. Finally, Cse1/RanGTP recycles importin α back to the cytoplasm in preparation for another round of import [45, 46].
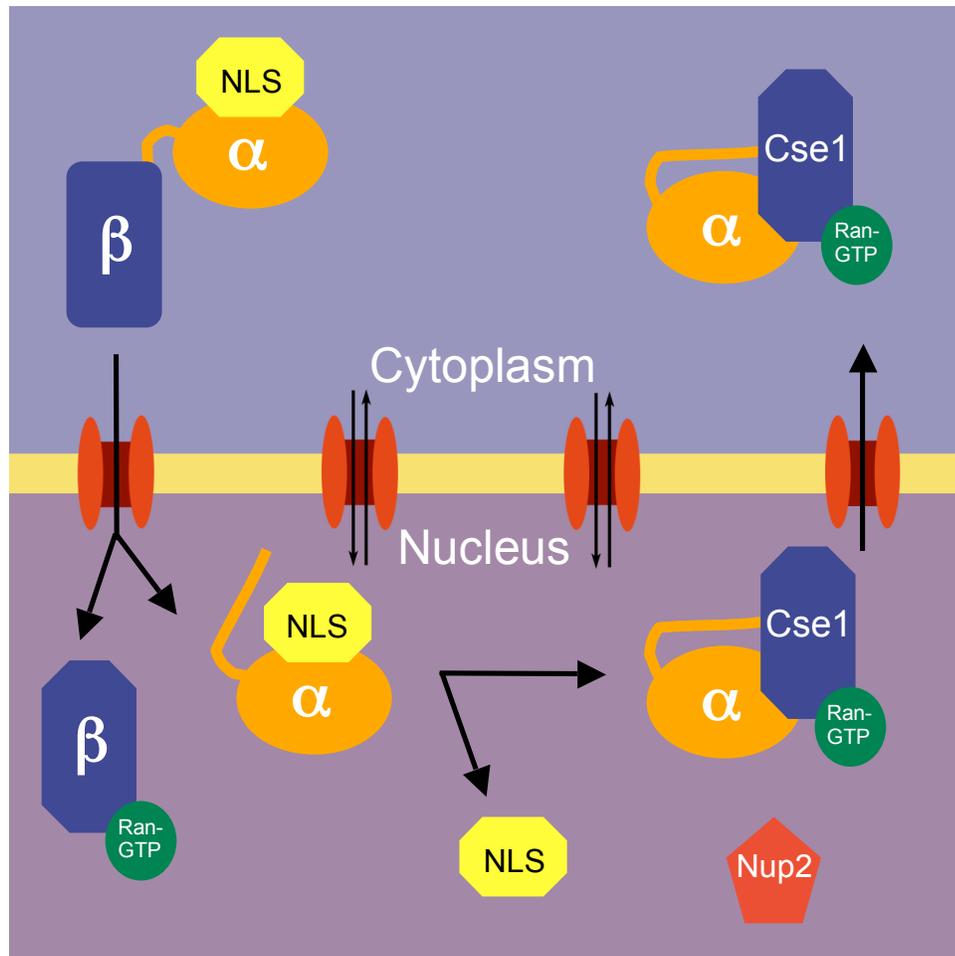
FIGURE 1.2  **The classical nuclear import cycle.**  In the cytoplasm, cargo containing a cNLS is bound by the heterodimeric import receptor, importin α/importin β [36, 47]. Importin α recognizes the cNLS and importin β mediates interactions with the nuclear pore during translocation.  Once inside the nucleus, RanGTP-binding causes a conformational change in importin β, which releases the IBB region of importin α.  This auto-inhibitory domain, together with Nup2 and Cse1, facilitates cNLS dissociation and delivery of the cNLS-cargo in the nucleus [40, 41].  Finally, importin α is recycled back to the cytoplasm by the export receptor, Cse1, in complex with RanGTP [45, 46].

**The Classical Nuclear Localization Signal**

The first step of nuclear import occurs when an importin discriminates between its cargo and other cellular proteins. Proteins destined for transport into the nucleus contain amino acid targeting sequences called <u>n</u>uclear <u>l</u>ocalization <u>s</u>ignals (NLSs). The best-characterized transport signal is the classical NLS (cNLS) for nuclear protein import, which consists of either one (monopartite) or two (bipartite) stretches of basic amino acids [4, 48, 49]. Monopartite cNLSs are exemplified by the SV40 large-T antigen NLS ([126]PKKKRRV[132]) and bipartite cNLSs are exemplified by the nucleoplasmin NLS ([150]<u>KR</u>PAATKKAGQA<u>KKK</u>[170]). Consecutive residues from the N-terminal lysine of the monopartite NLS are referred to as P1, P2, etc. Structural [50, 51] and thermodynamic [52] studies have identified many of the key requirements for a cNLS. These studies show that a monopartite cNLS requires a lysine in the P1 position, followed by basic residues in positions P2 and P4 to yield a loose consensus sequence of K(K/R)X(K/R). The established consensus sequence for a bipartite cNLS consists of two basic residues separated from at least three basic residues out of five by a ten amino acid linker [53]. However, importin α has been crystallized with peptides corresponding to the cNLS of human retinoblastoma protein, a classical bipartite NLS with a linker of 11 amino acids, and the cNLS of *Xenopus laevis* phosphoprotein N1N2, a bipartite cNLS with a linker of 12 residues [54], calling into question the validity of limiting the bipartite cNLS linker to 10aa. Work presented in Chapter 2 addresses this discrepancy and experimentally defines the boundaries of the bipartite cNLS linker.

Alanine scanning of the Myc, monopartite SV40, and artificial bipartite SV40 cNLSs revealed that the binding affinity of a cNLS for importin α measured *in vitro* correlates with the steady-state nuclear accumulation and import rate of the corresponding cNLS-cargo *in vivo* [52, 55]. These data establish a range for the binding capacity of a cNLS, since a functional cNLS has a binding constant of ~10 nM for importin α and a non-functional cNLS has a binding constant of ~1μM. Interestingly, non-functional near-cNLS sequences can either bind importin α too weakly to be effectively imported or too tightly to be efficiently released from the receptor in the nucleus, though karyopherin release factors such as Nup2 and Cse1 can aid in this cargo release step [41-44, 56]. Recent computer simulations also indicate that the import rate of cNLS-cargo depends strongly on the rate of formation of the import complex [26]. Therefore, the import and accumulation of cNLS-cargo in the nucleus is affected by both the affinity of the cNLS-cargo for importin α and by the concentration of the importin α receptor itself [26, 55]. This relationship is also characteristic of other NLS-cargo/receptor interactions [57].

**The Importin α/cNLS Interaction**

The molecular basis for recognition of a cNLS by importin α has been defined using x-ray crystallography techniques [50, 51, 58, 59] (Figure 1.3). These structural studies reveal that importin α is comprised of a large domain consisting of ten armadillo (ARM) motifs, each of which is constructed from three α-helices, and a flexible N-
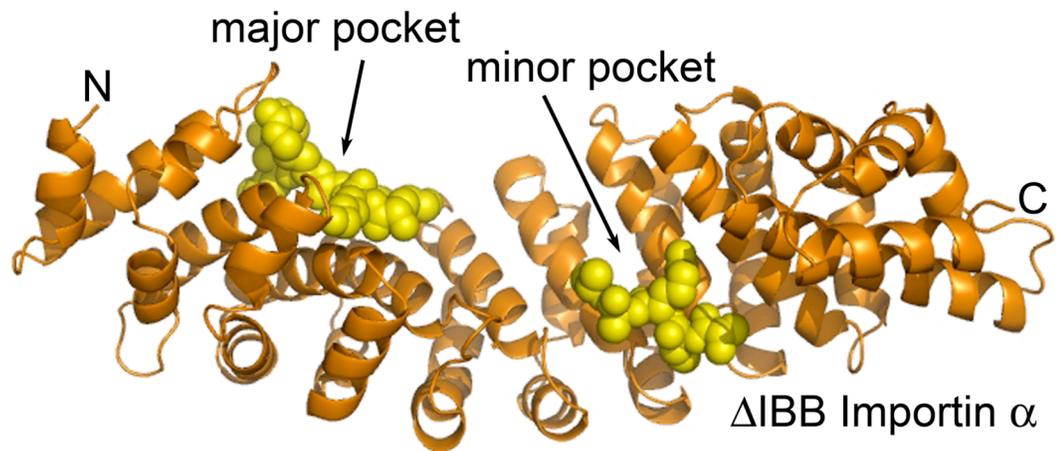
FIGURE 1.3 **The structure of importin α bound to classical NLS.** cNLS peptides bound to the major and minor binding pockets of *S. cerevisiae* importin α lacking the IBB domain [58]. Importin α (amino acids 88-530) is shown in an orange ribbon diagram. Two SV40 peptides bound to the major and minor binding pockets are shown in a yellow, space-filling model.

terminal domain (the IBB domain) required for both binding to importin β [36] and cargo dissociation [39, 60, 61].  The regular sequence of ARM repeats generates a gently curving, elongated molecule where the major and minor cNLS-binding sites are located within a shallow groove on the concave face.  Both pockets are formed by solvent-exposed, conserved tryptophans together with a set of invariant asparagines four residues downstream [39, 58].  The major pocket, which lies nearer the N-terminus, binds both monopartite cNLSs and the larger stretch of basic residues in bipartite cNLSs.  The minor pocket is more C-terminal and binds the smaller stretch of basic residues in bipartite cNLSs.  cNLS motifs bind to the NLS-binding pockets in an extended conformation with their main chains running anti-parallel to the direction of the importin α chain.  The key cNLS lysine side chains then lie between the stacked hydrophobic indole side chains of the conserved tryptophans and form salt bridges with negatively charged residues lining the binding groove, while the asparagines make key main chain contacts [39, 58].

**Dissociation of cNLS-Cargo from Importin α**

Once cargo molecules are recognized in the cytoplasm and translocated into the nuclear interior, they must be released in the nucleus before the receptors can be recycled for another round of transport.  For all modes of karyopherin-mediated import, RanGTP plays a major role in this complex dissociation.  However, in classical nuclear import, the presence of the adaptor molecule, importin α, adds to the complexity of the cargo delivery step.  Understanding how cargo is dissociated from importin α is important for two reasons.  First, a cNLS-cargo must be freed from importin α in order to carry out its

function in the nucleus [62]. Second, the export receptor Cse1 cannot interact with importin α when cNLS-cargo is bound [46, 63, 64], ensuring unidirectional import of cNLS-cargo and preventing futile transport events. Any defects in cargo release prevent importin α from being recycled to the cytoplasm and dramatically hinder classical nuclear import. Four factors have been implicated in this cargo release step: RanGTP; the auto-inhibitory region or IBB domain of importin α; the nucleoporin Nup2; and the export receptor for importin α, Cse1.

*RanGTP:* Once the cNLS-cargo/importin α/importin β trimeric complex translocates into the nucleus, it encounters the nuclear pool of RanGTP. Ran in the GTP-bound state binds to importin β, causing a conformational change [41, 65, 66] that increases the curvature of HEAT repeats 1-3 and decreases the curvature of HEAT repeats 12-19 relative to the IBB/importin β structure. The resulting C-terminal displacement of over 20Å frees the IBB domain and leaves a transient cNLS-cargo/importin α complex. Given that the interaction interface between importin β and the IBB domain is so extensive, it has been proposed that RanGTP actively "unzips" the importin α/importin β interface causing step-wise displacement [37]. RanGTP also binds sites in both the N- and C-termini of importin β, allosterically locking it in a position unable to bind cargo [37]. Therefore, RanGTP acts as a switch, modulating the flexibility and binding ability of importin β.

*Importin α IBB domain:* Structural studies first suggested that the newly freed IBB domain of importin α could modulate the interaction between importin α and cNLS-cargo by forming an intramolecular interaction with the major cNLS-binding pocket on importin α [39, 50, 58]. This competition is achieved through a cNLS-like sequence

within the IBB domain ($^{54}$KRR$^{56}$) that is essential for efficient release of cNLS-cargo into the nucleus [40, 61], supporting a model where the IBB domain serves as a regulatory switch between the cytoplasmic form of importin α, which has a high affinity for cNLS-cargo, and the nuclear form of importin α, which has a low affinity for NLS-cargo (Figure 1.4). In this model, importin β binds to the IBB domain of importin α in the cytoplasm, sequestering the auto-inhibitory region away from the cNLS-binding pocket and allowing cNLS-cargo to bind to importin α with high affinity [39, 60, 67]. When the import complex enters the nucleus, the RanGTP-triggered conformational change in importin β leads to release of the IBB domain, allowing it to fold over and bind to the major cNLS-binding pocket on importin α. Since both monopartite and bipartite cNLSs bind to the major pocket, the auto-inhibitory region affects the binding affinity of both types of cNLS, preventing re-binding and, therefore, facilitating the release of cNLS-cargo within the nucleus.

*Nup2 (Nup50/Npap60 in vertebrates):* Nup2 is a natively unfolded nucleoporin with an N-terminal importin α-binding domain, a central FXFG repeat domain that binds importin β and the nucleoporin Nup60, and a C-terminal RanGTP-binding motif [68-72]. Nup2 has putative vertebrate homologues, with Nup50 in humans and <u>n</u>uclear <u>p</u>ore <u>a</u>ssociated <u>p</u>rotein Npap60 in mouse having similar domain organization and sequence to Nup2 [73, 74]. Nup50/Npap60 can also functionally replace Nup2 in *S. cerevisiae* cells where Nup2 is essential [J. Hood, personal communication]. Studies have shown that Nup2 directly binds to importin α and that importin α accumulates in the nucleus of cells lacking Nup2 [70, 71, 75]. These studies implicate Nup2 in cNLS-cargo release in the nucleus because importin α cannot be exported by Cse1 while cargo-bound. Therefore, it
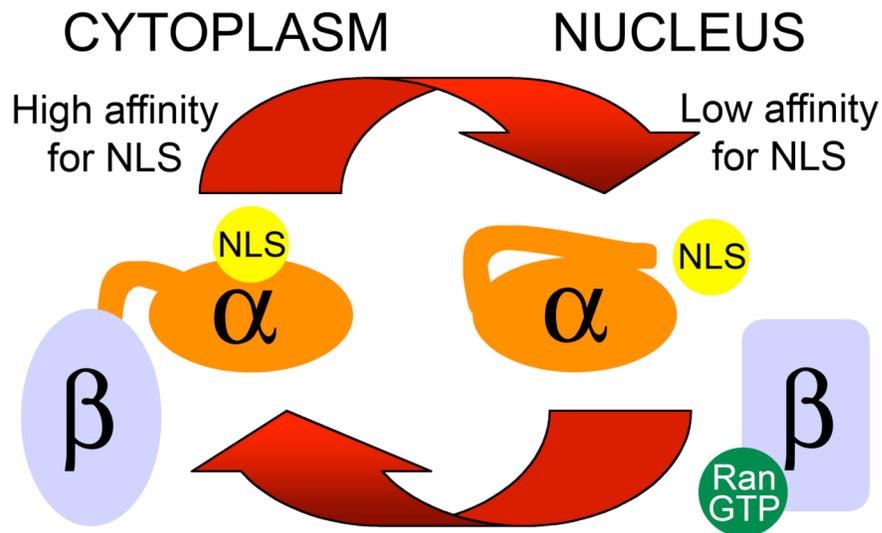
FIGURE 1.4  **The regulation of cNLS-cargo binding to importin α.**  In the cytoplasm, importin β is bound to the flexible auto-inhibitory domain of importin α, sequestering it from the NLS-binding pocket and allowing cNLSs to bind to importin α with high affinity [39, 60, 67].  Once in the nucleus, RanGTP binding to importin β causes a conformational change that releases the auto-inhibitory domain, which can then compete for binding to the NLS-binding pocket.  This competition contributes to a low affinity of the cNLS for importin α in the nucleus, facilitating cNLS-cargo delivery [40, 61].

is possible that the nuclear accumulation of importin α in *Δnup2* cells results from a loss

of the contribution of Nup2 to cargo release, causing accumulation of cargo-bound, and

therefore nuclear-bound, importin α complexes.  Genetic studies have also shown that

combining a deletion of the *NUP2* gene with an auto-inhibitory mutant of importin α

exacerbates the cold-sensitive phenotype of the importin α mutant, suggesting a potential

functional overlap [61].  Additionally, *in vitro* biochemical studies have shown that Nup2

and the related protein Nup1 can facilitate the dissociation of cNLS-cargo from importin

α [41].  The details of how Nup2 participates in protein cargo delivery in the nucleus is

unknown; however, our work to uncover the molecular mechanism underlying the

contribution of Nup2 to cNLS-cargo delivery is presented in Chapter 4.

 *Cse1 (CAS in vertebrates):*  Cse1, the importin α export receptor, is a member of

the β-karyopherin family of transport receptors [64, 76].  It is constructed from twenty

tandem HEAT (<u>h</u>untingtin, <u>e</u>longation factor 3, <u>A</u> subunit of PP2A, <u>T</u>OR) repeats and has

an N-terminal Ran-binding domain.  Prior to this dissertation, the interaction interface

between importin α and Cse1 had not been well-defined; however, two-hybrid

experiments involving the human homolog of Cse1, CAS, had suggested that importin α

interacts with the C-terminus of Cse1 [63].  As mentioned, the interaction between Cse1

and importin α only occurs once the cNLS-cargo has been released [46, 64], making

certain that importin α can only be recycled back to the cytoplasm once it has delivered

its cargo and, consequently, ensuring unidirectional import of cNLS-cargo.  The

mechanism for how Cse1 discriminates between the cargo-bound and cargo-free forms of

importin α was unknown prior to this dissertation, though we propose a "clamp" model

where the flexible IBB domain of importin α, in addition to the C-terminus, interacts

with Cse1, holding importin α in a closed state where the auto-inhibitory region is bound to the cNLS-binding pocket. This interaction could be involved in the actual displacement of cNLS-cargo and would also allow Cse1 to sense the presence or absence of cargo. Genetic data shows that mutations in Cse1 cause synthetic fitness defects when combined with mutations in the auto-inhibitory domain of importin α [61] and biochemical data shows that Cse1 can facilitate cNLS-cargo release from importin α *in vitro* [41], suggesting that Cse1 actively contributes to cNLS-cargo delivery. Chapter 4 describes our detailed analysis from structure to function of the likely role that Cse1 plays in cNLS-cargo release within the nucleus.

**The β-Karyopherin Family of Receptors**

As described, importin α binds to each of its cNLS-bearing cargos using a well-characterized and consistent mechanism. This rigid adherence to one motif reflects the fact that importin α has only one role, to serve as an adaptor between cNLS-cargo and importin β. In contrast, β-karyopherins not only must bind multiple different classes of cargos, but also must mediate interactions with the nuclear pore and respond to regulation by RanGTP [77, 78]. To integrate these diverse functions, importin β family members possess a number of binding options for a range of interactants. Recently, this binding flexibility and the question of how importin β directly interacts with its partners has been addressed by a number of structural studies. Six proteins have been co-crystallized with the importin β receptor (a fragment of importin α [79]; RanGTP [65]; the FxFG containing nucleoporins, Nsp1 [80] and Nup1 [81]; parathyroid hormone-related protein,

PTHrP [82]; and sterol-responsive element-binding protein 2, SREBP-2 [83]) and each interacts in a distinct manner.

Importin β is constructed from 19 tandem HEAT repeats with the A helices lining the convex outer face of the protein and the B helices lining the concave interior, providing an extensive interface that can be used to bind different cargos [79]. When bound to the positively charged N-terminal domain of importin α, importin β adopts a helicoidal, snail-like conformation with its highly acidic interior face wrapped around importin α. HEAT repeats 7-11 interact with an extended N-terminal moiety on importin α and HEAT repeats 12-19 interact with a more C-terminal α helix [79]. The non-classical NLS of PTHrP also binds to the concave surface of importin β, even overlapping the importin α-binding region in HEAT repeats 7-11, but it adopts an extended conformation, covering HEAT repeats 2-11 and adopting a different architecture [82]. In contrast, when five FxFG repeats from the nucleoporin Nsp1 are bound to importin β, they interact with the convex face on HEAT 5-6 and HEAT 7-8 [80]. Interestingly, this structure suggests a mechanism for RanGTP to release importin β from nucleoporins by moving the A helix of HEAT 5 relative to HEAT 6, causing a conformational change that obstructs the interaction site. Nup1 binds to importin β in a manner similar to Nsp1, burying phenylalanine residues into hydrophobic pockets between HEAT repeats 5, 6, 7, and 8. Hydrophobic residues adjacent to the phenylalanines and an additional linker region also contribute to binding, explaining the high affinity of Nup1 for importin β relative to other FxFG-containing nucleoporins [81].

The binding site for SREBP-2 on importin β is completely different to that for importin α, PTHrP, Nsp1, or Nup1. A dimer of SREBP-2 is held between HEAT 7 and

HEAT 17.  These particular HEAT repeats have long helices that hold the helix-loop-helix leucine zipper domain of SREBP-2 in a manner that has been equated to a pair of chopsticks [83].  In order to adopt this conformation, importin β must move into a more elongated and twisted shape than is seen in the importin α/importin β structure.  Lee et al. suggest that this apparent flexibility may be the key to the successful promiscuity of importin β [83].  Accordingly, members of the importin β family can be thought of as springs [84-86], where each HEAT repeat represents a coil and individual coils would only need to make small changes to cumulatively result in overall sizeable conformational changes.  These small changes in the relative position of different helices can also be paired with more global, hinge-like movements involving larger sections of the protein [87] to facilitate cargo recognition and interaction.  In this way, apparently prohibitively high binding energies can be overcome by small compensatory distortions by the molecules.  Small angle x-ray scattering (SAXS) data [88] supports this idea by showing extreme conformational variation between different β-karyopherin family members ranging from elongated S-shaped to globular.  One question left to answer is what the importin β family receptors look like in an unbound state, free from cargo.  These base structures could provide insight into how these molecules manage to so effectively bind cargo in one compartment and release it in another, constantly manipulating the switch between high and low affinity states.  In Chapter 5, we begin to answer this question by solving and analyzing the crystal structure of importin β in its unbound state.

Examining the diverse ways in which importin β binds to its cargos highlights the difficulty in establishing a consensus sequence for NLSs recognized by members of the

β-karyopherin superfamily. One of the few β-karyopherin consensus sequences described to date is the loosely-defined, leucine-rich nuclear export signal (NES) consensus sequence recognized by the export factor, Crm1, which consists of three or four hydrophobic residues [89, 90]. Additionally, in Chapter 3, we present studies examining the novel PY-NLS, which is recognized by the human β-karyopherin import receptor, Kapβ2 or transportin. Future work will need to focus on rigorously validating cargos for each transport receptor so that patterns to their NLS-recognition mechanisms can be elucidated. Establishing new consensus sequences for the NLS motifs recognized by these receptors will allow us to predict more precisely than ever before how the nuclear proteome is established and may even allow for precise targeting of these pathways for therapeutic purposes.

**Demonstrating that an NLS is Functional *in vivo***

In order for a putative NLS to be deemed a functional targeting sequence, it must meet four criteria [91]. First, the sequence must be necessary for import, meaning that transport of the protein of interest into the nucleus is dramatically hindered when the sequence is deleted or, preferably, altered. The most common approach to test whether a sequence is necessary is to mutate one or often more of the consensus residues to alanine and determine whether the nuclear localization of the protein decreases. Second, the sequence must be sufficient to target an unrelated protein to the nucleus. Typically, the suspected sequence is fused to the terminus of GFP that most closely mimics the native location of the NLS in the protein of interest and the location of the reporter protein is

assessed visually. Third, the protein of interest must directly interact with its putative import receptor and this interaction must be mediated by the identified sequence. The best way to address this particular question is to perform direct *in vitro* binding experiments with purified proteins in the presence of either RanGDP or RanGTP. Import complexes are dissociated by RanGTP, therefore in these binding assays, the protein of interest should interact with its import receptor in the presence of RanGDP, but not in the presence of RanGTP. Fourth, to determine which import pathway the protein of interest utilizes *in vivo*, it should be demonstrated that disabling the nuclear transport machinery in question disrupts import of the protein. For example, in *Saccharomyces cerevisiae*, the localization of the protein of interest could be assessed at the restrictive temperature in a strain containing a temperature-sensitive allele of the essential gene encoding the β-karyopherin. Unfortunately, testing the import pathway is much more complicated in non-yeast systems since specific inhibitors of nuclear import have not yet been developed. After a putative sequence has met these requirements, then it may be deemed a functional NLS for the protein of interest. However, it is important to note that a specific protein may have more than one nuclear targeting sequence that may be used by different pathways or respond to different types of signals, highlighting the complexity and intricacy of nuclear transport *in vivo*.

**Scope of this Dissertation**

As we have outlined, in the classical nuclear import pathway, importin α/importin β binds to proteins containing classical NLSs that consist of either one (monopartite) or

two (bipartite) stretches of basic amino acids. In the archetypal bipartite cNLS, the two basic regions are separated by a linker of 10aa; however, based on structural data suggesting that proteins with bipartite cNLSs containing linkers of up to 12aa can bind importin $\alpha$ [54], we hypothesize that the linker region of classical bipartite NLSs can vary from the canonical 10 residues. Using a combination of *in vivo* localization and *in vitro* binding studies, we find that bipartite cNLSs with linkers of 8-20 residues can bind importin $\alpha$ and effect import into the nucleus. Using a revised consensus sequence for the classical bipartite NLS, we then use a bioinformatics approach to probe the *S. cerevisiae* proteome and determine the prevalence of putative classical nuclear protein cargos in yeast.

The mechanisms of recognition of non-classical cargos are much less well-understood than the recognition of cNLS-cargos. Recently, however, the consensus sequence for the NLS recognized by a human importin, Kap$\beta$2 or transportin, was proposed. Based on the crystal structure of Kap$\beta$2 bound to the NLS of one of its cargos paired with *in vitro* binding experiments, Chook and colleagues defined the requirements for a PY-NLS: it is disordered, has a net basic charge, and has a hydrophobic or basic region upstream of a C-terminal $R/H/KX_{2-5}PY$ motif [92]. The apparent ortholog of Kap$\beta$2 in *S. cerevisiae* is Kap104 [93], which has two known cargos, the mRNA-binding proteins Nab2 (<u>n</u>uclear poly-<u>a</u>denylated RNA-<u>b</u>inding protein 2) and Hrp1 (<u>h</u>eterogeneous nuclear <u>r</u>ibonucleo<u>p</u>rotein 1). We uncover putative PY-NLS-like sequences within the primary sequence of both proteins and, using a combination of localization, binding, and functional studies, show that the PY-NLS within Hrp1 is necessary and sufficient for nuclear import of the protein and is required for protein

function; however, the less-conserved PY-NLS-like sequences within Nab2 are not required for Nab2 function or nuclear accumulation, suggesting that Kap104 may interact with its cargos using a variety of interface mechanisms. Importantly, these results demonstrate for the first time that the PY-NLS is functional *in vivo* and evolutionarily conserved across species.

Once an import complex reaches the nuclear interior, it must be dissociated so the protein cargo can carry out its duty in the nucleus and the receptor(s) can be recycled back to the cytoplasm. As described earlier, in most systems of nuclear import, the dissociation step is coordinated by the small GTPase, Ran. In the classical nuclear import pathway, Ran is aided by three other factors: a well-characterized auto-inhibitory region on importin α; the export factor for importin α, Cse1; and the nucleoporin, Nup2. The crystal structures of Cse1/RanGTP/importin α and Nup2/importin α have recently been solved [42-44], which provide detailed predictions for how Cse1 and Nup2 contribute to cNLS-cargo delivery in the nucleus. We aim to test the validity of these structural models *in vivo* by creating amino acid substitutions in residues predicted to be critical for complex formation and assaying their effects on protein localization, function, and interaction. We find that an interaction between the N-terminus of importin α and Cse1 is critical for the role of Cse1 in cargo release, which supports a "clamp" model of cNLS-cargo release where Cse1 holds importin α in a "closed" state, which is incompatible with cNLS-cargo re-binding. We also show that two regions on Nup2 are required for the function of Nup2 and for the ability of Nup2 to interact with importin α *in vivo*. One region interacts with the C-terminus of importin α and one region interacts with a cNLS-binding pocket within importin α. These interactions suggest that Nup2 may act

analogously to the auto-inhibitory domain, folding over the body of importin α and competing for binding with the cNLS-cargo, thereby aiding in the cargo release process.

Once an import receptor delivers its cargo in the nucleus, it is recycled to the cytoplasm to prepare for another round of import. The state of the import receptor in this pre-transport circumstance is poorly characterized; previous studies have mainly focused on the cargo-bound conformation of import receptors. To address this dearth, we collaborated to solve the crystal structure of the classical importin β receptor in a cargo-free state. We show that the molecule forms a highly compact ring structure, which is achieved through interactions in a hinge region of the protein and through interactions between the two termini of the protein. Amino acid substitutions to residues predicted to be critical for formation or maintenance of the circular structure have no discernable effect on the function of importin β or the import of a model classical cargo, though one importin β variant does show altered localization *in vivo*. These results suggest that the structure of free importin β is probably dependent on a series of interactions, any one of which can be altered without significantly disrupting the conformation or function of the protein.

Overall, this work focuses on teasing apart the elaborate processes involved in the mediation and regulation of transport into and out of the nucleus. By coupling detailed analyses of structures with thorough *in vivo* investigations, we advance the state of the field from the cartoon presented in Figure 1.2 to a structural understanding of the mechanisms that underlie nucleocytoplasmic transport, moving from abstract to crystal clear.

**CHAPTER 2**

**Redefining the Classical Nuclear Localization Signal (cNLS) and Exploring the Prevalence of cNLS-Containing Protein Cargo in *Saccharomyces cerevisiae***

This chapter is adapted from the following two papers:

**The length of the linker region in classical bipartite nuclear localization signals affects nuclear localization.**

Allison Lange*, Laura M. McLane*, Ryan E. Mills, Scott E. Devine, and Anita H. Corbett

Department of Biochemistry, Emory University, School of Medicine, Atlanta, GA 30322
*These authors contributed equally to the work.

**Classical nuclear localization signals: definition, function, and interaction with importin $\alpha$.**

Allison Lange[1], Ryan E. Mills[1], Christopher J. Lange[1], Murray Stewart[2], Scott E. Devine[1], and Anita H. Corbett[1]

[1]Department of Biochemistry, Emory University, School of Medicine, Atlanta, GA 30322
[2]MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

**Introduction**

In order for a protein to enter the nucleus, it must contain a nuclear localization signal (NLS), which is recognized by a class of soluble receptors called importins or karyopherins [17-19]. In the most well-characterized system of nuclear import, the classical import pathway, cargo proteins containing classical NLSs (cNLSs) are bound by the heterodimeric protein import receptor, importin α/importin β (reviewed in [53]). Importin α recognizes the cargo and importin β liaises with the nuclear pore [94]. Importin α consists of two main domains: a flexible N-terminal importin β-binding (IBB) domain, which contains an auto-inhibitory sequence that contributes to NLS-cargo delivery [40, 61], and a large domain that mediates binding to cNLSs. The large domain is comprised of ten armadillo repeats, which generate a gently curved molecule with two cNLS-binding pockets located in a shallow groove on the concave face [50, 51, 54, 58]. Monopartite cNLSs, which consist of one stretch of basic amino acids, bind in the major NLS-binding pocket of importin α. Structural [50, 51] and thermodynamic [52] studies have yielded a loose consensus sequence for the monopartite cNLS, which is K(K/R)X(K/R). Bipartite cNLSs, which consist of two stretches of basic amino acids separated by a linker region, bind to both the major and the minor NLS-binding pockets on importin α. The bipartite linker has traditionally been limited to 10aa based on the spacing found in a co-crystal structure of the bipartite cNLS of nucleoplasmin bound to ΔIBB importin α [50, 51], but bipartite cNLSs with linkers as long as 12aa have been shown to bind importin α *in vitro* [54]. Additionally, atypical bipartite cNLSs with much longer linkers have been proposed (for example, Smad4 [95], topoisomerase II [96], and

24

Ty1 integrase [97, 98]), though no rigorous studies have shown that these sequences are true bipartite cNLSs that actually mediate classical import *in vivo* [53].

The classical NLS is often thought of as the prototypical NLS. It was the first NLS to be characterized and, as such, many examples of proteins using the classical import pathway have been characterized. Because of the surfeit of known cNLS-containing proteins, many have assumed that this pathway is the most prevalent in the cell; however, no studies have established empirically the proportion of cargos imported via this mechanism. It is possible that other pathways account for a large amount of nuclear traffic and that examples are lacking simply because the methods for quick and easy identification of non-classical NLSs are lacking. To address this issue, we have taken advantage of the well-annotated *S. cerevisiae* genome and known consensus sequences for monopartite and bipartite cNLSs used by the predictive program, PSORT II, to estimate the prevalence of classical NLS-containing proteins using a bioinformatics approach.

The PSORT II consensus for bipartite cNLSs limits the length of the linker region to 10aa. Given the evidence suggesting that the linker length can vary, we aimed to define the boundaries for the length of the linker region in a bipartite cNLS. We created bipartite cNLS reporter proteins with linkers of various lengths, assessed their ability to mediate classical nuclear localization *in vivo* and their ability to bind importin α *in vitro,* and found that cNLSs with linker lengths other than 10aa are functional. We used this data to expand the conventional bipartite consensus sequence and conclude with a revised estimate of the prevalence of classical nuclear import in yeast. As suspected, the classical system is predicted to be the most prevalent in the cell, with about 60% of yeast proteins

potentially containing bipartite cNLS motifs; however, this plethora of prospective classical cargos does not preclude other mechanisms of nuclear transport from significantly contributing to the establishment the nuclear proteome.


**Experimental Procedures**


*Strains, Plasmids, and Chemicals*- All chemicals were obtained from US Biological or Sigma unless otherwise noted.  All media was prepared and all DNA manipulations were performed according to standard procedures [99, 100].  All yeast strains and plasmids used in this study are described in Table 2.1.

*Bioinformatics*- The algorithms from PSORT II [101] for monopartite and bipartite classical NLSs were used with ScanProsite [102] to search three sets of data: the 5850 proteins in the *S. cerevisiae* GenBank [103], the 1515 proteins localized to either the nucleus or the nucleolus in a comprehensive subcellular localization study performed with a global yeast GFP-fusion library [104], and the 224 proteins that interact with importin α according to the BioGRID database [105].  To determine the prevalence of putative cargos that contain a classical bipartite NLS with a linker of non-standard length, the PSORT II algorithm for bipartite NLSs was modified to have a linker of 8-9, 11-13, 14-15, 16-20, or 21-30 residues.

*Microscopy*- GFP-fusion proteins were localized in live cells using direct fluorescence and differential interference contrast (DIC) microscopy.  The GFP signal was visualized using a GFP-optimized filter (Chroma Technology) on an Olympus BX60 epifluorescence microscope equipped with a Photometrics Quantix digital camera.  Cells

TABLE 2.1 **Strains and plasmids used in Chapter 2**

| Strain/Plasmid | Description | Origin |
|---|---|---|
| FY23 (ACY192) | Wild type, MATa *ura3-52 leu2Δ1 trp1-63* | [106] |
| NOY672 (ACY1563) | *srp1-54,* MATa *ura3-1 leu2-3 trp1-1 his3-11 ade2-1 can1-100* | [107] |
| pAC493 | ΔIBB importin α, pProEX-HTB bacterial expression vector, *AMP$^R$* | [58] |
| pAC781 | GFP*,* pET28a bacterial expression vector, *KAN$^R$* | [60] |
| pAC1059 | Bipartite T3 SV40-GFP-GFP, *MET25* promoter, *CEN URA3 AMP$^R$* | [108] |
| pAC1069 | GFP-GFP, pET28a bacterial expression vector, *KAN$^R$* | [55] |
| pAC1306 | Bipartite T3 SV40-GFP 13aa, pET28a bacterial expression vector, *KAN$^R$* | This study |
| pAC1481 | Bipartite T3 SV40-GFP, pET28a bacterial expression vector, *KAN$^R$* | [52] |
| pAC2406 | Bipartite T3 SV40-GFP 30aa, pET28a bacterial expression vector, *KAN$^R$* | This study |
| pAC2407 | Bipartite T3 SV40-GFP-GFP 8aa, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |
| pAC2408 | Bipartite T3 SV40-GFP-GFP 13aa, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |
| pAC2409 | Bipartite T3 SV40-GFP-GFP 15aa, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |
| pAC2410 | Bipartite T3 SV40-GFP-GFP 20aa, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |
| pAC2411 | Bipartite T3 SV40-GFP-GFP 30aa, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |

expressing genes under the control of the *MET25* promoter were grown overnight in selective media, pelleted, washed, resuspended in fresh media lacking methionine, and induced for five hours prior to localization studies.

*Protein Expression and Purification*- The bipartite T3 SV40-GFP-GFP variant proteins and $\Delta$IBB importin $\alpha$ were purified essentially as previously described [60]. Proteins were expressed at 30°C in the *E. coli* strain BL21 (DE3). Cells were grown in LB media containing kanamycin to an $OD_{600}$ of 0.6 and expression was induced by addition of 0.2 mM isopropyl-1-thio-beta-D-galactopyranoside (IPTG) for 5 hours. Cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM beta-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, and 0.1% igepal). Following cell lysis with a French pressure cell and separation of cell debris by high-speed centrifugation, the soluble supernatant was loaded onto a HiTrap nickel chelator column (Amersham Pharmacia Biotech), washed with buffer B (50 mM $Na_2HPO_4$, pH 7.4, 0.25 M NaCl), and eluted with a 0.5 M imidazole gradient. The protein was stored at -80°C in phosphate-buffered saline (PBS) containing 10% glycerol.

In Vitro *Binding Assay*- Solid-phase binding assays were performed essentially as previously described [109]. Microtiter Immulux HB plates (Dynex) were coated with 100 $\mu$l/well of 25 nM GFP-proteins (GFP alone, bipartite T3 SV40-GFP 13aa, and bipartite T3 SV40-GFP 30aa) in coating buffer (1xPBS supplemented with 2 mM DTT and 0.2 mM phenylmethylsulfonyl fluoride) overnight at 4ºC on a rocker. Plates were washed three times in PBS and incubated in 100 $\mu$l of binding buffer (coating buffer supplemented with 3% BSA and 0.1% Tween 20) overnight at 4ºC on a rocker. The

following morning, 100 μl of 0-110 nM S-tagged ΔIBB importin α in binding buffer was added to each well and incubated for 2 hours at 4ºC to allow binding. Plates were washed three times in binding buffer lacking BSA. Proteins were then crosslinked for 15 min at room temperature in 1 mg/ml 1-ethyl-3-(3-dimethylaminopropyl) carboiimide (Pierce) in binding buffer lacking BSA. The wells were then subjected to a series of washes: 20 min in PBS-T (PBS supplemented with 0.2% Tween 20), 10 min with PBS-T containing 100 mM ethanolamine, and 10 min with PBS-T containing 3% BSA. The bound ΔIBB importin α was then detected by incubation with S-protein-horseradish peroxidase conjugate (Novagen) in coating buffer containing 1% BSA and 0.1% Tween 20 for 1 hour at 4ºC with rocking. The plates were washed three times in PBS. Following the washes, 100 μl/well of horseradish peroxidase substrate (1-Step™ Turbo TMB (3,3'5,5'-tetramethylbenzidine)-ELISA (Pierce)) was incubated for 30 min at room temperature. The reaction was stopped by addition of 100 μl of 2 M $H_2SO_4$. The absorbance of the samples was measured at 450 nm with an Ultra Microplate Reader with KCjunior software (Bio-Tek Instruments, Inc.). Average absorbance values at $OD_{450}$ were determined for GFP alone and for the bipartite T3 SV40-GFP proteins at each ΔIBB importin α concentration. Background GFP absorbance values were subtracted from those of the bipartite T3 SV40-GFP proteins. The absorbance values for the bipartite T3 SV40-GFP proteins at each ΔIBB importin α concentration were used to generate binding curves by non-linear regression and Kd values using Prism4 software (GraphPad Software, Inc.).

**Results**

*The prevalence of classical nuclear import in* Saccharomyces cerevisiae- The

search algorithm from PSORT II [101] for monopartite cNLSs and bipartite cNLSs was

used to query three sets of data.  First, we examined the known yeast proteome

represented by the 5850 proteins in the *S. cerevisiae* GenBank [103] to identify the

fraction of yeast proteins predicted to contain a cNLS and, hence, which have the

potential to enter the nucleus via the classical import pathway.  Second, we analyzed the

1515 proteins localized to either the nucleus or the nucleolus in a comprehensive

subcellular localization study [104].  We reasoned that these proteins were targeted to the

nucleus by some nuclear import pathway and that a fraction probably used classical

nuclear import.  Third, we analyzed the 224 proteins that interact with importin $\alpha$

according to the BioGRID database [105], since proteins that contain functional cNLSs

should interact with importin $\alpha$.  Given that the bipartite consensus sequence consists of

the monopartite consensus sequence plus a linker and upstream basic residues, all

proteins that contain a putative bipartite sequence also, by definition, contain a

monopartite consensus sequence.  To eliminate overlap, proteins containing a bipartite

cNLS were removed from the monopartite tally.

The complete results of this analysis are available on the Corbett Lab website and

are summarized in Figure 2.1.  In the GenBank set of 5850 proteins, 968 (16.5%) proteins

contain a predicted bipartite cNLS and 1671 (28.6%) contain a predicted monopartite

cNLS.  This result suggests that classical nuclear import may indeed be as prevalent as

the name implies since about 45% of the proteins in the cell have the potential to enter
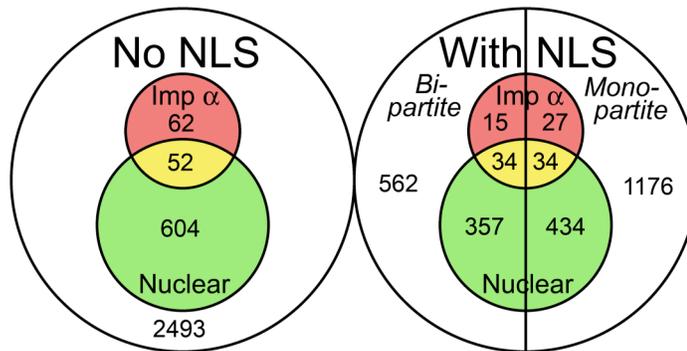
| Predicted NLS | Genome | | Nuclear | | Importin Alpha Interactors | |
|---|---|---|---|---|---|---|
| Bipartite | 968 | 16.5% | 391 | 25.8% | 49 | 21.9% |
| Monopartite | 1671 | 28.6% | 468 | 30.9% | 61 | 27.2% |
| None | 3211 | 54.9% | 656 | 43.3% | 114 | 50.9% |
| Total | 5850 | 100.0% | 1515 | 100.0% | 224 | 100.0% |

FIGURE 2.1 **The prevalence of classical nuclear import in *Saccharomyces cerevisiae.*** The PSORT II algorithm for monopartite and bipartite cNLSs was used to query three datasets: the 5850 proteins in the *S. cerevisiae* GenBank (*Genome*) [103]; the 1515 proteins localized to either the nucleus or the nucleolus in the global yeast GFP-fusion library (*Nuclear*, green) [104]; and the 224 proteins that interact with importin α according to the BioGRID database (*Importin α Interactors*, red) [105]. The Venn diagram indicates the number of proteins from each dataset with no predicted cNLS (left circle) or with a monopartite or bipartite cNLS (right circle). The corresponding percentages are listed in the table below.

the nucleus via the classical nuclear import pathway. Of the 1515 proteins that have been localized to the nucleus in the global GFP screen, 391 (25.8%) contain a putative bipartite cNLS and 468 (30.9%) contain a putative monopartite cNLS. Therefore, about 57% of steady-state nuclear proteins are predicted to use classical nuclear import, whereas about 43% may use other mechanisms to enter the nucleus. In the set of 224 proteins that interact with importin $\alpha$, 49 (21.9%) contain a predicted bipartite cNLS and 61 (27.2%) contain a predicted monopartite cNLS, meaning that about half do not contain predicted cNLSs.

*Expanding the classical bipartite linker*- Conventional bipartite cNLSs consist of two basic residues separated from a second stretch of basic residues by a linker of 10aa (reviewed in [53]). To test if the classical bipartite linker can vary from the established consensus sequence, we created variants of the bipartite T3 SV40-GFP-GFP reporter protein [52] with 8, 13, 15, 20, or 30 amino acids in the linker region. The sequences of those linkers are presented in Table 2.2 and are derived from the linker region of the artificial bipartite SV40 NLS, which is functional in both *in vivo* and *in vitro* studies [52, 55]. The T3 variant, which has a lysine to threonine substitution in the second basic residue of the C-terminal stretch of basic amino acids, was chosen because the monopartite T3 SV40 NLS binds to importin $\alpha$ too weakly to effect nuclear import *in vivo* [52]. Therefore, in order for the reporter proteins to be imported into the nucleus, the NLS must interact with importin $\alpha$ using both the upstream and downstream stretches of basic amino acids in the bipartite NLS, ensuring that the effect of the linker region is assessed in our analysis.

TABLE 2.2  **Linker sequences of bipartite T3 SV40 variants**

| Number of Residues in Linker | NLS Sequence |
| --- | --- |
| 10 | KR…**TADGSEFESP**……………………………..…..KTKKRKVE |
| 8 | KR…**TADEFESP**……………………….…………... KTKKRKVE |
| 13 | KR…**TADGSGSSEFESP**……………….…………... KTKKRKVE |
| 15 | KR…**TADGSTADGSEFESP**……………..…………… KTKKRKVE |
| 20 | KR…**TADGSEFESATADGSEFESP**………..………… KTKKRKVE |
| 30 | KR…**TADGSEFESATADGSEFESATADGSEFESP**...KTKKRKVE |

The localization of the bipartite T3 SV40-GFP-GFP reporter variants was assessed in wild-type cells (Figure 2.2).  As expected, bipartite T3 SV40-GFP-GFP with a 10aa linker was predominantly localized to the nucleus and GFP-GFP alone was predominantly localized to the cytoplasm.  The reporter proteins with linkers of 8aa and 13aa were primarily nuclear.  The reporter protein with a linker of 15aa showed slightly more cytoplasmic accumulation than the 13aa-linker reporter protein.  The reporter protein with a linker of 20aa had some nuclear accumulation with a strong cytoplasmic signal and the localization of the reporter protein with a 30aa linker resembled the localization of GFP-GFP alone.  Therefore, classical bipartite NLSs with linkers of between 8 and 20 amino acids can mediate nuclear import *in vivo*.

To verify that these bipartite T3 SV40-GFP-GFP reporter proteins are imported by the classical, importin α/importin β-mediated, system, the reporter protein variants were localized in *srp1-54* cells (Figure 2.3), which have a mutation in importin α just outside the minor NLS-binding pocket that affects how classical bipartite NLS-cargo proteins bind to the receptor [107].  All of the linker variants showed an increase in cytoplasmic signal in this mutant.  Since altering the ability of importin α to bind and import cargo altered the import of the linker variants, we conclude that the import of the bipartite T3 SV40-GFP-GFP reporter proteins is mediated by the classical import system.

Based on the *in vivo* localization data, we predict that a protein containing a bipartite NLS with a linker of 13aa should bind importin α and that a protein containing a bipartite NLS with a linker of 30aa should not.  To test this prediction, *in vitro* fusion proteins analogous to those created for the *in vivo* assays were created with linkers of 13aa or 30aa.  The ability of these bipartite T3 SV40-GFP proteins to bind ΔIBB importin

| No NLS | 10aa | 8aa | 13aa | 15aa | 20aa | 30aa |

GFP

DIC
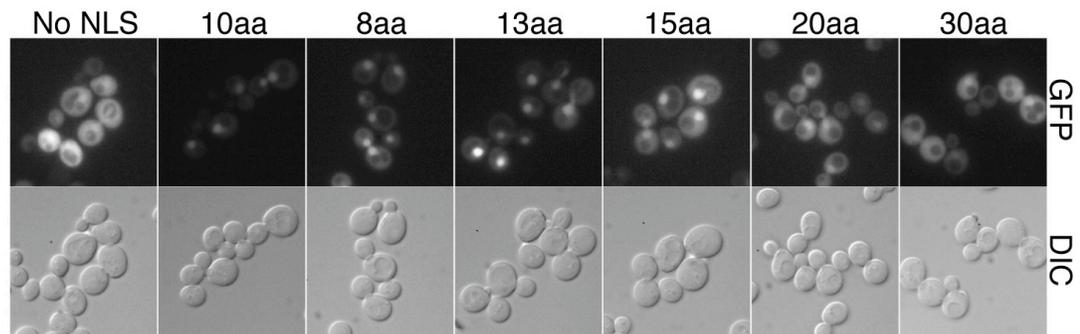
FIGURE 2.2  **Localization of bipartite T3 SV40-GFP-GFP with variable linker regions in wild-type cells.**  Wild-type cells (ACY192) expressing either GFP-GFP or bipartite T3 SV40-GFP-GFP reporter proteins with linkers of the indicated length were examined by direct fluorescence microscopy (GFP).  Corresponding differential interference contrast (DIC) images are shown.
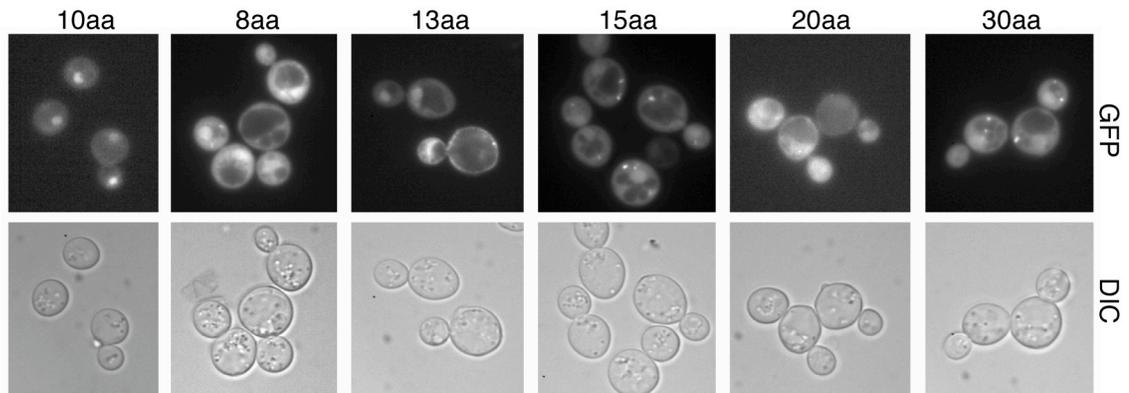
FIGURE 2.3  **Localization of bipartite T3 SV40-GFP-GFP with variable linker regions in *srp1-54* cells.**  *srp1-54* cells (ACY1563) expressing either GFP-GFP or bipartite T3 SV40-GFP-GFP reporter proteins with linkers of the indicated length were examined by direct fluorescence microscopy (GFP).  Corresponding differential interference contrast (DIC) images are shown.

α was assayed using an *in vitro* solid-phase binding assay (Figure 2.4). As predicted, the fusion protein with a linker of 13aa bound to importin α with a Kd of ~12 nM and the fusion protein with a linker of 30aa bound to importin α with a Kd of ~536 nM, a 45-fold change. For comparison, SV40-GFP with a linker of 10aa, which constitutes a functional NLS *in vivo*, binds to ΔIBB importin α with a kD of ~10 nM [52]. This binding data supports our *in vivo* findings; however, this result is preliminary and requires repetition.

*Revisiting the prevalence of putative bipartite cNLS-containing proteins in the yeast proteome-* The discovery that bipartite cNLSs with linkers of lengths other than 10aa bind importin α and function *in vivo* has the potential to greatly expand the pool of putative classical NLS-cargos. To update the prevalence of putative bipartite cargos in yeast, the PSORT II algorithm for bipartite cNLSs was modified to include linker lengths between 8 and 20 residues. The prevalence of new putative bipartite cargos is presented in Table 2.3, which shows that about 60% of the proteome may have the potential to be imported into the nucleus using the classical system. This number is obviously inflated since including a variable of up to 20aa casts a wide net. However, we can focus our attention by searching for putative bipartite cNLS-containing proteins that are known to interact with importin α and that are localized to the nucleus or the nucleolus. The results of this analysis are presented in Table 2.4.

Bipartite T3 SV40-GFP with 13aa Linker:  Kd = 12nM
Bipartite T3 SV40-GFP with 30aa Linker:  Kd = 536nM
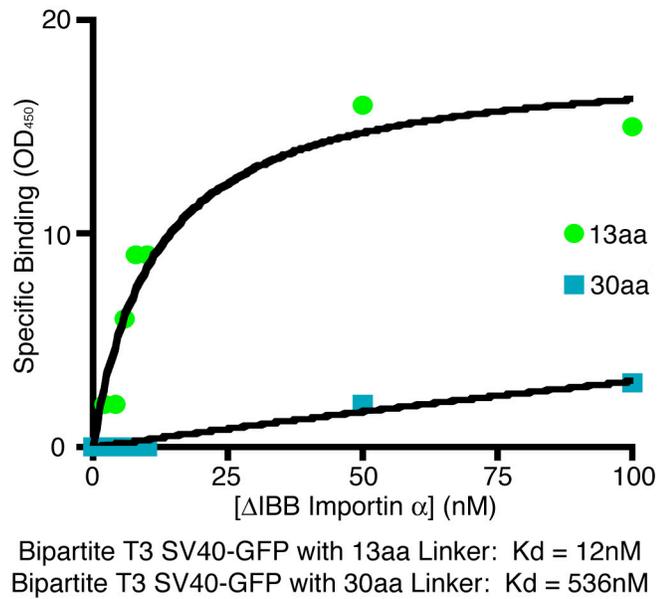
FIGURE 2.4 *In vitro* **binding of bipartite T3 SV40-GFP variants to ΔIBB importin α.** A solid-phase binding assay [109] was used to assess the binding of bipartite T3 SV40-GFP with a linker of 13aa or bipartite T3 SV40-GFP with a linker of 30aa to S-tagged ΔIBB importin α.  The NLS-GFP was conjugated to plates and incubated with ΔIBB importin α.  After washing, bound importin α was detected with an HRP-conjugated anti-S-protein antibody.

TABLE 2.3  **Predicted prevalence of proteins containing a classical bipartite NLS with a linker of the indicated length**

| Predicted Bipartite Linker Length | Genomic Prevalence | |
|---|---|---|
| 8→9 | 533 | 9.1% |
| 10 | 968 | 16.5% |
| 11→13 | 674 | 11.5% |
| 14→15 | 285 | 4.9% |
| 16→20 | 1018 | 17.4% |
| Total | 3478 | 59.4% |

TABLE 2.4 **Prevalence of proteins that have a predicted classical bipartite NLS with a linker of the indicated size that interact with importin α [105] and are nuclear or nucleolar at steady-state [104]**. Functions are adapted from the Saccharomyces Genome Database. Asterisked selections are unique to that category. As noted in Figure 2.1, there are 34 predicted proteins for the 10aa linker.

| 8aa→9aa linker, 6 predicted proteins | | |
| --- | --- | --- |
| YDR318W | MCM21 | Involved in minichromosome maintenance, component of the COMA complex (Ctf19, Okp1, Mcm21, Ame1) that bridges kinetochore subunits |
| *YGL238W | CSE1 | Mediates the nuclear export of importin alpha (Srp1p), required for accurate chromosome segregation |
| YGL175C | SAE2 | Protein with a role in accurate meiotic and mitotic double-strand break repair |
| YKL108W | SLD2 | Required for DNA replication, phosphorylation is essential for DNA replication |
| YLR298C | YHC1 | Component of the U1 snRNP complex required for pre-mRNA splicing |
| YMR001C | CDC5 | Polo-like kinase that has multiple functions in mitosis and cytokinesis through phosphorylation of substrates |
| **11aa→13aa, 15 predicted proteins** | | |
| YCL026C-A | FRM2 | Protein of unknown function, involved in the integration of lipid signaling pathways with cellular homeostasis |
| YDL005C | MED2 | Subunit of the RNA polymerase II mediator complex, essential for transcriptional regulation |
| YDR480W | DIG2 | Regulatory protein of unknown function, pheromone-inducible, involved in the regulation of mating-specific genes and the invasive growth pathway |
| *YDR510W | SMT3 | Ubiquitin-like protein of the SUMO family, conjugated to lysine residues of target proteins |
| YGL254W | FZF1 | Transcription factor involved in sulfite metabolism |
| *YGR218W | CRM1 | Major karyopherin, involved in export of proteins, RNAs, and ribosomal subunits from the nucleus |
| *YKL130C | SHE2 | RNA-binding protein that binds specific mRNAs, part of mRNA localization machinery that restricts accumulation of certain proteins to the bud |
| YKL108W | SLD2 | Required for DNA replication, phosphorylation is essential for DNA replication |
| YLR298C | YHC1 | Component of the U1 snRNP complex required for pre-mRNA splicing |
| YMR001C | CDC5 | Polo-like kinase that has multiple functions in mitosis and cytokinesis through phosphorylation of substrates |
| YNL232W | CSL4 | Subunit of the exosome, which is an essential complex present in both nucleus and cytoplasm that mediates RNA processing and degradation |
| YNL102W | POL1 | Catalytic subunit of the DNA polymerase alpha-primase complex |
| YOR001W | RRP6 | Exonuclease component of the nuclear exosome; contributes to the quality-control system that retains and degrades aberrant mRNAs in the nucleus |
| YPL153C | RAD53 | Protein kinase, required for cell-cycle arrest in response to DNA damage, plays a role in initiation of DNA replication |
| *YPL111W | CAR1 | Arginase, responsible for arginine degradation |
| **14aa→15aa linker, 4 predicted proteins** | | |
| YGL175C | SAE2 | Protein with a role in accurate meiotic and mitotic double-strand break repair |
| YJR056C | | Uncharacterized ORF |
| YMR047C | NUP116 | Subunit of the nuclear pore complex (NPC) that is localized to both sides of the pore, interacts with mRNA export factor Mex67p and with karyopherin Kap95p |
| YMR125W | STO1 | Large subunit of the nuclear mRNA cap-binding protein complex, interacts with Npl3p to carry nuclear poly(A)+ mRNA to cytoplasm |

| 16aa→20aa linker, 21 predicted proteins | | |
|---|---|---|
| YAL054C | ACS1 | Acetyl-coA synthetase isoform which is the nuclear source of acetyl-coA for histone acetylation |
| YCL011C | GBP2 | Poly(A+) RNA-binding protein, involved in the export of mRNAs from the nucleus to the cytoplasm |
| YDL005C | MED2 | Subunit of the RNA polymerase II mediator complex, essential for transcriptional regulation |
| YDR240C | SNU56 | Component of U1 snRNP required for mRNA splicing via spliceosome |
| *YER095W | RAD51 | Strand exchange protein, forms a helical filament with DNA that searches for homology, involved in the recombinational repair of double-strand breaks |
| YGL175C | SAE2 | Protein with a role in accurate meiotic and mitotic double-strand break repair |
| YJL157C | FAR1 | Cyclin-dependent kinase inhibitor that mediates cell cycle arrest in response to pheromone |
| YJR056C | | Uncharacterized ORF |
| YJR068W | RFC2 | Subunit of heteropentameric Replication factor C (RF-C), which is a DNA binding protein and ATPase that acts as a clamp loader of the proliferating cell nuclear antigen (PCNA) processivity factor for DNA polymerases δ and ε |
| YKL108W | SLD2 | Protein required for DNA replication, phosphorylation is essential for DNA replication and for complex formation with Dpb11p |
| YLR082C | SRL2 | Protein of unknown function; overexpression suppresses the lethality caused by a rad53 null mutation |
| YLR298C | YHC1 | Component of the U1 snRNP complex required for pre-mRNA splicing |
| YLR383W | SMC6 | Protein involved in structural maintenance of chromosomes, essential subunit of Mms21-Smc5-Smc6 complex |
| YMR001C | CDC5 | Polo-like kinase that has multiple functions in mitosis and cytokinesis through phosphorylation of substrates |
| YMR047C | NUP116 | Subunit of the nuclear pore complex (NPC) that is localized to both sides of the pore, interacts with mRNA export factor Mex67p and with karyopherin Kap95p |
| YMR125W | STO1 | Large subunit of the nuclear mRNA cap-binding protein complex, interacts with Npl3p to carry nuclear poly(A)+ mRNA to cytoplasm |
| YNL232W | CSL4 | Subunit of the exosome, which is an essential complex present in both nucleus and cytoplasm that mediates RNA processing and degradation |
| YOL135C | MED7 | Subunit of the RNA polymerase II mediator complex, associates with core polymerase subunits to form the RNA polymerase II holoenzyme |
| YOL051W | GAL11 | Component of the Mediator complex; interacts with RNA polymerase II and the general transcription factors to form the RNA polymerase II holoenzyme |
| YOR001W | RRP6 | Exonuclease component of the nuclear exosome; contributes to the quality-control system that retains and degrades aberrant mRNAs in the nucleus |
| YPL153C | RAD53 | Protein kinase, required for cell-cycle arrest in response to DNA damage, plays a role in initiation of DNA replication |

**Discussion**

In this study, we began with an analysis of the prevalence of classical nuclear import in *S. cerevisiae* using the established parameters for what constitutes a classical NLS. We found that almost half of the known proteins in yeast contain a putative cNLS, hinting at the potential importance of the classical system of nuclear import to establishing the nuclear proteome. Given that only 25.8% of proteins localize to the nucleus at steady-state when chromosomally tagged with GFP [104], this fraction is revealed as an especially sizeable proportion. Of the proteins known to interact with importin α, about 22% were found to contain putative monopartite cNLSs and about 27% were found to contain putative bipartite cNLSs. At first, these numbers seem surprisingly low, since cNLSs are conventionally thought to mediate interactions with importin α; however, many of the interactions in the BioGRID database are genetic in nature and do not require or even necessarily imply a direct physical interaction. In addition, many proteins that interact with importin α in non-cargo roles are included, such as importin β (the import partner for importin α) and Cse1 (the export factor for importin α), which interact with importin α through non-NLS-mediated interfaces [44, 79]. The current definition of a cNLS also may need refining, since some of these proteins may contain cNLSs and utilize traditional importin α-mediated import, but escape recognition by the standard algorithm. For example, the primary sequence of the STAT1 (signal transducers and activators of transcription 1) protein does not contain a functional classical targeting sequence; however, upon dimerization, each subunit contributes basic residues that form a cNLS recognized by importin α [110].

Bipartite cNLSs have traditionally been defined as having a linker length of 10 residues. However, our data, which shows that protein cargo bearing bipartite cNLSs with linkers ranging between eight and twenty amino acids can be imported by the classical transport system and can bind importin α, suggests that this established definition may be artificially limiting. Mining the yeast proteome for putative classical cargos using an updated definition of the bipartite cNLS revealed that about 60% of the known proteins contain putative bipartite cNLS motifs. Presumably, this number is unreasonably high and further studies are necessary to further refine the bipartite consensus. Current studies are aimed at identifying real-world examples of classical cargos that contain cNLS motifs with expanded or contracted linker regions by localizing a selection of the predicted atypical bipartite cNLS-containing cargos mentioned in Table 2.4. Additionally, we aim to ascertain the effect of the sequence of the linker on the efficacy of the cNLS. In our studies, the linkers tested were all derived from a single linker sequence. Though this sequence has previously been shown to be functional at 10aa [52, 55], it is possible that the particular sequence chosen is simply incompatible with assuming the conformation necessary for a longer bipartite cNLS to bind importin α and that other linker sequences may bestow functionality upon expanded bipartite cNLSs. Therefore, though the model bipartite cNLS with a 30aa linker proved incapable of effecting import, one can imagine how a different linker sequence might have the ability to assume a position compatible with importin α-binding and might contribute to a functional cNLS *in vivo*.

To conclude, in this study, we have reconsidered the bipartite cNLS archetype and found that the traditional consensus is unnecessarily restrictive. Expanding the classical

NLS definition allowed us to reprobe the yeast proteome and discover that the prevalence of classical nuclear transport may be much greater than previously imagined.  Indeed, a large fraction of the proteins examined in these three databases show evidence for import by the classical system; however, our analyses also collectively indicate that alternative nuclear localization mechanisms also are likely to be prevalent.  Thus, further studies and novel experimental approaches will be necessary to fully explore the contributions of these additional systems.

**CHAPTER 3**

**A PY-NLS Nuclear Targeting Signal is Required for Nuclear Localization and Function of the *Saccharomyces cerevisiae* mRNA-Binding Protein, Hrp1**

Allison Lange, Ryan E. Mills, Scott E. Devine, and Anita H. Corbett

Department of Biochemistry, Emory University, School of Medicine, Atlanta, GA 30322

**Introduction**

A main focus of inquiry in the nuclear transport field has been understanding the mechanism by which β-karyopherin receptors recognize their particular host of cargo proteins.  In the most well-characterized system of nuclear transport, the classical nuclear import pathway, a specific β-karyopherin receptor termed importin β binds to an adaptor protein, importin α, which interacts with cargos carrying classical NLSs (cNLSs).  These cNLSs consist of either one or two clusters of basic amino acids and are epitomized by the SV40 large T antigen NLS ([126]PKKKRRV[132]) and the nucleoplasmin NLS ([155]KRPAATKKAGQAKKKK[170]) [53].  A large number of classical substrates have been experimentally verified and examination of their nuclear targeting signals has led to the development of a useful consensus sequence [53].

In contrast to the classical nuclear import pathway, for most β-karyopherin receptors, only a handful of substrates have been identified making generalization of binding schemes difficult.  Compounding the problem, these cargos often lack sequence or structural homology and, in some cases, even interact with a particular β-karyopherin receptor using distinct binding sites and diverse interaction modes (for example [79, 82, 83]).  However, a mechanism underlying recognition of a specific class of cargos by a transport receptor that does not mediate cNLS-protein import was recently determined.  Chook and colleagues reported the structure of a human β-karyopherin called karyopherin β2 or transportin (Kapβ2) bound to the NLS of its best-characterized cargo, the mRNA-binding protein hnRNP A1 [92].  Combining the structural information with biochemical studies involving other known Kapβ2 cargos, they then proposed a set of

predictive rules, which outline the requirements for a putative Kapβ2 NLS: 1) the NLS is structurally disordered, 2) the NLS has a net basic charge, and 3) the NLS has a hydrophobic or basic region upstream of a C-terminal $R/H/KX_{2-5}PY$ motif. The invariant proline and tyrosine residues contained at the end of the consensus led them to term these Kapβ2-recognition sequences "PY-NLSs". Using these rules, they then searched the human proteome, identified 81 putative Kapβ2 cargos, and showed that a selection of these PY-NLS proteins bound to Kapβ2 in a RanGTP-dependent manner *in vitro*. These binding experiments suggest that the proteins are substrates of Kapβ2; however, *in vivo* functional and localization experiments are still necessary to demonstrate that these PY-NLS-containing proteins are imported by Kapβ2 in a PY-NLS-dependent manner.

The apparent ortholog of Kapβ2 in *S. cerevisiae* is Kap104 (karyopherin 104) [93], which has two known cargos, the mRNA-binding proteins Nab2 (nuclear poly-adenylated RNA-binding protein 2) and Hrp1 (heterogeneous nuclear ribonucleoprotein 1) [111, 112]. Previous work has defined the general regions of these cargo proteins that are required for Kap104-mediated nuclear import [112, 113]. In Nab2, a region containing an RNA-binding motif consisting of a series of arginine-glycine-glycine (RGG) repeats is necessary and sufficient for binding to Kap104 and for nuclear import *in vivo* [112, 113]. Deletion of the CCCH zinc finger region of Nab2 also reduces binding to Kap104 *in vitro* [112]. In Hrp1, the arginine/glycine-rich carboxyl terminus of the protein is required for binding to Kap104 and is sufficient to mediate import *in vivo* [112]. Thus, although the protein domains in Hrp1 and Nab2 that mediate Kap104 binding have been determined, no specific nuclear import sequences have been identified. Much like many Kapβ2 cargos [92], Hrp1 and Nab2 are involved in mRNA biogenesis,

with Hrp1 contributing to cleavage and polyadenylation of pre-mRNA 3' ends [114] and Nab2 aiding in poly(A) tail length regulation and mRNA export [115, 116]. Interestingly, the human protein most closely related to Hrp1 is hnRNP A1 (PBI-BLAST).

The goal of this study was to examine the function of the novel PY-NLS sequence *in vivo*. Examination of the primary sequences of Hrp1 and Nab2 uncovered putative PY-NLS-like sequences in both proteins. Using a combination of localization, function, and binding studies we found that the PY-NLS-like sequence within Hrp1 is necessary and sufficient for nuclear import of the protein and is required for protein function. However, the less conserved PY-NLS-like sequences within Nab2 are not required for nuclear accumulation of Nab2. These results demonstrate that the PY-NLS consensus sequence is conserved in *S. cerevisiae*, but also reveal that Kap104 likely utilizes additional mechanisms of cargo recognition. Finally, identification of the functional PY-NLS in yeast allowed us to interrogate the yeast proteome to identify new putative cargos of Kap104. Therefore, we conclude with a preliminary prediction of the likely prevalence of the PY-NLS in *S. cerevisiae*.

**Experimental Procedures**

*Strains, Plasmids, and Chemicals*- All chemicals were obtained from US Biological or Sigma unless otherwise noted. All media was prepared and all DNA manipulations were performed according to standard procedures [99, 100]. All yeast strains and plasmids used in this study are described in Table 3.1.

TABLE 3.1 **Strains and plasmids used in Chapter 3**

| Strain/Plasmid | Description | Origin |
|---|---|---|
| FY23 (ACY192) | Wild type, MATa *ura3-52 leu2Δ1 trp1* | [106] |
| ESY41-4C (ACY251) | ΔKAP104::HIS3 [KAP104 *CEN URA3 AMP$^R$*], MATα *ade trp leu his* | P.A. Silver |
| ACY427 | ΔNAB2::HIS3 [NAB2 *CEN URA3 AMP$^R$*], MATa *leu lys ade his* | This study |
| SVL182/PSY1224 (ACY1571) | ΔHRP1::HIS3 [HRP1 *CEN URA3 AMP$^R$*] | S.R. Valentini |
| pRS315 (pAC3) | *CEN LEU2 AMP$^R$* | [117] |
| pPS808 (pAC23) | GFP, *GAL1-10* promoter, *2μ URA3 AMP$^R$* | P.A. Silver |
| pAC717 | NAB2, *CEN LEU2 AMP$^R$* | [116] |
| pAC753 | NAB2-GFP, *CEN LEU2 AMP$^R$* | [116] |
| pAC1069 | GFP-GFP, *MET25* promoter, *CEN URA3 AMP$^R$* | [55] |
| pPS1358 (pAC1725) | GFP-HRP1, *GAL1-10* promoter, *2μ URA3 AMP$^R$* | [114] |
| pAC2023 | NOP1-GFP, *CEN URA3 AMP$^R$* | This study |
| pAC2325 | HRP1, *CEN LEU2 AMP$^R$* | This study |
| pAC2329 | GFP-HRP1 R525A/P531A/Y532A, *GAL1-10* promoter, *2μ URA3 AMP$^R$* | This study |
| pAC2330 | GFP-HRP1 P531A/Y532A, *GAL1-10* promoter, *2μ URA3 AMP$^R$* | This study |
| pAC2344 | HRP1 R525A/P531A/Y532A, *CEN LEU2 AMP$^R$* | This study |
| pAC2345 | HRP1 P531A/Y532A, *CEN LEU2 AMP$^R$* | This study |
| pAC2440 | HRP1 Y532V, *CEN LEU2 AMP$^R$* | This study |
| pAC2441 | GFP-HRP1 Y532V, *GAL1-10* promoter, *2μ URA3 AMP$^R$* | This study |
| pAC2442 | NAB2 P332A, *CEN LEU2 AMP$^R$* | This study |
| pAC2443 | NAB2-GFP PV332A, *CEN LEU2 AMP$^R$* | This study |
| pAC2444 | NAB2 P407A, *CEN LEU2 AMP$^R$* | This study |
| pAC2445 | NAB2-GFP P407A, *CEN LEU2 AMP$^R$* | This study |
| pAC2451 | GFP-GFP-HRP1 aa503-534, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |
| pAC2452 | GFP-GFP-HRP1 aa522-534, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |
| pAC2472 | GFP-GFP-NAB2 aa320-333, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |
| pAC2473 | GFP-GFP-NAB2 aa389-408, *MET25* promoter, *CEN URA3 AMP$^R$* | This study |
| pAC2527 | NAB2 P332A/P407A, *CEN LEU2 AMP$^R$* | This study |
| pAC2528 | GFP-HRP1 P531A, *GAL1-10* promoter, *2μ URA3 AMP$^R$* | This study |
| pAC2538 | NAB2-GFP P332A/P407A, *CEN LEU2 AMP$^R$* | This study |
| pAC2539 | HRP1 P531A, *CEN LEU2 AMP$^R$* | This study |
| pAC2551 | GST-KAP104, pGEX-6P3 *AMP$^R$* | M.P. Rout |

*Microscopy*- GFP-fusion proteins were localized in live cells using direct fluorescence microscopy. The GFP signal was visualized using a GFP-optimized filter (Chroma Technology) on an Olympus BX60 epifluorescence microscope equipped with a Photometrics Quantix digital camera. For most experiments, cells expressing genes expressed from their own promoters were grown overnight in selective media, diluted in fresh media, and grown for three hours prior to localization studies. Cells expressing genes under the control of the *GAL1-10* promoter were grown overnight in selective media with glucose as a sugar source, pelleted, washed, resuspended in fresh media with galactose as a sugar source, and induced for six hours prior to localization studies. Cells expressing genes under the control of the *MET25* promoter were grown overnight in selective media, pelleted, washed, resuspended in fresh media lacking methionine, and induced for five hours prior to localization studies. Experiments involving the *ΔKAP104* deletion strain (ACY251) required an initial growth phase of five nights to reach the appropriate cell density.

In Vitro *GST-binding Assay*- For binding studies, Kap104 was expressed in *E. coli* as a GST fusion protein [118], then GST-Kap104 beads were incubated with yeast lysate from cells expressing wild-type GFP-Hrp1 or R525A/P531A/Y532A GFP-Hrp1. To prepare the GST-Kap104 beads, *E. coli* cells expressing GST-Kap104 (pAC2551) were resuspended in PBS supplemented with 0.5 mM phenylmethylsulfonyl fluoride (PMSF), PLAC (3 mg/ml each of Pepstatin A, Leupeptin, Aprotinin, Chymostatin), and 2 mM β-mercaptoethanol and lysed by sonication. Lysates were cleared by centrifugation and incubated with 200 µl of prepared glutathione Sepharose beads (GE Healthcare). Beads were collected and washed three times with PBS. To prepare yeast lysates, wild-

type cells (ACY192) containing plasmids encoding wild-type GFP-Hrp1 (pAC1725) or

R525A/P531A/Y532A GFP-Hrp1 (pAC2329) were grown overnight in selective media

with glucose as a sugar source, pelleted, washed, and grown overnight in selective media

with galactose as a sugar source to induce expression of the GFP-Hrp1 fusion proteins.

Cells were collected and washed twice in $dH_2O$ and once in PBSMT (PBS, 5 mM $MgCl_2$,

0.5% Triton X-100).  Glass bead lysis was conducted in PBSMT supplemented with

protease inhibitors (PMSF and PLAC).  Lysates were cleared by centrifugation and total

protein concentration was assessed by Bradford assay.  One milligram of total yeast

lysate and 50 µl of GST-Kap104 bead slurry were incubated overnight with agitation.

The unbound fraction was removed, then the bound fraction was washed three times with

PBSMT + 500 mM NaCl and once with PBS.  Bound and unbound samples were loaded

on a 10% denaturing SDS-PAGE gel.  The proteins were transferred to a nitrocellulose

membrane and probed with either primary α-GFP (rabbit, 1:3,000 dilution) and

secondary α-rabbit-HRP conjugated (1:5,000) antibodies to detect the GFP-Hrp1 fusion

proteins or primary α-GST (mouse, 1:5,000) and secondary α-mouse-HRP conjugated

(1:5,000) antibodies to detect the GST-Kap104 protein.

In Vivo *Functional Analyses*- The *in vivo* function of each of the Hrp1 and Nab2

variants was tested using a plasmid shuffle technique [119].  Plasmids encoding wild-

type or variant Hrp1 or Nab2 proteins were transformed into either *ΔHRP1* (ACY1571)

or *ΔNAB2* (ACY427) cells containing a wild-type *HRP1* or *NAB2 URA3* maintenance

plasmid.  Single transformants were grown to saturation in liquid culture, serially diluted

(1:10) in $dH_2O$, and spotted on control ura⁻ leu⁻ glu plates or on selective leu⁻ glu plates

containing 5-fluoroorotic acid (5-FOA).  Plates were incubated at 18°C, 25°C, 30°C, or

37°C for 3-5 days. For growth curve analysis, cells picked from the 25°C 5-FOA Hrp1 plasmid shuffle plate, which express either wild-type or P531A Hrp1 as the only copy of Hrp1, were grown overnight at 25°C, normalized to equal starting concentrations, diluted 1:10 in a 96-well plate, and monitored for growth over time using an ultra microplate reader (Bio-Tek Instruments, Inc.). Cells were incubated at 37°C with shaking and the optical density was measured at 600nm every 30 minutes.

*Bioinformatics*- A custom library was constructed with different PY-NLS profiles: the consensus sequence for a hydrophobic PY-NLS [92], [LIMHFYVPQ]-[GAS]-[LIMHFYVPRQK]-[LIMHFYVPRQK]-X(7,12)-[RKH]-X(2,5)-PY; the sequence for a basic PY-NLS [92], [KR]-X(0,2)-[KR]-[KR]-X(3,10)-[RKH]-X(1,5)-PY; and the sequence for the C-terminal PY-core of the PY-NLS, [RKH]-X(2,5)-PY. This library was then used with ScanProsite [102] to query two data sets: the yeast proteome represented by the 5,850 proteins in the *S. cerevisiae* GenBank database [120] and the 1,515 proteins localized to either the nucleus or the nucleolus according to a comprehensive subcellular localization study using the global yeast GFP-fusion library [104]. The results are summarized in a Venn diagram with corresponding percentages tabulated in a chart. A detailed table of results is presented on the Corbett Lab website.

**Results**

*Hrp1 and Nab2 contain PY-NLS-like sequences*- Scanning the protein sequences of the known Kap104 cargos, Hrp1 and Nab2 [111, 112], revealed that Hrp1 contains one sequence and Nab2 contains two sequences that are similar to the established Kapβ2-

binding sequence termed the PY-NLS [92] (Figure 3.1A). The PY-NLS-like sequence of

Hrp1 (residues 506-532) is located at the extreme C-terminus of the protein, within the

region previously shown to be required for import [112]. The PY-NLS-like sequences of

Nab2 are both contained in the C-terminal CCCH zinc finger domain. PY-NLS-like

sequence 1 (residues 320-333) lies between the second and third zinc fingers and

sequence 2 (residues 389-408) is located between the fourth and fifth zinc finger. The

sequence in Hrp1 very closely resembles the mammalian PY-NLS, containing the core C-

terminal portion of the consensus sequence preceded by a basic stretch of amino acids. In

contrast, the putative sequences in Nab2 both contain upstream basic or hydrophobic

stretches, but they vary from the PY-NLS consensus in the C-terminal PY-core as both

sequences contain a lysine residue followed within two to five residues by a proline and a

valine, yielding PV rather than the eponymous PY sequence. This change in the final

residue raises the possibility that, although Hrp1 and Nab2 are both established cargos of

Kap104 [111, 112], they may interact with the receptor using different sets of interactions

or perhaps even in fundamentally different ways.

*Hrp1 and Nab2 are imported by Kap104*- To verify that nuclear localization of

Hrp1 and Nab2 depends on Kap104 [111], the localization of GFP-tagged Hrp1 or Nab2

was assessed in wild-type or *ΔKAP104* cells (Figure 3.1B). Hrp1 was tagged with GFP

at the N-terminus of the protein to preserve the C-terminal location of the PY-NLS-like

sequence. Nab2 was tagged at the C-terminus with GFP since this fusion protein has

previously been shown to function *in vivo* and to localize to the nucleus [116]. In wild-

type cells, both GFP-Hrp1 and Nab2-GFP are localized to the nucleus; however, in

*ΔKAP104* cells, both GFP-Hrp1 and Nab2-GFP are localized throughout the cell,
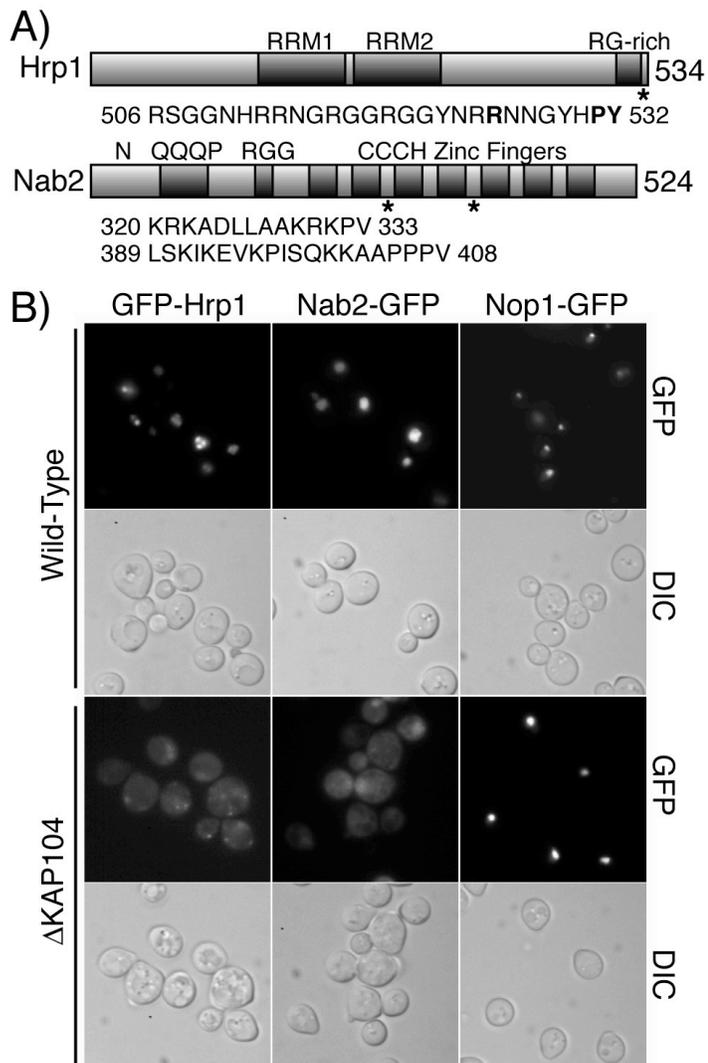
FIGURE 3.1 **Hrp1 and Nab2 contain putative PY-NLS-like sequences.** *A*, domain structures of Hrp1 and Nab2. The positions of the PY-NLS-like sequence(s) are indicated by the asterisks and sequences are listed below each protein. *B*, nuclear localization of Hrp1 and Nab2 is dependent on Kap104. Wild-type and *ΔKAP104* cells expressing GFP-Hrp1, Nab2-GFP, or the control protein Nop1-GFP were examined by direct fluorescence microscopy (GFP). Corresponding differential interference contrast (DIC) images are shown.

indicating that proper nuclear localization of Hrp1 and Nab2 is dependent on Kap104. This experiment also demonstrates that GFP-tagged versions of the proteins are valid reporters. A non-Kap104-dependent cargo, Nop1-GFP [111, 118], is properly localized within the nucleus in both wild-type and *ΔKAP104* cells, showing that all nuclear import is not impaired in *ΔKAP104* cells.

*The PY-NLS-like sequence within Hrp1 is necessary and sufficient for nuclear import and is necessary for Kap104 binding-* To test whether the putative PY-NLS-like sequence within Hrp1 is necessary for the import of Hrp1 into the nucleus, specific amino acid changes were created in the C-terminal PY-core of the PY-NLS-like sequence of GFP-Hrp1 and the resulting GFP-Hrp1 variants were localized in wild-type cells (Figure 3.2A). Wild-type GFP-Hrp1 was localized to the nucleus; however, when either the PY (P531A/Y532A) or both the upstream arginine and the PY (R525A/P531A/Y532A) residues were changed to alanine, GFP-Hrp1 was mislocalized to the cytoplasm, indicating that the PY-NLS-like sequence of Hrp1 is necessary for nuclear accumulation of Hrp1. Both the wild-type and the mutant GFP-Hrp1 partially localized in discrete puncta. These accumulations do not seem to correlate with particular nuclear or cytoplasmic bodies and their cause is unknown. Immunoblotting verified that all of the GFP-Hrp1 proteins were expressed at approximately equal levels (data not shown).

To determine whether the PY-NLS-like sequence within Hrp1 is required for the interaction between Hrp1 and Kap104, we tested whether recombinant GST-Kap104 could interact with either wild-type or mutant GFP-Hrp1 in yeast lysate. For this experiment, GST-Kap104 beads were incubated with yeast lysate from cells expressing either wild-type GFP-Hrp1 or R525A/P531A/Y532A GFP-Hrp1 (Figure 3.2B). Wild-
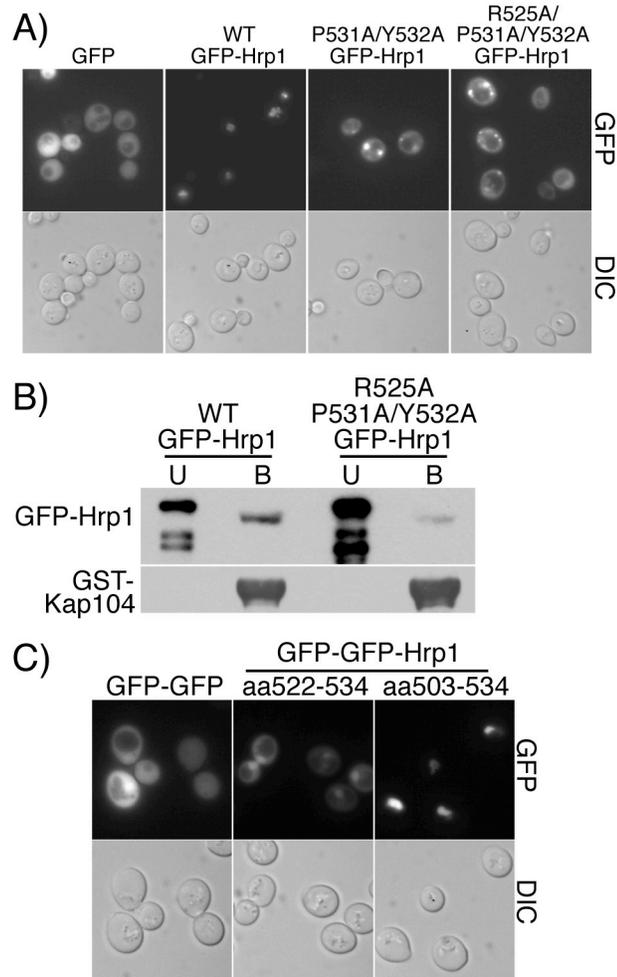
FIGURE 3.2 **The PY-NLS-like sequence within Hrp1 is necessary and sufficient for Hrp1 import and is necessary for Kap104 binding.** *A*, wild-type cells expressing GFP alone, wild-type GFP-Hrp1, P531A/Y532A GFP-Hrp1, or R525A/P531A/Y532A GFP-Hrp1 were examined by direct fluorescence microscopy (GFP). Corresponding differential interference contrast (DIC) images are shown. *B, in vitro* binding between GST-Kap104 and either wild-type GFP-Hrp1 or R525A/P531A/Y532A GFP-Hrp1 was examined using glutathione beads as described in Experimental Procedures. The unbound (U) and bound (B) fractions were probed with an anti-GFP antibody to detect GFP-Hrp1 fusion proteins or with an anti-GST antibody to detect GST-Kap104. *C*, wild-type cells expressing a GFP-GFP control, GFP-GFP-Hrp1 aa522-534 (containing the Hrp1 PY-core), or GFP-GFP-Hrp1 aa503-534 (containing the entire Hrp1 PY-NLS) were examined by direct fluorescence microscopy (GFP). Corresponding differential interference contrast (DIC) images are shown.

type GFP-Hrp1 robustly bound to GST-Kap104 while R525A/P531A/Y532A GFP-Hrp1 showed greatly reduced binding, demonstrating that the PY-NLS-like sequence within Hrp1 is important for receptor binding.

To test if the PY-NLS-like sequence within Hrp1 is sufficient to mediate import of a non-nuclear protein, either the C-terminal PY-core (residues 522-534) or the entire PY-NLS-like sequence of Hrp1 (residues 503-534) was fused to GFP-GFP (Figure 3.2C). Two GFP molecules were used to create a protein that was too large to efficiently diffuse through the nuclear pore into the nucleus. Accordingly, in order to accumulate in the nucleus, these fusion proteins must utilize an active system of transport. As expected, GFP-GFP alone was mainly localized to the cytoplasm. The C-terminal PY-core fused to GFP-GFP showed significant nuclear accumulation with some remaining cytoplasmic signal. The complete PY-NLS-like sequence fused to GFP-GFP mediated very efficient nuclear targeting with the reporter protein localized exclusively to the nucleus. These results indicate that the PY-NLS-like sequence of Hrp1 is sufficient to mediate import into the nucleus and that even the small C-terminal PY-core lacking the upstream basic or hydrophobic residues can mediate nuclear targeting to some degree. Both Hrp1 GFP-GFP-PY-NLS reporters were mislocalized to the cytoplasm in *kap104-16* mutant cells [111] (data not shown), showing that import of these reporter proteins is dependent on Kap104.

*The PY-NLS-like sequences within Nab2 are neither necessary nor sufficient for Nab2 import into the nucleus-* An approach similar to that described for Hrp1 was taken to determine if the PY-NLS-like sequences within Nab2, which actually contain terminal PV residues rather than PY residues, are required for Nab2 nuclear localization. To test

if either or both of these PY-NLS-like sequences within Nab2 are required for Nab2 nuclear import, wild-type Nab2-GFP, P332A Nab2-GFP, P407A Nab2-GFP, and P332A/P407A Nab2-GFP were localized in wild-type cells.  As shown in Figure 3.3A, wild-type Nab2-GFP is localized exclusively to the nucleus.  Nab2-GFP with amino acid changes in either or both of the two putative PY-NLS-like sequences is also localized exclusively to the nucleus, indicating that neither PY-NLS-like sequence is required for Nab2 import.

To test if either of the putative PY-NLS-like sequences within Nab2 is sufficient to mediate nuclear import, either Nab2 PY-NLS-like sequence 1 (residues 320-333) or Nab2 PY-NLS-like sequence 2 (residues 389-408) was fused to GFP-GFP and these reporter proteins were localized in wild-type cells (Figure 3.3B).  Both Nab2 PY-NLS-GFP-GFP reporters localized throughout the cell, demonstrating that neither Nab2 PY-NLS-like sequence alone is sufficient to mediate import into the nucleus.

*The PY-NLS-like sequence is required for Hrp1 function but not for Nab2 function*- Since the PY-NLS-like region of Hrp1 is both necessary and sufficient to mediate Hrp1 nuclear import, this sequence should be required for the function of the essential nuclear Hrp1 protein [114] *in vivo*.  In contrast, since neither of the PY-NLS-like sequences within Nab2 is required for import, we predict that neither of the sequences should be required for Nab2 function.  Accordingly, to test whether the PY-NLS-like sequences within Hrp1 or Nab2 are essential for protein function *in vivo*, plasmid shuffle assays were performed.  *ΔHRP1* or *ΔNAB2* cells containing a wild-type *URA3* maintenance plasmid were transformed with plasmids encoding wild-type Hrp1 or Nab2, mutant PY-NLS Hrp1 or Nab2, or vector alone and were plated on control plates

FIGURE 3.3 **The PY-NLS-like sequences within Nab2 are neither necessary nor sufficient for Nab2 nuclear localization.** *A*, wild-type cells expressing GFP alone, wild-type Nab2-GFP, P332A Nab2-GFP, P407A Nab2-GFP, or P332A/P407A Nab2-GFP were examined by direct fluorescence microscopy (GFP). Corresponding differential interference contrast (DIC) images are shown. *B*, wild-type cells expressing a GFP-GFP control, GFP-GFP-Nab2 aa320-333, or GFP-GFP-Nab2 aa389-408 were examined by direct fluorescence microscopy (GFP). Corresponding differential interference contrast (DIC) images are shown.

59

or plates containing 5-fluoroorotic acid (5-FOA). 5-FOA is a toxic analog of uracil, which selects against *URA*-containing maintenance plasmids [119]. *HRP1* and *NAB2* are essential genes; therefore, only cells containing a functional copy of *HRP1* or *NAB2* are able to grow on plates containing 5-FOA. As seen in Figure 3.4, cells expressing P531A/Y532A Hrp1 or R525A/P531A/Y532A Hrp1 are not viable, indicating that the PY-NLS-like sequence within Hrp1 is essential for protein function, presumably because import of Hrp1 is required for proper mRNA processing [114]. Cells expressing P332A Nab2, P407A Nab2, or P332A/P407A Nab2 as the only copy of Nab2 grow as well as cells expressing wild-type Nab2, indicating that neither PY-NLS-like region within Nab2 is required for cell viability.

*Functional dissection of the PY-NLS motif*- To assess the contribution of the proline and the tyrosine residues to the function of the PY-NLS motif, we created several variants within the PY-NLS of Hrp1. First, we changed the tyrosine to valine (Y532V Hrp1). In the structure of Kapβ2 bound to hnRNP A1, the tyrosine residue of the PY-NLS makes extensive hydrophobic and polar contacts with the receptor [92]. This change to a nonpolar residue should disrupt these interactions. This amino acid change also mimics the PY-NLS-like sequences of Nab2, allowing a further test of whether valine can substitute for tyrosine in *S. cerevisiae* PY-NLS sequences. Second, we changed the proline to alanine (P531A Hrp1). The proline of the hnRNP A1 PY-NLS makes hydrophobic contacts with Kapβ2 [92] and may contribute to the unstructured nature of the NLS while bound to Kapβ2. Since alanine is a much smaller, nonpolar sidechain, this amino acid change tests whether those interactions and structural changes are required for proper PY-NLS function. Y532V Hrp1 and P531A Hrp1 were tagged at
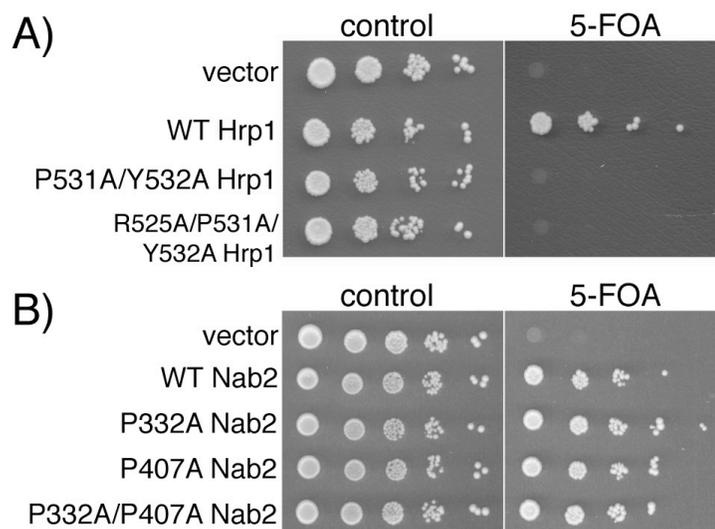
FIGURE 3.4 **The PY-NLS-like sequence within Hrp1 is required for protein function.** Protein function *in vivo* was assessed by a plasmid shuffle assay as described in Experimental Procedures. *A*, *ΔHRP1* cells (ACY1571) maintained by a plasmid encoding wild-type Hrp1 and expressing either wild-type or mutant Hrp1 proteins were serially diluted, spotted onto control or 5-FOA plates, and grown at 30°C for 3 days. *B*, *ΔNAB2* cells (ACY427) maintained by a plasmid encoding wild-type Nab2 and expressing either wild-type or mutant Nab2 proteins were serially diluted, spotted onto control or 5-FOA plates, and grown at 30°C for 3 days.

the N-terminus with GFP and localized in wild-type cells (Figure 3.5A). Much like the other Hrp1 PY-NLS mutants, GFP-Hrp1 Y532V was mislocalized to the cytoplasm, indicating that proper function of the PY-NLS-like sequence was disrupted. Therefore, valine cannot substitute for tyrosine in a functional yeast PY-NLS-like sequence. GFP-Hrp1 P531A was partially mislocalized to the cytoplasm, but also retained some nuclear accumulation. This intermediate phenotype suggests that the proline residue may play a smaller, but still important, role in the Hrp1 PY-NLS. Immunoblotting verified that the GFP-Hrp1 variants were expressed at approximately equal levels (data not shown).

The function of the Y532V and P531A Hrp1 mutants was assessed using the plasmid shuffle assay (Figure 3.5B). Cells expressing Y532V Hrp1 as the only form of Hrp1 are non-viable, indicating that the tyrosine residue in the C-terminal PY-core of the Hrp1 PY-NLS-like sequence is required for Hrp1 function. Cells expressing P531A Hrp1 are viable, but have a slight growth defect at 30°C. To determine whether cells expressing P531A Hrp1 as the only copy of Hrp1 are temperature-sensitive for growth, a quantitative growth assay was performed at 37°C (Figure 3.5C). Results of the analysis demonstrate that P531A Hrp1 cells grow much more slowly than wild-type cells at an elevated temperature.

*The prevalence of the PY-NLS in* S. cerevisiae- This study has analyzed the PY-NLS-like sequences of the Kap104 cargos, Nab2 and Hrp1. The PY-NLS-like sequences of Nab2 are not functional NLSs *in vivo*; however, it is important to note that Nab2 would not have been predicted to contain a putative PY-NLS. Both of the PY-NLS-like sequences within Nab2 vary from the consensus in the final tyrosine, a residue which we have shown to be essential for proper PY-NLS function *in vivo* (Figure 3.5A and 3.5B).
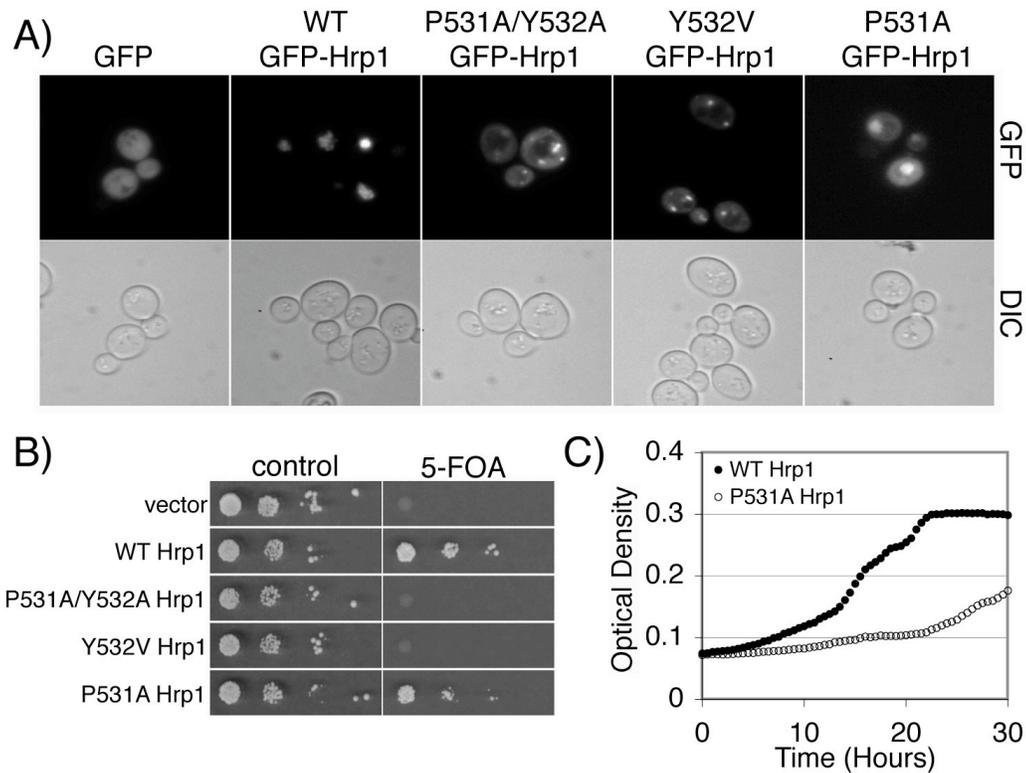
FIGURE 3.5 **Functional dissection of the PY-NLS motif.** *A*, wild-type cells expressing GFP alone, wild-type GFP-Hrp1, GFP-Hrp1 P531A/Y532A, GFP-Hrp1 Y532V, or GFP-Hrp1 P531A were examined by direct fluorescence microscopy (GFP). Corresponding differential interference contrast (DIC) images are shown. *B*, *ΔHRP1* cells (ACY1571) maintained by a plasmid encoding wild-type Hrp1 and expressing either wild-type or mutant Hrp1 protein were spotted onto control or 5-FOA plates and grown at 30°C for 3 days. *C,* wild-type Hrp1 or P531A Hrp1 cells were monitored for growth over time at 37°C as described in Experimental Procedures.

In contrast, we find that Hrp1 does contain a functional PY-NLS, demonstrating that the PY-NLS motif is conserved in yeast. The conservation of the PY-NLS expands our current arsenal of yeast NLS consensus sequences and, combined with a bioinformatics approach, allows us to undertake a preliminary search for new putative Kap104 cargos in *S. cerevisiae*. Additionally, in Figure 3.2C, we showed that not only could the entire PY-NLS-like region of Hrp1 mediate nuclear import, but also that a minimal C-terminal PY-core could effect nuclear localization. Therefore, given that both of these sequences can mediate nuclear targeting *in vivo*, to determine the prevalence of putative PY-NLS-containing proteins in yeast, we queried two datasets using the established hydrophobic and basic PY-NLS consensus sequences as well as a shorter sequence consisting of just the core $R/H/KX_{2-5}PY$ motif (Figure 3.6). First, we searched the yeast proteome as represented by the 5,850 proteins in the *S. cerevisiae* GenBank database [120] to reveal the entire complement of potential Kap104 cargos. Second, we scanned the 1,515 proteins localized at steady-state to either the nucleus or the nucleolus as determined by a comprehensive localization study utilizing the global yeast GFP-fusion library [104]. This more focused dataset was chosen because all nuclear proteins should contain nuclear targeting sequences, a subset of which should contain PY-NLSs. If one considers the C-terminal PY-core sufficient to mediate import, 1,404 proteins or 24% of the total yeast proteome contain at least one predicted PY-NLS sequence. Additionally, 374 nuclear proteins, about 25% of the nuclear proteome, contain putative PY-NLSs. A selection of proteins that are likely candidates for Kap104 import are presented in Table 3.2, which contains a list of yeast proteins that are nuclear or nucleolar, contain a putative

hydrophobic or basic PY-NLS, and lack a putative classical NLS [53].  A detailed table

containing all of the results is available on the Corbett Lab website.
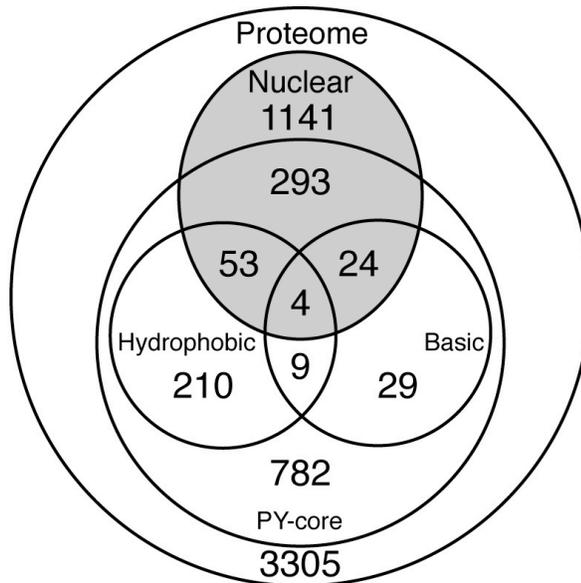
| Predicted PY-NLS | Proteome | | Nuclear | |
|---|---|---|---|---|
| C-terminal PY-core only | 1075 | 18.4% | 293 | 19.3% |
| Hydrophobic | 263 | 4.5% | 53 | 3.5% |
| Basic | 53 | 0.9% | 24 | 1.6% |
| Hydrophobic and Basic | 13 | 0.2% | 4 | 0.3% |
| None | 4446 | 76.0% | 1141 | 75.3% |
| Total | 5850 | 100% | 1515 | 100% |

FIGURE 3.6 **The prevalence of predicted PY-NLS proteins in *S. cerevisiae.*** Algorithms for hydrophobic and basic PY-NLSs or for the C-terminal PY-NLS core motif (see Experimental Procedures) were used to search the 5,850 proteins in the yeast proteome (Proteome) and the 1,515 proteins that are nuclear or nucleolar at steady-state (Nuclear). The results of this preliminary analysis are plotted in a Venn diagram and summarized in the chart below. Hydrophobic and basic PY-NLSs, by definition, also contain the C-terminal PY-core motif, so proteins denoted as containing "Hydrophobic" or "Basic" PY-NLSs also contain the PY-NLS core.

TABLE 3.2 **Nuclear or nucleolar yeast proteins that contain putative hydrophobic or basic PY-NLSs, but lack putative classical NLSs**

| Gene | Systematic Name | Function (adapted from the Saccharomyces Genome Database) |
|---|---|---|
| APC4 | YDR118W | Subunit of the Anaphase-Promoting Complex/Cyclosome (APC/C) |
| APD1 | YBR151W | Protein of unknown function, required for normal localization of actin patches and for normal tolerance of sodium ions and hydrogen peroxide |
| ENP1 | YBR247C | Protein associated with U3 and U14 snoRNAs, required for pre-rRNA processing and 40S ribosomal subunit synthesis |
| FPR1 | YNL135C | Peptidyl-prolyl cis-trans isomerase (PPIase) |
| GAL80 | YML051W | Transcriptional regulator involved in the repression of GAL genes in the absence of galactose |
| HRR25 | YPL204W | Protein kinase involved in regulating diverse events including vesicular trafficking, DNA repair, and chromosome segregation |
| NRG1 | YDR043C | Transcriptional repressor that recruits the Cyc8p-Tup1p complex to promoters; mediates glucose repression |
| PPX1 | YHR201C | Exopolyphosphatase, hydrolyzes inorganic polyphosphate (poly P) into Pi residues |
| PRP11 | YDL043C | Subunit of the SF3a splicing factor complex, required for spliceosome assembly |
| QNS1 | YHR074W | Glutamine-dependent NAD(+) synthetase |
| RHR2 | YIL053W | Constitutively expressed isoform of DL-glycerol-3-phosphatase; involved in glycerol biosynthesis |
| RPO21 | YDL140C | RNA polymerase II largest subunit B220, part of central core; phosphorylation regulates association with transcription and splicing factors |
| RPS14B | YJL191W | Ribosomal protein 59 of the small subunit, required for ribosome assembly and 20S pre-rRNA processing |
| RRP7 | YCL031C | Essential protein involved in rRNA processing and ribosome biogenesis |
| SCC2 | YDR180W | Subunit of cohesin loading factor (Scc2p-Scc4p), a complex required for the loading of cohesin complexes onto chromosomes |
| SGV1 | YPR161C | Cyclin-dependent protein kinase that functions in transcriptional regulation; phosphorylates the C-terminal domain of Rpo21p, which is the largest subunit of RNA polymerase II |
| SIN4 | YNL236W | Subunit of the RNA polymerase II mediator complex; associates with core polymerase subunits to form the RNA polymerase II holoenzyme |
| SOK1 | YDR006C | Protein whose overexpression suppresses the growth defect of mutants lacking protein kinase A activity; involved in cAMP-mediated signaling |
| SPT21 | YMR179W | Protein required for normal transcription at several loci including HTA2-HTB2 and HHF2-HHT2; involved in telomere maintenance |
| TAH11 | YJR046W | DNA replication licensing factor, required for pre-replication complex assembly |
| TFB3 | YDR460W | Subunit of TFIIH and nucleotide excision repair factor 3 complexes, involved in transcription initiation, required for nucleotide excision repair |
| URA5 | YML106W | One of two orotate phosphoribosyltransferase isozymes that catalyze the fifth enzymatic step in de novo biosynthesis of pyrimidines |
| URH1 | YDR400W | Uridine nucleosidase, cleaves N-glycosidic bonds in nucleosides; involved in recycling pyrimidine deoxy- and ribonucleosides via the pyrimidine salvage pathway |
| YSH1 | YLR277C | Putative endoribonuclease, subunit of the mRNA cleavage and polyadenylation specificity complex required for 3' processing of mRNAs |

**Discussion**

In this study, we demonstrate that the PY-NLS, a human Kapβ2-binding nuclear

targeting sequence [92], is conserved in *S. cerevisiae* and show for the first time in any

organism that the PY-NLS is a functional nuclear import signal *in vivo*. Through

localization, function, and binding studies, we show that the PY-NLS-like sequence

within Hrp1, a cargo of the yeast ortholog of Kapβ2, is necessary and sufficient for

nuclear import and is required for receptor binding and for protein function. These

experiments indicate that the PY-NLS-like sequence within Hrp1 is a true NLS and,

significantly, build on previous *in vitro* binding studies to demonstrate that this PY-NLS

can actually mediate import of a protein cargo *in vivo*.

In contrast, we found that the putative PY-NLS-like sequences within Nab2 are

not functional NLS motifs, suggesting that Nab2 must have alternative mechanisms for

gaining entry into the nucleus. These mechanisms may involve several different import

receptors; however, since Figure 3.1B and previous studies [112, 113] show that proper

Nab2 localization requires Kap104, these results more directly suggest that Kap104 can

interact with different classes of cargo proteins using different mechanisms. In the first

mode of interaction, Kap104 recognizes cargo proteins like Hrp1 via PY-NLS motifs and

in the second mode of interaction, Kap104 recognizes cargo proteins like Nab2 via

interactions like that mapped within the RGG domain of Nab2 [112, 113].

Our study also assessed the contribution of the signature residues of the PY-NLS,

the proline and tyrosine, to the function of the Hrp1 PY-NLS. We found that changing

the proline residue to an alanine caused partial mislocalization of Hrp1 accompanied by

temperature sensitive growth, suggesting that, though not essential, the proline is required for proper localization and function of the Hrp1 protein. This moderate effect on Hrp1 localization and function may be due to the position of the PY-NLS only two residues from the C-terminus of Hrp1. Typically, proline residues are assumed to contribute to the unstructured nature of NLSs while bound to their receptors. The proline within the PY-NLS of Hrp1 may not be absolutely required for PY-NLS function because the end of the protein may already lack any significant structure. However, a proline residue could make more significant contributions to PY-NLS sequences within other cargos where the motif is located in a more structured region of the protein. Changing the tyrosine in the Hrp1 PY-NLS to a valine completely abrogated PY-NLS function, suggesting that the change in the terminal residue of the PY-NLS-like sequences within Nab2 (PV versus PY) may be responsible for the lack of function of these putative PY-NLSs *in vivo*.

Interestingly, the PY-NLS sequence of Hrp1 does not exactly match the consensus proposed by Lee et al. for human Kapβ2 binding [92]; the C-terminal PY-core precisely matches, but the pattern of upstream basic residues differs from the established motif. Combined with the results of Figure 3.2C where the C-terminal PY-core of Hrp1 alone can mediate import, this variability suggests that the requirements for recognition of yeast PY-NLSs by Kap104 may be slightly less stringent than the requirements for recognition of a PY-NLS by human Kapβ2. It is important to note, however, that the Hrp1 PY-core alone does not mediate proper Hrp1 localization—there is still significant cytoplasmic signal in addition to the nuclear accumulation. For proteins that require strict nuclear localization, the C-terminal PY-core may not be sufficient for proper function even in yeast; however, for many proteins, establishing a nuclear presence to

any degree may be sufficient for the protein to accomplish its role in the cell.  Therefore, we propose that the minimal PY-NLS in yeast consists of the C-terminal portion of the human PY-NLS, R/H/KX$_{2-5}$PY, with upstream basic or hydrophobic residues enhancing the ability of the NLS to mediate nuclear import.

Given the variability in the evolving definition of the consensus sequence for the PY-NLS, caution should be used in predicting PY-NLS motifs.  However, a preliminary estimate of the likely prevalence of the PY-NLS in yeast using the consensus sequence for the minimal yeast PY-NLS as well as the established consensus sequences for basic and hydrophobic PY-NLSs revealed that 24% of the *S. cerevisiae* proteome and about 25% of the proteins that localize to the nucleus or nucleolus at steady-state contain a putative PY-NLS.  Importantly, this analysis predicts that Hrp1 should contain a putative PY-NLS and, indeed, our studies show that this PY-NLS is functional *in vivo*.  In contrast, the non-functional PV-containing PY-NLS-like sequences in Nab2 are not identified in this analysis.  If we consider only the longer, more stringent hydrophobic or basic PY-NLS consensus sequences, about 6% of the yeast proteome and about 5% of the nuclear proteins contain at least one PY-NLS.  For comparison, about 45% of the yeast proteome and about 57% of nuclear proteins contain a putative classical NLS [53].  Obviously, the prevalence of putative PY-NLS-containing proteins is much lower than the prevalence of putative cNLS-containing proteins, but the fraction of putative yeast PY-NLS cargos is still substantial.  Examination of the twenty-four yeast proteins that are nuclear, contain a putative hydrophobic or basic PY-NLS, and lack a cNLS revealed that seven are implicated in RNA biogenesis or trafficking (Table 3.2).  Many Kapβ2 cargos are also involved in RNA processing in higher eukaryotes [92].  This similarity brings up

the tantalizing possibility that nuclear import via a PY-NLS motif may be a feature common to RNA-related processes in all eukaryotes. However, more functional analyses of predicted PY-NLS cargos are clearly required before we can begin to reliably define the contribution of this import pathway to establishing the nuclear proteome *in vivo*.

# CHAPTER 4

## The Role of Cse1 and Nup2 in Classical NLS-Cargo Release

This chapter is adapted from the following two papers:

**The role of Cse1 and Nup2 in classical NLS-cargo release**

Allison Lange[1], Murray Stewart[2], and Anita H. Corbett[1]

[1]Department of Biochemistry, Emory University, School of Medicine, Atlanta, GA 30322
[2]MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

In preparation.

**Structural basis for Nup2p function in cargo release and karyopherin recycling in nuclear import**

Yoshiyuki Matsuura[1], Allison Lange[2], Michelle T. Harreman[2], Anita H. Corbett[2] and Murray Stewart[1]

[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
[2]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322-3050, USA

**Introduction**

In most systems of nuclear transport, the delivery of protein cargo is coordinated

by the small GTPase, Ran (reviewed in [121]). Ran exists in a GDP-bound state in the

cytoplasm and a GTP-bound state in the nucleus [34, 35]. This asymmetric distribution is

the result of the compartmentalization of the Ran GTPase activating protein (RanGAP) in

the cytoplasm and the Ran guanine nucleotide exchange factor (RanGEF) in the nucleus

[30-33]. As a consequence of this uneven localization, Ran acts like a switch, controlling

the ability of transport receptors to bind their cargo. Therefore, import complexes form

in the cytoplasm in the absence of RanGTP and are dissociated in the nucleus in the

presence of RanGTP and, conversely, export complexes form in the presence of RanGTP

and dissociate upon hydrolysis of RanGTP to RanGDP in the cytoplasm.

In the most well-studied system of nuclear import, the classical nuclear import

pathway, NLS-bearing protein cargo is bound by a heterodimer consisting of importin β

and an adaptor protein, importin α [53]. Importin α recognizes classical NLSs (cNLSs),

which consist of either one (monopartite) or two (bipartite) stretches of basic amino acids

[4, 48, 49]. Much like in other systems of nuclear protein import, Ran is implicated in

the dissociation of cNLS-cargos from their import receptor. Classical protein cargo binds

to the heterodimeric import receptor and translocates into the nucleus where it encounters

RanGTP. This GTP-bound form of Ran binds to importin β, causing a conformational

change [41, 65, 66] that releases importin β, resulting in a transient cNLS-cargo/importin

α complex. The newly-freed auto-inhibitory region on importin α (also called the

importin β-binding region or IBB) then competes for binding with the cNLS, delivering the protein cargo into the nucleus [39, 40, 50, 58, 122].

The effect of RanGTP coupled with the auto-inhibitory domain is sufficient to modulate the affinity of cNLS-cargo for importin α *in vitro*; however, the spontaneous dissociation of the cNLS-cargo/importin α/importin β complex occurs too slowly to account for the rapid release of protein cargo required *in vivo* [41]. Interestingly, two other karyopherin release factors (KaRFs) have been implicated in the release of cNLS-cargo: the export factor for importin α, Cse1, and the nucleoporin, Nup2. Cse1 is a member of the β-karyopherin superfamily of receptors, has an N-terminal RanGTP-binding domain, and is constructed from twenty tandem HEAT repeats [78]. Nup2 is a natively unfolded, mobile nucleoporin consisting of an N-terminal domain, a central FG-repeat domain that binds importin β and the nucleoporin Nup60, and a C-terminal Ran-binding motif [68-72]. Cse1 and Nup2 can both accelerate cNLS-cargo release *in vitro* [41] and mutations in both proteins cause synthetic growth defects when paired with auto-inhibitory mutations in importin α, pointing to their involvement in nuclear protein delivery *in vivo* [61].

Crystal structures of the Cse1/RanGTP/importin α complex and of the Nup2/importin α complex have recently been solved [42-44], which provide detailed predictions of how Cse1 and Nup2 effect cargo release in the nucleus. However, all crystal studies are necessarily *in vitro* and exist outside of the context of the cell. In this study, we test the validity of these structural models *in vivo* by creating amino acid substitutions in residues predicted to be essential for complex formation and assaying their effects on protein localization, function, and interaction. We find that reversing the

charge of the R44 residue of importin α or the D220 residue of Cse1, which targets an interaction between the IBB domain of importin α and the N-terminal arch of Cse1, results in loss of function of the two proteins. The importin α and Cse1 reversal-of-charge variants also fail to interact *in vivo*. Amino acid substitutions in basic residues of the N-terminus of Nup2, which target two regions of Nup2 predicted to interact with either the minor cNLS-binding pocket or with the C-terminus of importin α, disrupt both the function of Nup2 and its interaction with importin α *in vivo*. These results suggest that the structures of Cse1/RanGTP/importin α and Nup2/importin α are representative of the state of the complexes *in vivo* and allow creation of a mechanistic model of cNLS-cargo delivery in the nucleus.

**Experimental Procedures**

*Strains, Plasmids, and Chemicals*- All chemicals were obtained from US Biological or Sigma unless otherwise noted. All media was prepared and all DNA manipulations were performed according to standard procedures [99, 100]. All yeast strains and plasmids used in this study are summarized in Table 4.1.

*Microscopy*- GFP-fusion proteins were localized in live cells using direct fluorescence microscopy. The GFP signal was visualized using a GFP-optimized filter (Chroma Technology) on an Olympus BX60 epifluorescence microscope equipped with a Photometrics Quantix digital camera. Cells expressing genes from their own promoters were grown overnight in selective media, diluted in fresh media, and grown for 3 hours prior to localization studies.

TABLE 4.1 **Strains and plasmids used in Chapter 4**

| Strain/Plasmid | Description | Origin |
|---|---|---|
| FY23 (ACY192) | Wild type, MATa *ura3-52 leu2Δl trp1* | [106] |
| ACY324 | ΔSRP1::HIS3 [SRP1 *CEN URA Amp$^R$*], MATα *leu lys* | [40] |
| ACY685 | ΔSRP1::HIS3 [SRP1 *CEN URA Amp$^R$*], MATa *leu trp* | This Study |
| ACY712 | ΔNUP2::*Kan$^R$* Importin α-GFP::HIS3 [SRP1 *CEN URA Amp$^R$*], *leu* | [42] |
| ACY789 | ΔNUP2 srp1-31 [SRP1 *CEN URA Amp$^R$*], MATa *ura leu his trp lys* | [42] |
| ACY1237 | ΔCSE1::*Kan$^R$* [CSE1 *CEN URA Amp$^R$*], MATa *his3Δl leu2Δ0* | This Study |
| ACY1552 | ΔNUP2::HIS3 ΔNUP133::*Kan$^R$* [NUP2 *CEN URA Amp$^R$*], MATa *leu ura* | This Study |
|  |  |  |
| pRS315 (pAC3) | *CEN LEU Amp$^R$* | [117] |
| pRS305 (pAC7) | *2µ LEU Amp$^R$* | [117] |
| pAC592 | Importin β, *2µ TRP Amp$^R$* | [40] |
| pAC855 | Importin α A3, *CEN LEU Amp$^R$* | [40] |
| pAC856 | Importin α, *CEN LEU Amp$^R$* | [40] |
| pAC858 | Importin α A55, *CEN LEU Amp$^R$* | [40] |
| pAC883 | Importin α-GFP, *CEN URA Amp$^R$* | [40] |
| pAC890 | Importin α-GFP A3, *CEN URA Amp$^R$* | [40] |
| pAC891 | Importin α-myc, *CEN URA Amp$^R$* | [40] |
| pPS1176 (pAC958) | Cse1, *2µ LEU Amp$^R$* | P.A. Silver |
| pAC1215 | 1-51 Nup2-GFP, *CEN LEU Amp$^R$* | [42] |
| pAC1268 | Nup2 Δ50, *CEN LEU Amp$^R$* | [42] |
| pAC1271 | Cse1-GST, *Amp$^R$* | P.A. Silver |
| pAC1298 | Nup2, *CEN URA Amp$^R$* | [42] |
| pAC1303 | Cse1, *2µ TRP Amp$^R$* | [40] |
| pAC1342 | Nup2, *CEN LEU Amp$^R$* | [42] |
| pAC1385 | Nup2, *2µ TRP Amp$^R$* | [40] |
| pAC1394 | Nup2-GFP Δ50, *2µ URA Amp$^R$* | [42] |
| pAC1395 | Nup2-GFP, *2µ URA Amp$^R$* | This Study |
| pAC1880 | Importin α R44A, *CEN LEU Amp$^R$* | This Study |
| pAC1881 | Importin α R44E, *CEN LEU Amp$^R$* | This Study |
| pAC1882 | Importin α R44D, *CEN LEU Amp$^R$* | This Study |
| pAC1883 | Importin α R44K, *CEN LEU Amp$^R$* | This Study |
| pAC1884 | Importin α-GFP R44A, *CEN URA Amp$^R$* | This Study |
| pAC1885 | Importin α-GFP R44E, *CEN URA Amp$^R$* | This Study |
| pAC1886 | Importin α-GFP R44D, *CEN URA Amp$^R$* | This Study |
| pAC1887 | Importin α-GFP R44K, *CEN URA Amp$^R$* | This Study |
| pAC1892 | Cse1 D220A, *2µ LEU Amp$^R$* | This Study |
| pAC1893 | Cse1 D220R, *2µ LEU Amp$^R$* | This Study |
| pAC1894 | Cse1 D220K, *2µ LEU Amp$^R$* | This Study |
| pAC1895 | Cse1 D220N, *2µ LEU Amp$^R$* | This Study |
| pAC1896 | Cse1-GST D220A, *Amp$^R$* | This Study |
| pAC1897 | Cse1-GST D220R, *Amp$^R$* | This Study |
| pAC1898 | Cse1-GST D220K, *Amp$^R$* | This Study |
| pAC1899 | Cse1-GST D220N, *Amp$^R$* | This Study |
| pAC2283 | Nup2 K3E/R4D, *CEN LEU Amp$^R$* | This Study |
| pAC2286 | Nup2 R38A/R39A R46A/R47A, *CEN LEU Amp$^R$* | This Study |
| pAC2289 | Nup2 K3E/R4D R38A/R39A R46A/R47A, *CEN LEU Amp$^R$* | This Study |
| pAC2290 | Nup2-GFP K3E/R4D, *2µ URA Amp$^R$* | This Study |
| pAC2293 | Nup2-GFP R38A/R39A R46A/R47A, *2µ URA Amp$^R$* | This Study |
| pAC2296 | Nup2-GFP K3E/R4D R38A/R39A R46A/R47A, *2µ URA Amp$^R$* | This Study |

In Vivo *Functional Analysis*- The *in vivo* function of each of the importin α, Nup2, or Cse1 variants was tested using a plasmid shuffle technique [119]. Plasmids encoding wild-type or variant importin α, Nup2, or Cse2 were transformed into *ΔSRP1* (ACY324), *ΔCSE1* (ACY1237), or *ΔNUP2/ΔNUP133* (ACY1552) cells containing a wild-type *SRP1*, *CSE1,* or *NUP2 URA3* plasmid. Single transformants were grown to saturation in liquid culture, serially diluted (1:10) in dH$_2$O, and spotted on control ura- leu- glu plates or on selective leu- glu plates containing 5-fluoroorotic acid (5-FOA). Plates were incubated at 18°C, 25°C, 30°C, or 37°C for 3-5 days. To clarify the results of the Nup2 function assay, cells were picked from the 25°C 5-FOA plate, grown to saturation at 25°C, serially diluted, spotted on leu- glu plates, and incubated at 30°C.

*Expression and purification of the Nup2/importin α complex*- His/S-tagged Nup2 residues 1–51 (Nup2N) were expressed from pET30a (Novagen). Untagged ARM repeat domain of importin α (residues 88–530, ΔIBB importin α) was cloned into pET30a and the Y397D point mutation introduced using Quikchange PCR (Stratagene). For crystallization, Nup2N and ΔIBB importin α were expressed separately in BL21-Gold(DE3) cells and, after lysis and clarification, crude bacterial extracts were mixed in buffer A (20 mM Tris-HCl pH 7.5, 0.3 M NaCl, 1 mM PMSF, 3 mM β-mercaptoethanol), with ΔIBB importin α in excess, and applied to Ni-NTA (Qiagen). The resin was washed extensively with buffer A, followed by buffer A containing 25 mM imidazole, and then exchanged into 20 mM Tris–HCl pH 7.5, 50 mM NaCl, 2 mM CaCl2, 3 mM β-mercaptoethanol and incubated with 7 U/ml enterokinase (Novagen) overnight. The Nup2/importin α complex released from the resin was applied to enterokinase capture

resin (Novagen) and the flow-through fraction concentrated to 39 mg/ml in 10 mM Tris–HCl pH 8.0, 100 mM NaCl, 10% glycerol.

*Crystallization, data collection and structure determination*- Crystals of the Nup2/importin α complex were obtained at 18°C by streak seeding hanging drops containing 2 mg/ml protein, 50 mM HEPES pH 6.8, 0.15 M NaCl, 24% PEG3350 and 2% PEG400.  The plate-shaped crystals had P21212 symmetry with a = 129.8 Å, b = 140.1 Å and c = 63.9 Å. A 200×200×400 µm crystal was cryo-protected in 50 mM HEPES pH 6.8, 0.15 M NaCl, 27% PEG3350 and 12% PEG400 and flash-frozen in liquid nitrogen.  Mosaic spread was reduced by freeze-thaw annealing and a 2.6 Å dataset collected on beamline 14.2 at Daresbury (UK) using a MAR CCD detector and 0.978 Å wavelength radiation.  Molecular replacement solutions were found using the Crystallography & NMR System (CNS) [123] for two independent ΔIBB importin α molecules in the asymmetric unit using a chain from the c-myc NLS complex [50] as search model.  The structure was initially refined using CNS.  After rigid body refinement, simulated annealing, positional refinement using conjugate gradient minimization with NCS restraints enforced throughout the molecule, alternating with local rebuilding, the free R-factor was 32.7% (R-factor 29.6%) and unambiguous difference density appeared for the Nup2 fragment along the central groove of ΔIBB importin α on both chains and a model was built for residues 36–51.  The model was then refined using REFMAC5 [124] and strong NCS restraints based on individual ARM repeats.  After iterative cycles of refinement and rebuilding and the addition of 158 waters, the final R-factor was 21.6% (R-free 25.7%).

**Results**

*The crystal structure of importin α/Cse1/RanGTP-* Matsuura, Stewart and

colleagues have solved the crystal structure of importin α bound to its export receptor

Cse1/RanGTP [44] (Figure 4.1A and Figure 4.1B). In this structural model of the export

complex, RanGTP and the C-terminus of importin α are nestled between two arched

regions of Cse1. The IBB domain of importin α then folds back over the body of the

importin α molecule and binds to the major and minor cNLS-binding pockets.

Importantly, the IBB domain also interacts with Cse1 in two key regions. First, the

extreme N-terminus of the auto-inhibitory domain interacts with Cse1 holding the entire

molecule in a "closed" position and second, a region between the two basic clusters of the

IBB domain interacts with the N-terminal Cse1 arch. We predict that the interactions of

the IBB domain with Cse1 are critical to cNLS-cargo delivery into the nucleus and to

recycling of importin α back to the cytoplasm. Presumably, these interactions could

serve two important roles: facilitating the release of cNLS-cargo by clamping importin α

in a closed state, which reinforces the effect of the auto-inhibitory domain and prevents

cargo from re-binding, and providing a mechanism for Cse1 to distinguish between

cargo-bound and cargo-free importin α, which prevents fruitless rounds of transport.

Based on the crystal structure, we hypothesize that the R44 residue of importin α and the

D220 residue of Cse1 are integral to formation of the tertiary complex (Figure 4.1C). To

test this prediction *in vivo*, we created amino acid substitutions in both importin α and

Cse1, which either neutralize or reverse the charge of the relevant amino acid (R44A,

R44E, R44D, or R44K importin α and D220A, D220R, D220K, or D220N Cse1),
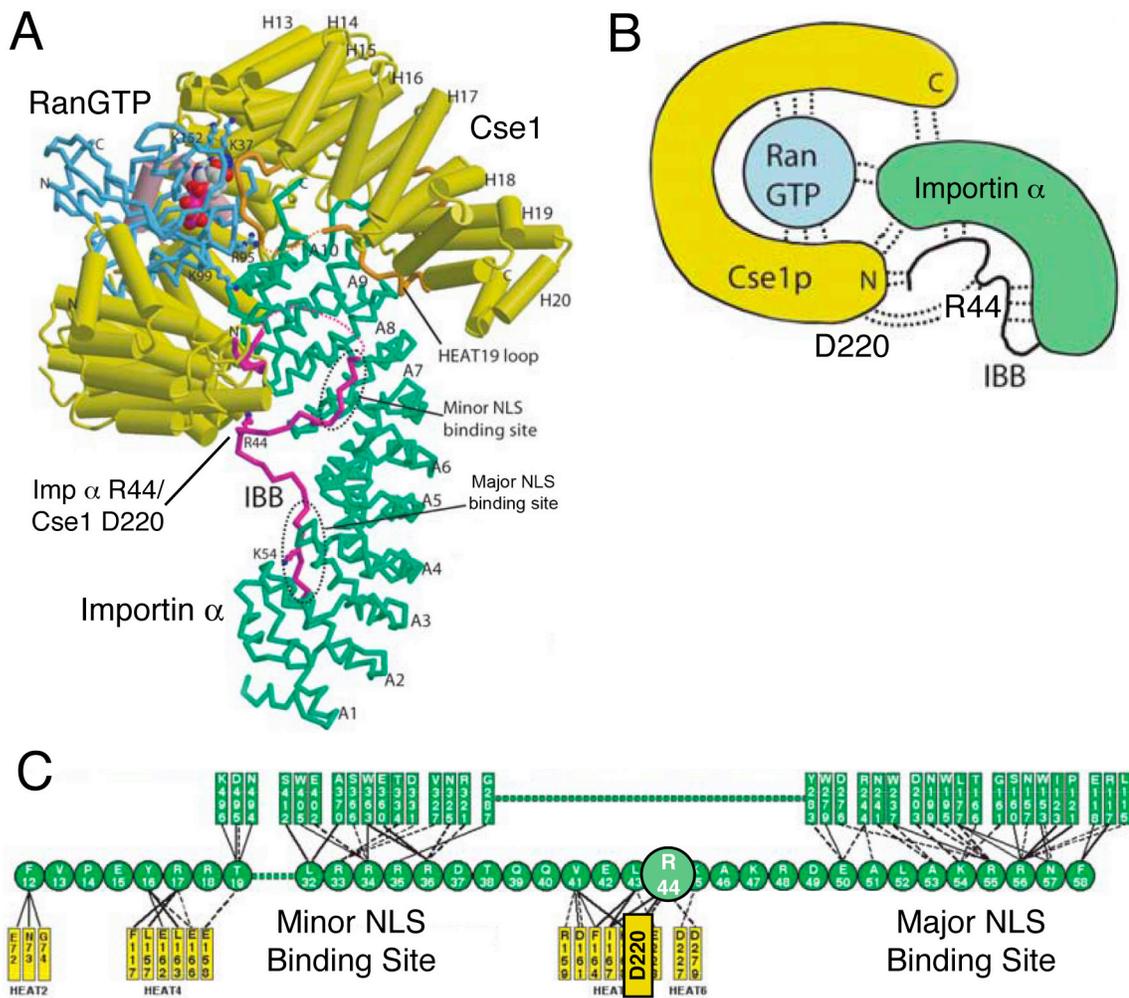
FIGURE 4.1 **Schematic of the importin α/Cse1/RanGTP complex.** *A,* crystal structure of the importin α/Cse1/RanGTP complex [44]. Importin α is shown in green with the IBB domain denoted in purple, Cse1 is shown in yellow, and RanGTP is shown in blue. *B,* cartoon depiction of the trimeric complex. Dotted lines represent interactions. *C,* schematic of the interaction between the IBB domain of importin α (green circles), the body of importin α (green rectangles), and Cse1 (yellow rectangles). Solid lines represent hydrophobic interactions and dashed lines represent H-bonds and salt bridges. All portions of this figure are adapted from [44].

and assessed the function, localization, and interaction capabilities of the variant proteins.

*Importin α-GFP variant localization and expression-* The amino acid changes created in the N-terminus of importin α are within the importin β-binding region of importin α and could potentially disrupt the interaction with importin β, preventing these importin α variants from entering the nucleus.  Since Cse1 interacts with importin α after it has reached the nucleus, this potential mislocalization could preclude interaction between importin α and Cse1.  To verify that the importin α changes do not affect the localization of the variant proteins, each importin α variant was tagged at the C-terminus with GFP and the resulting fusion proteins were visualized in wild-type cells (Figure 4.2A).  Wild-type importin α-GFP and all of the importin α-GFP variants were localized to the nucleus at steady-state, suggesting that the interaction with importin β is not affected in these variants.  Immunoblotting verified that all of the importin α-GFP proteins were expressed at approximately equal levels (Figure 4.2B).

In vivo *function of the importin α and Cse1 variants-* To test the function of the importin α and Cse1 variants *in vivo*, plasmid shuffle assays were performed (Figure 4.3A and Figure 4.3B).  Cells lacking the gene for importin α, *SRP1*, or the gene for Cse1, *CSE1*, and containing wild-type *URA3* maintenance plasmids were transformed with plasmids encoding wild-type importin α or Cse1, variant importin α or Cse1, or vector alone and were plated on control plates or plates containing 5-fluoroorotic acid (5-FOA).  5-FOA is a toxic analog of uracil, which selects against the *URA3*-containing maintenance plasmid [119].  *SRP1* and *CSE1* are essential, so only cells with a functional copy of the gene will grow on selective plates.  *ΔSRP1* cells expressing R44A or R44K
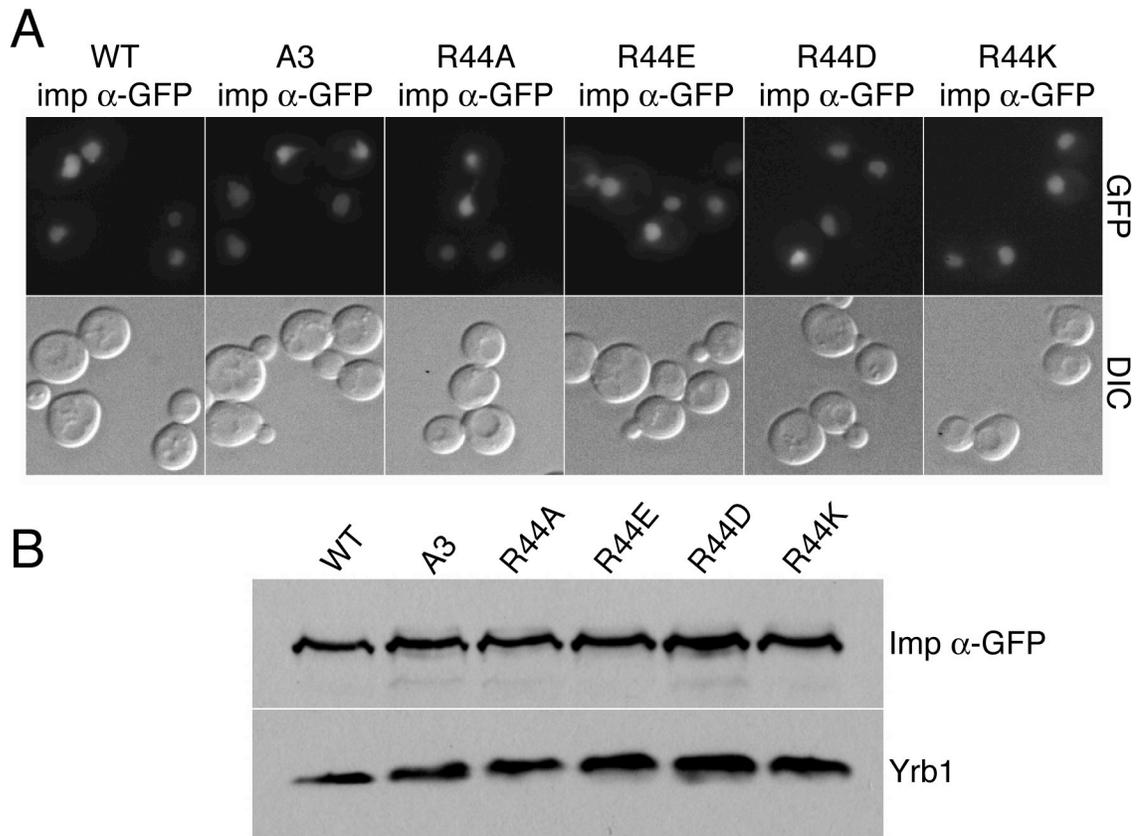
FIGURE 4.2 **Localization and expression level of importin α-GFP variants.** *A*, wild-type cells expressing importin α-GFP, A3 importin α-GFP, or one of the R44 importin α-GFP variants were examined by direct fluorescence microscopy (GFP). Corresponding differential interference contrast (DIC) images are shown. *B*, the importin α-GFP variants are expressed at approximately equal levels. Lysates from wild-type cells expressing each of the importin α-GFP variants were probed with anti-GFP antibodies to detect the importin α fusion proteins and with anti-Yrb1 antibody [125] as a loading control.
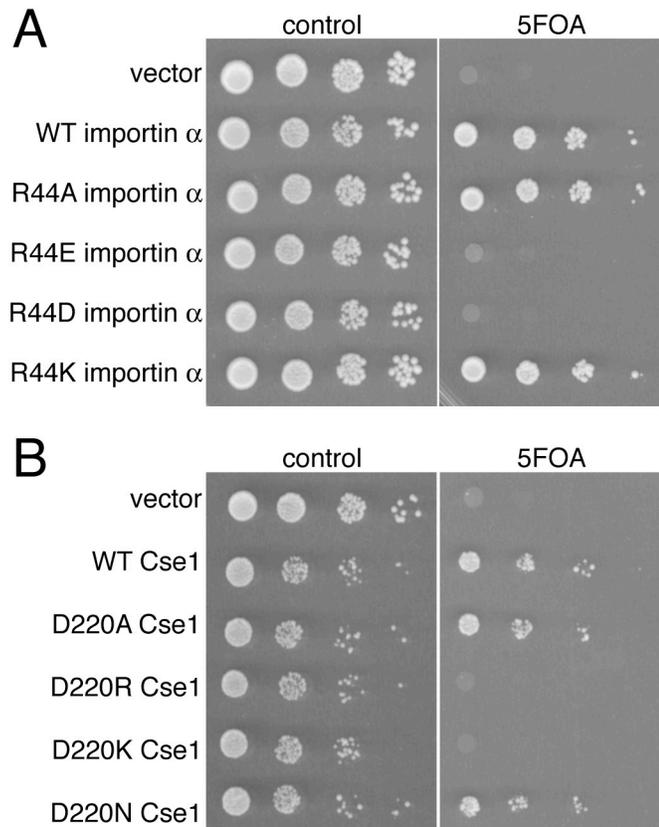
FIGURE 4.3 **Reversing the charge of the importin α R44 or the Cse1 D220 residue renders the protein non-functional.** Protein function *in vivo* was assessed by a plasmid shuffle assay [119] as described in Experimental Procedures. *A, ΔSRP1* cells (ACY324) maintained by a plasmid encoding wild-type importin α and expressing either wild-type or variant importin α were serially diluted, spotted onto control or 5-FOA plates, and grown at 30°C for 3 days. *B, ΔCSE1* cells (ACY1237) maintained by a plasmid encoding wild-type Cse1 and expressing either wild-type or variant Cse1 were serially diluted, spotted onto control or 5-FOA plates, and grown at 30°C for 3 days.

importin α grew like wild-type cells; however, cells expressing R44E or R44D importin α as the sole copy of importin α were unable to grow.  Similarly, *ΔCSE1* cells expressing D220A or D220N Cse1 grew like wild-type cells and cells expressing D220R and D220K Cse1 were inviable.  These results suggest that reversing the charge of the importin α R44 or Cse1 D220 residues interrupts an essential interaction between the proteins.

*Localization of importin α-GFP variants with wild-type Cse1 overexpression*- To assess the interaction between importin α and Cse1 *in vivo*, we can exploit the fact that Cse1 is the export receptor for importin α and assay the steady-state localization of importin α as a measure of the *in vivo* interaction between the two proteins.  In cells with wild-type levels of Cse1, importin α-GFP is localized to the nucleus; however, in cells with elevated levels of Cse1, importin α-GFP is localized throughout the cell, presumably because the excess Cse1 allows for more rapid recycling of importin α-GFP to the cytoplasm [40].  Importantly, this export step requires interaction between the cargo, importin α, and its receptor, Cse1.

In Figure 4.4A, variants of importin α-GFP were localized either in cells with endogenous levels of Cse1 or in cells overexpressing wild-type Cse1.  All of the importin α-GFP variant proteins were localized to the nucleus in cells expressing endogenous Cse1 (data not shown and Figure 4.2A).  In cells overexpressing Cse1, wild-type importin α-GFP was localized throughout the cell and a control, A3 importin α-GFP, which has reduced binding to Cse1 due to a defect in cargo release [40], remained nuclear.  The R44K importin α-GFP variant was localized throughout the cell, indicating that export of this variant was facilitated by excess Cse1 and suggesting that a functional interaction
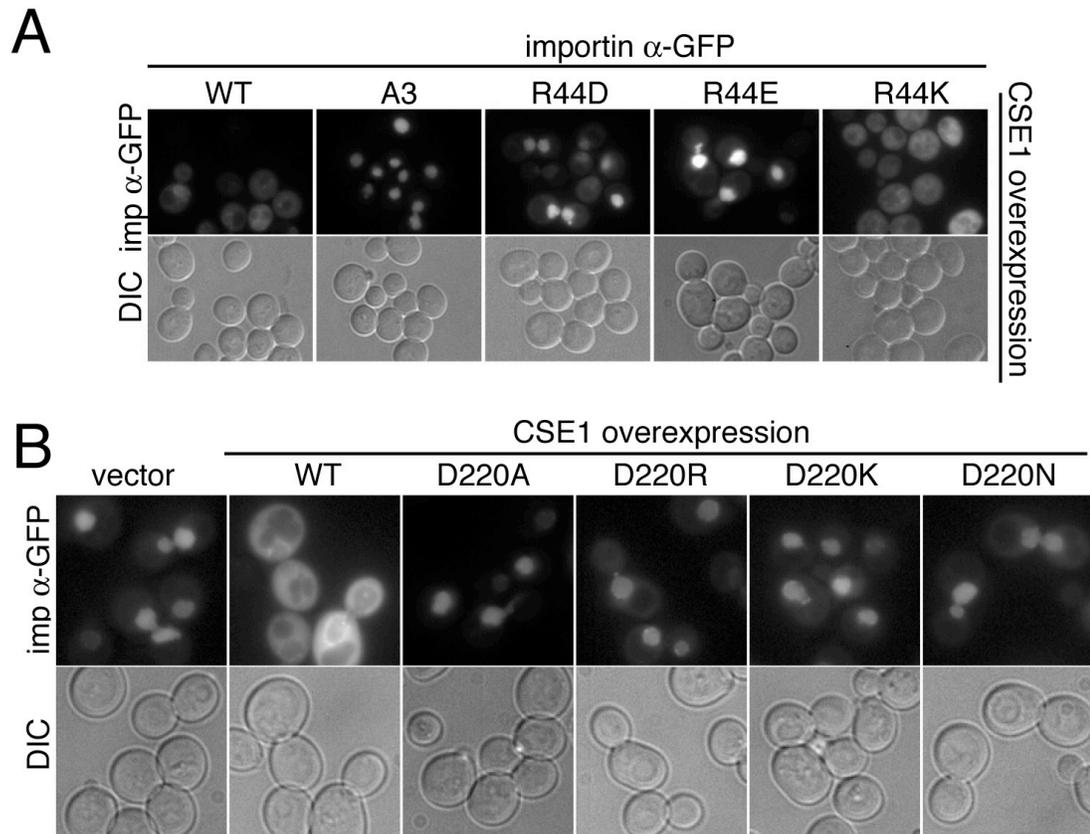
FIGURE 4.4 ***In vivo* assessment of the interaction between importin α and Cse1.** *A,* wild-type cells expressing each of the importin α-GFP variants and overexpressing wild-type Cse1 were examined by direct fluorescence microscopy (imp α-GFP). Corresponding differential interference contrast (DIC) images are shown. *B*, wild-type cells expressing wild-type importin α-GFP and overexpressing each of the Cse1 variants were examined by direct fluorescence microscopy (imp α-GFP). Corresponding differential interference contrast (DIC) images are shown.

with Cse1 was maintained.  In contrast, R44D and R44E importin $\alpha$-GFP were both localized to the nucleus even in cells with Cse1 overexpression, indicating that reversing the charge of the R44 residue of importin $\alpha$ prevents proper interaction with Cse1 *in vivo*.

*Localization of wild-type importin $\alpha$-GFP with Cse1 variant overexpression-* The converse experiment was performed to assess the role of the Cse1 D220 residue in the interaction with importin $\alpha$ *in vivo*.  Wild-type importin $\alpha$-GFP was visualized in cells overexpressing wild-type or D220 variant Cse1 (Figure 4.4B).  As expected, importin $\alpha$-GFP was localized to the cytoplasm in cells overexpressing wild-type Cse1.  However, in cells overexpressing D220A, D220R, D220K, or D220N Cse1, importin $\alpha$-GFP was localized to the nucleus, suggesting that the Cse1 D220 residue is required for productive interaction with wild-type importin $\alpha$ *in vivo*.

*Importin $\alpha$-GFP variant localization with compensatory Cse1 variant overexpression-* Based on the crystal structure of Cse1/RanGTP/importin $\alpha$, we predict that the interaction between the importin $\alpha$ R44 residue and the Cse1 D220 residue is significantly ionic in nature.  This prediction is supported by our finding that reversing the charge of either the R44 or D220 residue results in a loss of function of the importin $\alpha$ or Cse1 protein and in a loss of interaction between the two proteins *in vivo*.  We predict that creating a compensatory amino acid substitution in either R44 or D220 could rescue the interaction between the two proteins.  Accordingly, the interaction between R44D importin $\alpha$-GFP and D220R Cse1 was assessed *in vivo* using the previously described importin $\alpha$ relocalization assay (Figure 4.5).  As was found earlier, in cells overexpressing wild-type Cse1, R44D importin $\alpha$-GFP remains localized to the nucleus.
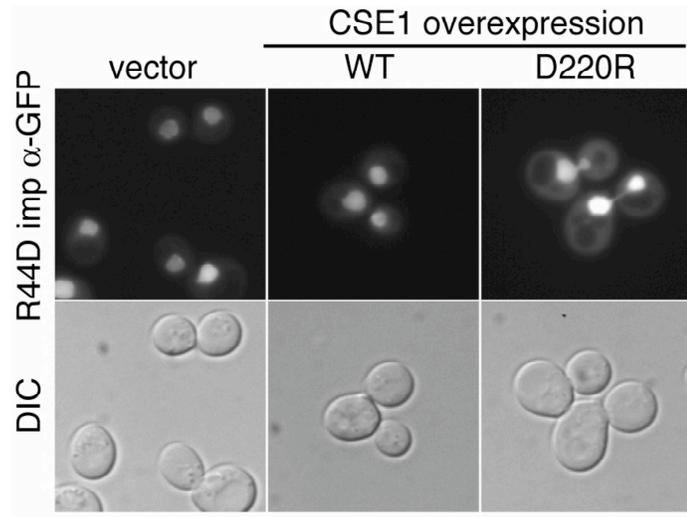
FIGURE 4.5 **Compensatory mutations.** Wild-type cells expressing R44D importin α-GFP and overexpressing either vector, wild-type Cse1, or D220R Cse1 were examined by direct fluorescence microscopy (R44D imp α-GFP). Corresponding differential interference contrast (DIC) images are shown.

However, in cells overexpressing D220R Cse1, R44D importin α-GFP is both nuclear and cytoplasmic. The appearance of this cytoplasmic pool of R44D importin α-GFP suggests that the compensatory amino acid substitution in Cse1 (D220R) can partially rescue the interaction between importin α and Cse1 *in vivo*, indicating that there is indeed a significant ionic component to the R44-D220 interaction.

      *The N-terminus of Nup2 is required for function* in vivo- The nucleoporin Nup2 has also been implicated in NLS-cargo delivery, therefore, we next wanted to explore its contribution to cNLS-cargo release in *S. cerevisiae*. Hood et al. showed that deletion of residues 1-175 inhibited Nup2 targeting to the nuclear envelope as well as binding to importin α and suggested a role for the Nup2 N-terminus in providing an initial NPC docking site along the Cse1/RanGTP-mediated importin α export pathway [71]. Subsequent work suggested that the first 50 residues of Nup2 were critical for importin α binding *in vitro* [126]. To test the functional importance of Nup2 residues 1-50 *in vivo*, a plasmid encoding wild-type Nup2, 1-51aa Nup2, or Δ50 Nup2 or vector alone were transformed into *ΔNUP2 srp1-31* double mutant cells (where Nup2 is absolutely essential for viability [75]) maintained with a *URA3 SRP1* plasmid. The maintenance plasmid was then removed by plasmid shuffle [119] (Figure 4.6A). Expression of full-length Nup2 complemented *ΔNUP2 srp1-31* cells, but expression of Δ50 Nup2 or 1-51aa Nup2 did not, consistent with residues 1-50 of Nup2 being functionally necessary, but not sufficient *in vivo*.

      We exploited the observation that while integrated importin α-GFP is usually localized to the nuclear rim, in cells that lack Nup2, importin α-GFP accumulates within the nuclear interior [70, 71, 75] to probe if residues 1-50 of Nup2 are required for Nup2-
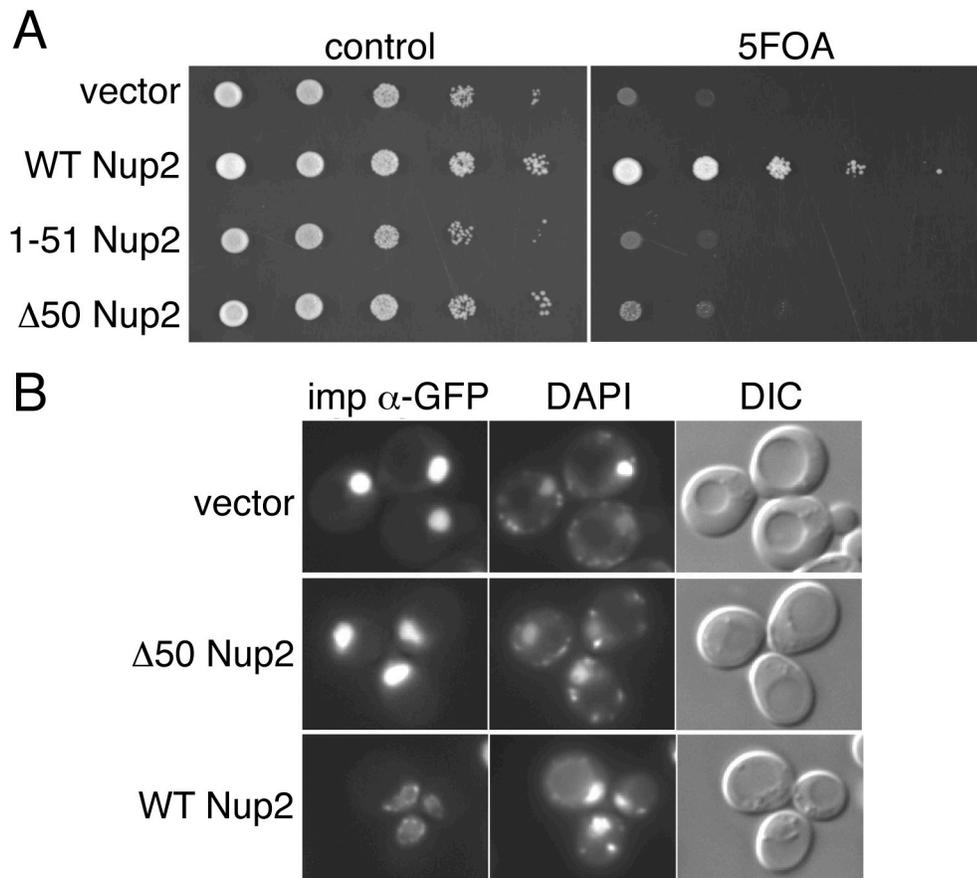
FIGURE 4.6 ***In vivo* functional analysis of the N-terminus of Nup2.** *A*, *ΔNUP2 srp1-31* cells maintained by a plasmid encoding importin α and expressing either full-length Nup2, 1-51 Nup2, or Δ50 Nup2 were spotted onto control plates lacking uracil or 5-FOA plates. 5-FOA eliminates the *URA3* maintenance plasmid encoding importin α [119]. *B*, the Nup2 N-terminus promotes docking of importin α to the nuclear envelope and is required for efficient recycling of importin α to the cytoplasm. Importin α-GFP was integrated at the endogenous SRP1 locus of *ΔNUP2* cells. The cells were then transformed with plasmids encoding either full-length Nup2, Δ50 Nup2, or vector alone, and importin α-GFP was visualized by direct fluorescence microscopy. Corresponding DIC and DAPI images are shown.

dependent recycling of importin α to the cytoplasm. *ΔNUP2* cells were transformed with

plasmids encoding full-length Nup2, Δ50 Nup2, or vector alone; co-stained with DAPI to

mark the position of the nucleus; and visualized (Figure 4.6B). Integrated importin α-

GFP was concentrated at the nuclear rim in cells expressing Nup2, as described [70, 75],

but was localized throughout the nucleus of cells expressing Δ50 Nup2 or vector alone,

consistent with the first 50aa of Nup2 being essential for efficient recycling of importin α

to the cytoplasm. Binding assays have indicated that Δ50 Nup2 does not bind importin α

[126] and that residues 1-51 of Nup2 retain the ability to bind tightly to importin α [42],

suggesting that the effects seen *in vivo* were due, at least in part, to decreased interaction

between Nup2 and importin α.

*The crystal structure of the N-terminus of Nup2 bound to importin α-* Our *in vivo*

data coupled with previous *in vitro* data have indicated that the N-terminal 50aa of Nup2

are involved in the interaction of Nup2 with importin α. Seeking the precise mechanism

by which Nup2 facilitates cNLS-cargo release from importin α, we collaborated to solve

the crystal structure of Nup2 bound to ΔIBB importin α [42, 43] (Figure 4.7A). In this

structure, Nup2 binds to importin α in two main regions: one near the center of importin

α, which partially overlaps the minor cNLS-binding pocket and one on the C-terminus of

importin α, which overlaps the Cse1/RanGTP-binding site. Amino acid substitutions in

the mammalian homolog of Nup2, Nup50/Npap60, combined with *in vitro* binding

studies suggested that basic residues in both binding regions of Nup2 are involved in

binding importin α and dissociating cNLS-cargo [43]. Therefore, we propose that Nup2

acts analogously to the auto-inhibitory domain of importin α, binding to the C-terminus

of importin α, then folding over to compete with cNLS-cargo for binding to importin α,

A Nup2 — Ala2    ΔIBB Importin α

Arg46

B

```
        10          20          30          40          50
MAKRVADAQIQRETYDSNESDDDVTPSTKVASSAVMNRRKIAMPQRRMAFK...
   **                                **        **
```

Binds to NLS pocket          Binds to C-terminus
   of importin α                of importin α

FIGURE 4.7  **The crystal structure of the N-terminus of Nup2 bound to importin α.**
*A*, the crystal structure of the N-terminus of Nup2 bound to ΔIBB importin α [42, 43].
ΔIBB importin α is shown in yellow and Nup2 is shown in green.  *B*, amino acid
sequence of the N-terminus of Nup2.  Residues targeted for amino acid substitution are
denoted with asterisks.  Both parts of this figure are adapted from [43].

thereby contributing to cargo release. To test if the basic residues of Nup2 are required for proper function of the protein and to test the model of Nup2-mediated cargo release *in vivo*, we have created a series of amino acid substitutions in the N-terminus of Nup2 (Figure 4.7B) and will assess their effects on the localization and function of Nup2 and the interaction with importin α.

*Localization of Nup2-GFP variants*- Nup2 shuttles between the nucleus and the cytoplasm, but is nuclear at steady-state [70, 127]. Since Nup2 interacts with importin α in the nucleus as a cargo release factor, it is important that the variant Nup2 proteins maintain some nuclear localization. To test the localization of the Nup2 variant proteins, GFP was fused to the C-terminus of Nup2 and the resulting fusion proteins were visualized in wild-type cells (Figure 4.8A). Wild-type Nup2-GFP, K3E/R4D Nup2-GFP, and R38A/R39A/R46A/R47A Nup2-GFP are all localized to the nucleus. Cells expressing Δ50 Nup2-GFP and K3E/R4D/R38A/R39A/R46A/R47A Nup2-GFP have some cytoplasmic signal, but still show significant nuclear accumulation. Immunoblotting verified that the Nup2-GFP variants were expressed at approximately equal levels (Figure 4.8B)

In vivo *function of Nup2 variants*- To test the function of the Nup2 variants, a plasmid shuffle assay [119] was performed using *ΔNUP2/ΔNUP133* cells containing a wild-type *NUP2 URA3* maintenance plasmid. This strain was used because *NUP2* is not an essential gene, but the combination of *ΔNUP2* and *ΔNUP133* is synthetically lethal [128] at 30°C, creating a genetic background where Nup2 is essential. Cells were transformed with plasmids encoding wild-type Nup2, mutant Nup2, or vector alone and were tested on control or 5-FOA plates (data not shown). To clarify the results of the
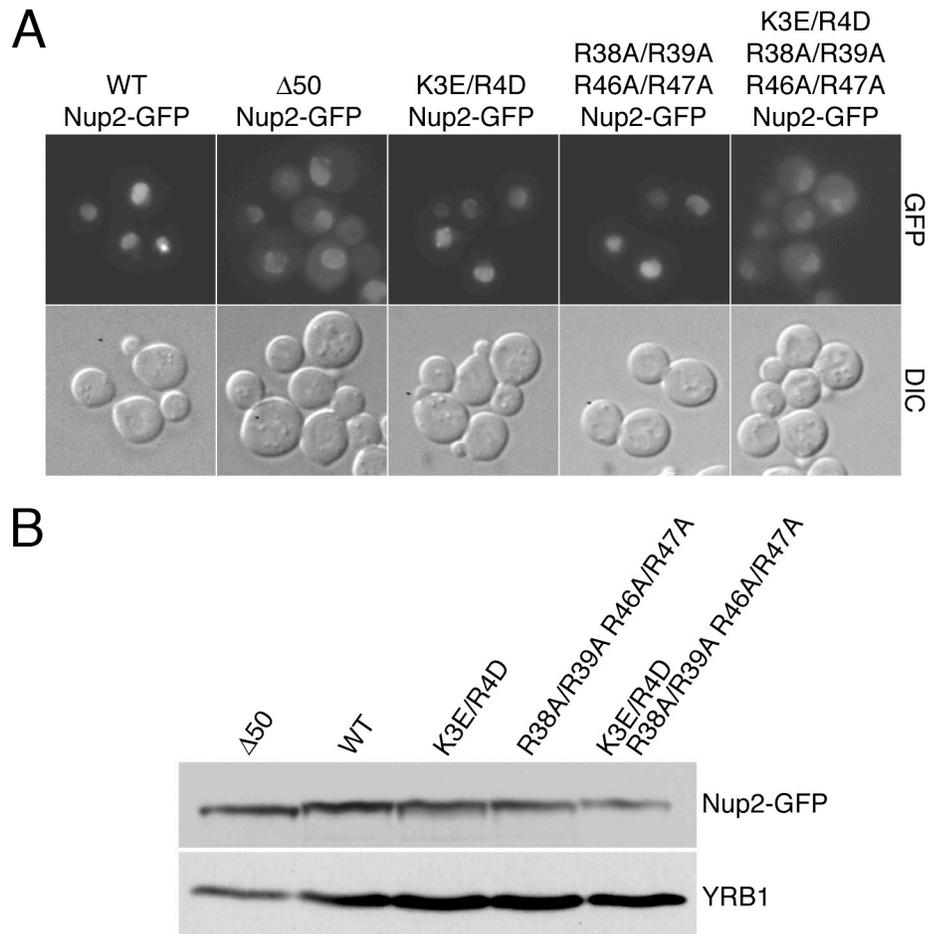
FIGURE 4.8 **Nup2-GFP variant localization and expression levels.** *A*, wild-type cells expressing wild-type Nup2-GFP, Δ50 Nup2-GFP, or one of the Nup2-GFP variants were examined by direct fluorescence microscopy (GFP). Corresponding differential interference contrast (DIC) images are shown. *B,* the Nup2-GFP variants are expressed at approximately equal levels. Lysates from wild-type cells expressing each of the Nup2-GFP variants were probed with anti-GFP antibodies to detect the Nup2 fusion proteins and with anti-Yrb1 antibody [125] as a loading control.

functional assay, cells from the 25°C 5-FOA plate were grown to saturation, diluted, spotted on leu- glu plates, and grown at 30°C (Figure 4.9A). As expected, cells containing Δ50 Nup2 showed slow growth compared to cells containing wild-type Nup2. Cells expressing variants of Nup2 showed a range of growth phenotypes between wild-type and Δ50 Nup2, with K3E/R4D Nup2 cells having a slight growth defect, R38A/R39A/R46A/R47A Nup2 cells having an intermediate phenotype, and the combination K3E/R4D/R38A/R39A/R46A/R47A Nup2 variant cells growing most slowly.

*Effect of Nup2 variants on the localization of importin α-GFP*- To assess the interaction between Nup2 and importin α *in vivo*, we again took advantage of the integrated importin α-GFP relocalization assay. If the point mutations in Nup2 affect the interaction with importin α *in vivo*, we predict that integrated importin α-GFP will be mislocalized to the nucleus in cells expressing the Nup2 variant as the only copy of Nup2. Accordingly, integrated importin α-GFP was localized in *ΔNUP2* cells containing either vector alone or plasmids bearing the Nup2 variants (Figure 4.9B). As expected, integrated importin α-GFP was localized to the nucleus in cells lacking Nup2 and in cells expressing Δ50 Nup2 and was localized to the nuclear rim in cells expressing wild-type Nup2. In cells expressing any of the Nup2 variants, integrated importin α-GFP was mislocalized to the nucleus, suggesting that mutation of either or both of the binding regions of Nup2 disrupts the interaction with importin α *in vivo*.

*Reincarnation*- We have shown that reversing the charge of the R44 residue in importin α results in a loss of function (Figure 4.3A). Further testing suggested that this effect is due to a loss of proper interaction with Cse1 (Figure 4.4A), which could result in

A



B
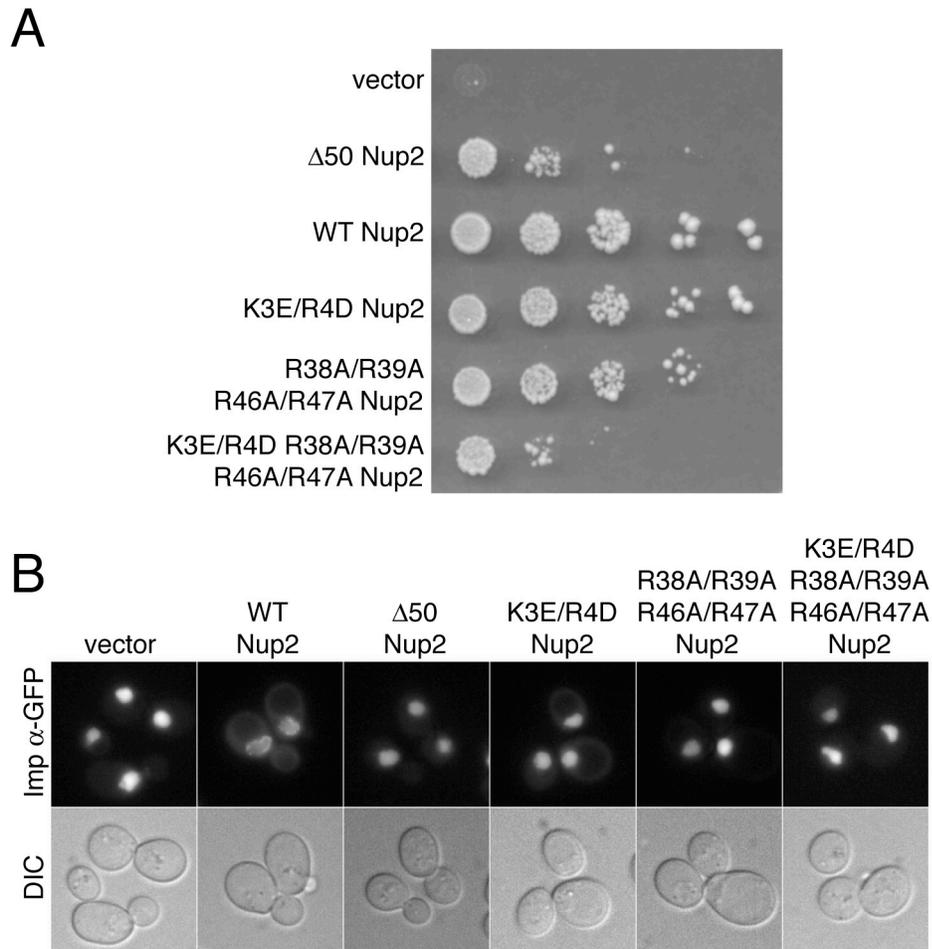


FIGURE 4.9 **Functional analysis of Nup2 variants.** *A*, Nup2 variant protein function *in vivo* was assessed by a plasmid shuffle assay [119] as described in Experimental Procedures. *ΔNUP2/ΔNUP133* cells (ACY1552) maintained by a plasmid encoding wild-type Nup2 and expressing either wild-type Nup2, Δ50 Nup2, or mutant Nup2 were spotted onto 5-FOA plates and grown at 25°C. Cells from the 5-FOA plates were grown to saturation at 25°C, serially diluted, spotted onto selective plates, and grown at 30°C. *B*, *ΔNUP2* cells transformed with wild-type Nup2, Δ50 Nup2, or one of the Nup2 variants and expressing integrated importin α-GFP were examined by direct fluorescence microscopy (imp α-GFP). Corresponding differential interference contrast (DIC) images are shown.

slower rates of cNLS-cargo delivery in the nucleus.  If a decline in the rate of cargo delivery is the cause of the defect, we predict that overexpression of a karyopherin release factor such as Nup2 or Cse1 could rescue the growth phenotype of the R44 importin α reversal-of-charge variants.  Therefore, to test if overexpression of a karyopherin release factor can rescue the R44 importin α variant phenotypes, cells expressing either wild-type or variant importin α were transformed with plasmids that encode either Cse1 or Nup2 and were plated on control plates or plates containing 5-FOA to remove the maintenance plasmid [119] (Figure 4.10).  An empty vector or a plasmid encoding importin β were also transformed as controls.

Cells expressing wild-type importin α grew well at both 18°C and 30°C with and without overexpression of the karyopherin release factors.  As controls, cells expressing A3 importin α, which has a major change in the auto-inhibitory region and pronounced defects in cargo release [40], are dead at both temperatures regardless of which factors are overexpressed and the cold-sensitive phenotype of cells expressing A55 importin α, which has a smaller change in the auto-inhibitory region [61], is rescued by overexpression of either Cse1 or Nup2, but not importin β.  As expected, cells expressing R44A or R44K importin α grow like wild-type cells.  At 30°C, cells expressing R44D or R44E importin α are still inviable; however, at 18°C, the growth of cells expressing R44D or R44E importin α is restored, not only when Cse1 is overexpressed, but also when Nup2 is overexpressed.  This result supports the hypothesis that the R44 importin α amino acid substitutions cause a defect in cNLS-cargo release because upregulating the expression of not only Cse1, but also other karyopherin release factors such as Nup2, can compensate for the effects of the reversal of charge and restore some function.
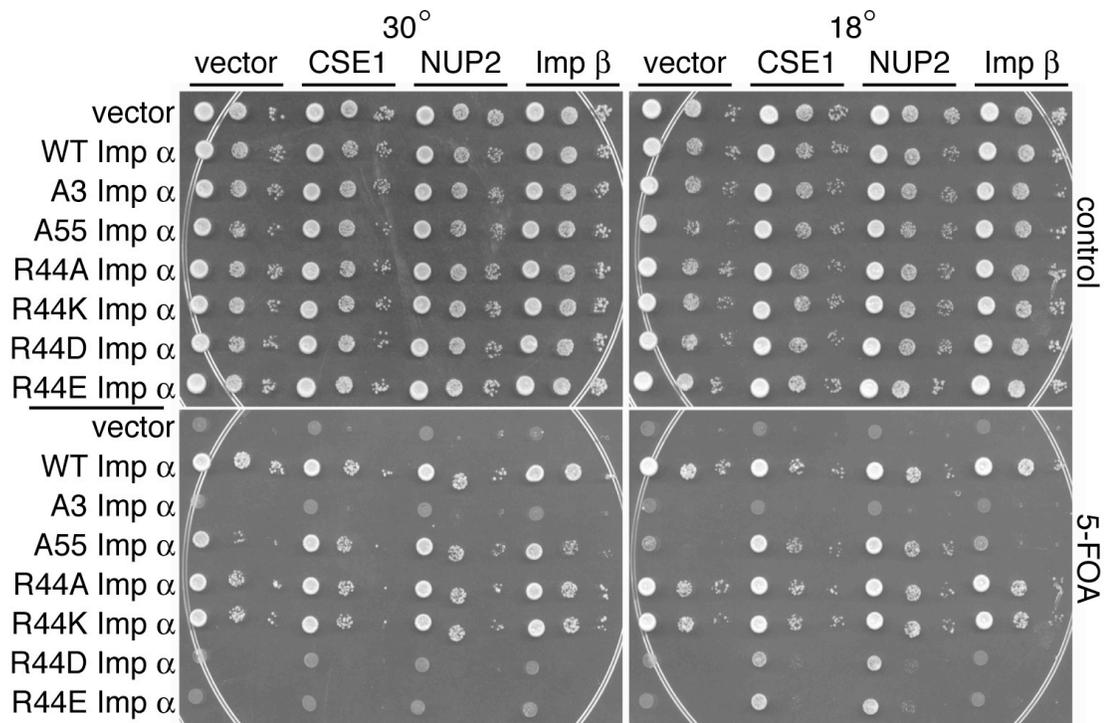
FIGURE 4.10 **Reincarnation.** *ΔSRP1* cells (ACY685) maintained by a plasmid encoding wild-type importin α, expressing either wild-type or variant importin α, and overexpressing either Cse1, Nup2, or importin β were serially diluted, spotted onto control or 5-FOA plates [119], and grown at 30°C or 18°C.

**Discussion**

During classical nuclear protein import, the importin α import receptor must bind to its cNLS-cargo with high affinity in the cytoplasm, cooperate with importin β to escort the cargo into the nucleus, then release the protein in the nuclear interior. The question of how importin α so precisely modulates its affinity for protein cargo is a key one in the nuclear transport field. Previous studies showed how RanGTP releases importin β from the import complex, freeing an auto-inhibitory domain on importin α to compete for binding to the cNLS-binding pockets. They also implicated Nup2 and Cse1 in the delivery of cNLS-cargo within the nucleus, but the precise mechanism of their involvement was unknown.

The crystal structures of Cse1/RanGTP/importin α and Nup2/importin α provided detailed models of how these karyopherin release factors might contribute to cNLS-cargo release [42-44]. Using the predictions provided by these structures, we aimed to test the validity of these models *in vivo*. In the trimeric export complex, Cse1 was predicted to form key interactions with both the C-terminus and the IBB domain of importin α. Previous studies had implicated the C-terminus of importin α in this interaction, but the IBB domain had heretofore only been conjured in hypothetical models. In this study, we hypothesized that the R44 residue of importin α and the D220 residue of Cse1 formed an interaction that was essential to the interaction between the two proteins. We found that reversing the charge of either of the residues caused importin α or Cse1 to be non-functional. The R44 importin α reversal-of-charge variants were also unable to interact with Cse1 *in vivo*. Similarly, substitution of the D220 residue of Cse1 to asparagine or

lysine rendered the protein incapable of interacting with importin α; however, D220A and D220N Cse1 were also unable to bind importin α *in vivo*. This result suggests that the D220 residue of Cse1 may be essential for maintenance of the tertiary structure. Interestingly, D220N is the point mutation contained in the *cse1-1* mutant. This variant of Cse1 was one of the first to be characterized and was found in a screen for chromosome segregation mutants [129], which gave the export factor its name. The cause of the *cse1-1* phenotype had previously not been explored in detail, but our studies suggest that a defect in cNLS-cargo delivery may contribute to the chromosome segregation errors, perhaps by affecting import of a crucial component for the segregation process.

The structural model of the N-terminus of Nup2 bound to ΔIBB importin α revealed that Nup2 interacts with two areas of importin α: the minor cNLS-binding pocket and the C-terminus. We found that amino acid substitutions within Nup2 predicted to affect binding to either or both of these regions caused varied growth defects and disrupted the interaction with importin α *in vivo*. These findings suggest that basic residues in the N-terminus of Nup2 contribute to the function of the protein in cells, suggesting that the model predicted by the structure is representative of how Nup2 functions *in vivo*. Before now, two main models had been proposed to explain the role of Nup2 in cNLS-cargo release. First, Nup2 could serve as a scaffold, facilitating cNLS-cargo release and importin α recycling by enhancing the local concentrations of importin α, RanGTP, and other dissociation factors [42]. The dynamic association of Nup2 with Nup60 on the nuclear face of the pore supports this model because Nup2 could serve as a docking site for incoming import complexes, catalyzing cargo release through proximity

without actively participating in dissociation [81]. Second, Nup2 could act analogously to the auto-inhibitory domain, electrostatically binding to the C-terminus of importin α, folding over the groove of the minor cNLS-binding pocket and extending towards the major cNLS-binding pocket, thereby competing for binding with both mono- and bipartite cNLSs [42, 43]. Nup2 itself could then be displaced from the C-terminus of importin α by Cse1 [56]. The structures and *in vivo* confirmations presented in this study support this second model, but do not preclude the first. For example, while the Nup2 N-terminus competes for binding with the cNLS, the Nup2 Ran-binding domain could recruit RanGTP to facilitate the dissociation of importin β and the freeing of the auto-inhibitory domain, and the Nup2 FXFG repeats could recruit Cse1 to prepare for its role in dissociation and export. Future studies will focus on establishing whether cNLS-cargo delivery in the nucleus occurs in a sequential or a concerted manner, advancing in a step-wise progression or proceeding with collaborative, cooperative style.

**CHAPTER 5**

**A Structural and Functional Analysis of the Classical Import Receptor, Importin β: Getting to the Heart of Nuclear Transport**

Allison Lange[1], Callie Preast[1], Anita H. Corbett[1], Murray Stewart[2], Bostjan Kobe[3], and Jade Forwood[4]

[1]Department of Biochemistry, Emory University, School of Medicine, Atlanta, GA 30322
[2]MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK
[3]Institute for Molecular Bioscience and Special Research Centre for Functional and Applied Genomics, The University of Queensland, Brisbane, Queensland, Australia
[4]School of Biomedical Sciences, Charles Sturt University, Wagga Wagga, Australia

**Introduction**

The transport of macromolecules from the cytoplasm into the nucleus is an essential eukaryotic process that allows transcription factors, DNA and RNA polymerases, histones, and other proteins to gain specific access to the nucleus [1-3]. Importantly, regulation of nucleocytoplasmic transport provides the cell with an additional level of control in regulating processes such as cell signaling and cell division, which are conserved in higher eukaryotes. Aberrations in the transport process are correlated with the formation of cancer [130-132], Alzheimer disease [133], mental illness [134], and other diseases such as genetic sex reversal [135].

The active transport of proteins across the nuclear envelope is dependent on a superfamily of carrier proteins known as importins or karyopherins [17, 18]. Importins recognize a diverse range of proteins through their interaction with nuclear localization signals (NLSs), which are contained within proteins destined for the nucleus (reviewed in [53]). The binding of an importin to its NLS-cargo is the first step of the nuclear import cycle, followed by docking and translocation through the nuclear pore complex. Once in the nucleus, the importin/cargo complex is dissociated through a high affinity interaction with RanGTP and the action of karyopherin release factors, which delivers the NLS-cargo protein and allows recycling of the importin back to the nucleus.

Importin β plays a central role in the classical nucleocytoplasmic transport process. It is constructed from 19 tandem HEAT repeats [79, 136], which provide ample interface for binding a range of interactors such as importin α, Ran, nucleoporins, and cargo proteins. Details of these interactions have been elucidated in structural studies,

predominantly by x-ray crystallography [79-83]. These studies reveal that importin β uses a variety of interface techniques that involve the inner and outer faces of the protein, varied degrees of extension, and even individual HEAT repeats. An emerging theme is that importin β can act as a spring [84-86], with its flexibility playing an important role in mediating binding with such varied partners.

Despite the range of importin β/protein complex structures that have been solved over the last decade, a high-resolution structure of unbound importin β has yet to be determined. Here, we describe the structure of the importin β nuclear import receptor in its unbound state. We show that the molecule is tightly coiled, with a number of intra-molecular interactions contributing to the circular structure. Mutations predicted to disrupt the closed structure of the protein resulted in aberrant localization of importin β-GFP demonstrating that altering these residues affects importin β *in vivo*; however, neither amino acid substitutions within residues predicted to enhance or disrupt the closed state of importin β affected the function of the protein *in vivo* or the localization of a model cNLS-cargo. Together, these results suggest that this closed importin β structural model probably is representative of the *in vivo* unbound state and that the flexibility of importin β probably affords it great tolerance of mutation.

**Experimental Procedures**

*Strains, Plasmids, and Chemicals-* All chemicals were obtained from US Biological or Sigma unless otherwise noted. All media was prepared and all DNA manipulations were performed according to standard procedures [99, 100]. All yeast

strains and plasmids used in this study are described in Table 5.1. *RSL1* is the gene that encodes importin β in *S. cerevisiae*.

*Preparation of Importin β-* Full-length GST-importin β was expressed in *E. coli* strain BL21(DE3)pLysS and purified by glutathione affinity chromatography. The GST fusion-tag was cleaved using thrombin protease and removed by size exclusion chromatography. Fractions containing the purified recombinant importin β were passed through glutathione sepharose affinity matrix to remove residual traces of GST and uncleaved GST-importin β. The protein was concentrated to 40 mg/ml and stored in 20 mM Tris (pH 7.4), 50 mM NaCl at $-80^{\circ}$C.

*Crystallization, Data Collection, and Analysis-* Crystals were grown by the hanging drop vapor diffusion method using 2 μl of protein solution with an equal volume of reservoir solution containing 20% PEG 4K, 0.1 M MES (pH 6.5), 20 mM $MgCl_2$, 125 mM NaCl, and 12% MPD. After two days at 17°C, large rod shaped crystals were transiently soaked in reservoir solution and flash-cooled under a 100 K nitrogen stream (Cryocool, Cryo Industries). Diffraction data were collected from a single crystal on a RaxisIV++ image plate detector using CuKα radiation produced from a Rigaku FR-E rotating anode generator (Rigaku). Raw data were auto-indexed, integrated and scaled using the HKL2000 package. Crystals displayed $P2_1$ symmetry, with unit cell dimensions of $a = 58.17$ Å, $b = 127.25$ Å, $c = 68.52$ Å; B = 102.23. The asymmetric unit contained one importin β molecule. Initial phases were obtained by molecular replacement using Phaser and the Kap95p:Nup1p structure as a search model. Local rebuilding using COOT and refinement with REFMAC from the CCP4 program suite yielded a final model with good overall stereochemistry.

TABLE 5.1 **Strains and plasmids used in Chapter 5**

| Strain/Plasmid | Description | Origin |
|---|---|---|
| FY23 (ACY192) | Wild type, MATa *ura3-52 leu2Δ1 trp1* | [106] |
| PSY884 (ACY209) | ΔRSL1::HIS3 [RSL1 *CEN URA3 AMP$^R$*], MATa *ade2 leu2 trp1 ura3* | [137] |
| ACY639 | srp1-31, MATα *trp- ura- leu- lys- his-* | [138] |
| pRS315 (pAC3) | *CEN LEU2 AMP$^R$* | [117] |
| pAC62 | RSL1, *CEN LEU2 AMP$^R$* | This Study |
| pAC511 | RSL1-GFP, *CEN TRP1 AMP$^R$* | This Study |
| pAC1065 | SV40NLS-GFP-GFP, *CEN URA3 AMP$^R$* | [55] |
| pAC1069 | GFP-GFP, *CEN URA3 AMP$^R$* | [55] |
| pAC2460 | S74K RSL1-GFP, *CEN TRP1 AMP$^R$* | This Study |
| pAC2461 | D167K RSL1-GFP, *CEN TRP1 AMP$^R$* | This Study |
| pAC2462 | E341A RSL1-GFP, *CEN TRP1 AMP$^R$* | This Study |
| pAC2463 | K468D RSL1-GFP, *CEN TRP1 AMP$^R$* | This Study |
| pAC2464 | F514K RSL1-GFP, *CEN TRP1 AMP$^R$* | This Study |
| pAC2465 | E737K RSL1-GFP, *CEN TRP1 AMP$^R$* | This Study |
| pAC2466 | S74K RSL1, *CEN LEU2 AMP$^R$* | This Study |
| pAC2467 | D167K RSL1, *CEN LEU2 AMP$^R$* | This Study |
| pAC2468 | E341A RSL1, *CEN LEU2 AMP$^R$* | This Study |
| pAC2469 | K468D RSL1, *CEN LEU2 AMP$^R$* | This Study |
| pAC2470 | F514K RSL1, *CEN LEU2 AMP$^R$* | This Study |
| pAC2471 | E737K RSL1, *CEN LEU2 AMP$^R$* | This Study |
| pAC2536 | S74K/F514K RSL1-GFP, *CEN TRP1 AMP$^R$* | This Study |
| pAC2537 | S74K/F514K RSL1, *CEN LEU2 AMP$^R$* | This Study |

*Microscopy*- GFP-fusion proteins were localized in live *S. cerevisiae* cells using direct fluorescence microscopy on an Olympus BX60 epifluorescence microscope equipped with a Photometrics Quantix digital camera. To localize the importin β variants, wild-type cells (ACY192) expressing wild-type or variant importin β-GFP under its own promoter were grown overnight in selective media, diluted in fresh media, and grown for 3 hours to log phase prior to localization. To visualize the effect of the importin β mutants on the localization of a classical nuclear import cargo versus GFP-GFP, *ΔRSL1* cells containing a wild-type *RSL1 URA3* plasmid (ACY208) were transformed with plasmids encoding each importin β variant. The resulting transformants were plated on plates containing 5-fluoroorotic acid (5-FOA) [119] to remove the maintenance plasmid and were transformed with a plasmid encoding either SV40NLS-GFP-GFP (pAC1065) or a GFP-GFP control (pAC1069) under the control of the *MET25* promoter. Cells were grown overnight in selective media, washed once in dH$_2$O, resuspended in media lacking methionine to induce expression of the reporter proteins, and incubated overnight prior to localization studies.

*Immunoblot Analysis*- Immunoblot analysis was performed essentially as previously described [139]. Wild-type cells (ACY192) transformed with plasmids encoding the importin β-GFP variants were grown to log phase, collected by centrifugation, and washed once with dH$_2$O and twice with cold PBSMT (PBS, 5mM MgCl$_2$, 0.5% Triton X-100). Glass bead lysis was conducted in PBSMT in the presence of the protease inhibitors phenylmethylsulfonyl fluoride (0.5mM PMSF) and PLAC (3μg/ml each of Pepstatin A, Leupeptin, Aprotinin, Chymostatin). Lysates were cleared

by centrifugation and total protein concentration was assessed by Bradford assay.  Thirty microgram of protein was resolved by SDS-PAGE, transferred to a nitrocellulose membrane, and probed with an anti-GFP antibody (1:3,000, rabbit) to detect the importin β-GFP proteins and an anti-Yrb1 antibody (1:50,000, rabbit, [125]) as a loading control.

In Vivo *Functional Analyses*- The *in vivo* function of each of the importin β mutants was assessed using a plasmid shuffle technique [119].  For the plasmid shuffle, vector alone or plasmids encoding the wild-type or variant importin β proteins were transformed into *ΔRSL1* cells containing a wild-type *RSL1 URA3* maintenance plasmid (ACY208).  Single transformants were streaked onto control ura- leu- glu plates or on selective leu- glu plates containing 5-FOA.  Plates were incubated at 18°C, 25°C, 30°C, or 37°C for 3-5 days.

*NLS Import Assay*- The NLS import assay was performed essentially as previously described [55, 140].  Cells containing the importin β variants as the only copy of importin β and expressing the SV40NLS-GFP-GFP reporter protein were grown and induced.  After collection by centrifugation, cells were pelleted, resuspended in glucose-free media containing sodium azide and 2-deoxy-D-glucose, and incubated at 30°C for 45 minutes.  The cells were washed with ice-cold dH$_2$O, resuspended in glucose-containing media pre-warmed to 30°C, and incubated at 30°C.  Following resuspension, samples were removed every 2.5 minutes and imaged using direct fluorescence microscopy.  At least 100 cells were analyzed for each time point.  Cells were scored as "nuclear" if the nucleus was discernable, i.e. if the nucleus was brighter than the cytoplasm and if a nuclear-cytoplasmic boundary was visible.

**Results**

*Structural determination of free importin β*- Structural analysis of the nuclear

import receptor, importin β, in its unbound state was undertaken through crystallization

of recombinantly expressed protein purified by affinity and size exclusion

chromatography. Protein crystals diffracted to 2.5 Å resolution and the structure was

solved by molecular replacement using the structure of importin β bound to the C-

terminal region of Nup1 [81]. The final model consisted of all 861 residues of importin β

and had good overall stereochemistry. Recombinantly expressed importin β has

previously been shown to function in *in vitro* studies [60].

The x-ray crystallographic structure of importin β revealed that all 19 HEAT

repeats are arranged in a tightly coiled, compact structure (Figure 5.1A). HEAT repeats

1-17 form a ring structure, with the C-terminal HEAT repeats folding back to make

contacts with the N-terminus of the molecule. Overall, the compacted conformation

appears to be mediated through interactions involving the two termini of the protein

(between N-terminal HEAT repeats, H2 and H4, and the distal part of the molecule, H17)

and through interactions in a hinge-region of the protein (between HEAT repeat H8 and

HEAT repeats H11 and H12).

In vivo *importin β variants*- Analyzing the crystal structure of unbound importin

β allows us to make specific predictions about what residues may be critical for

formation or maintenance of the closed structure *in vivo*. As noted above, the circular

arrangement of free importin β seems to be achieved through two main interaction

A
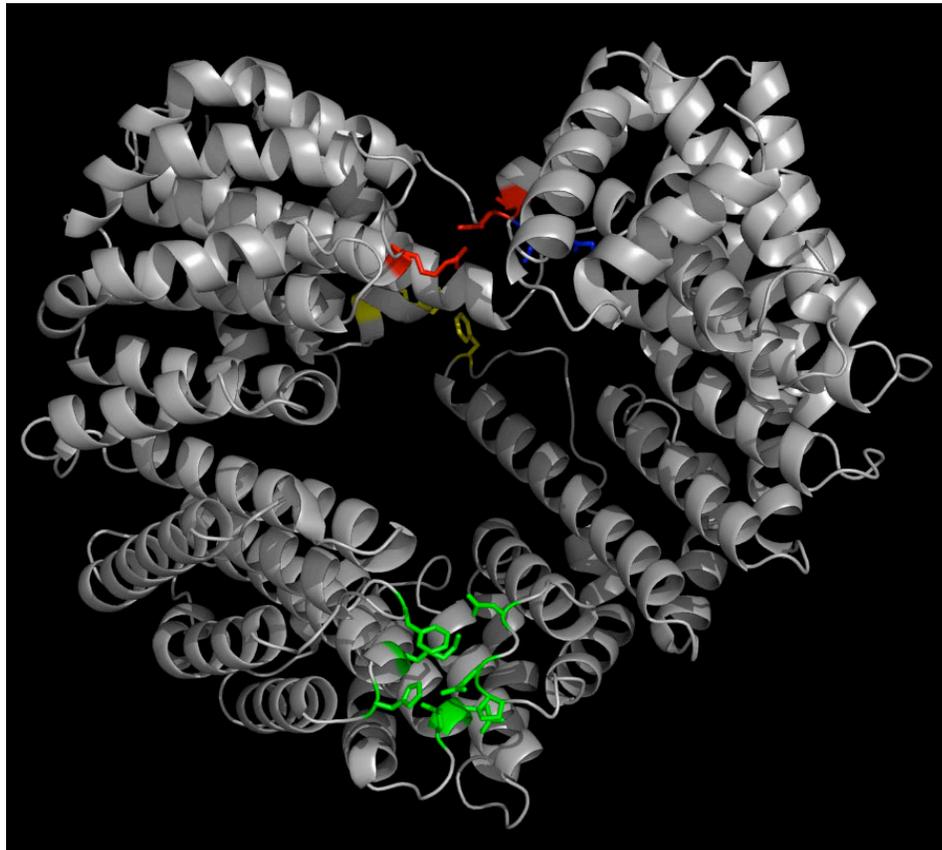


B
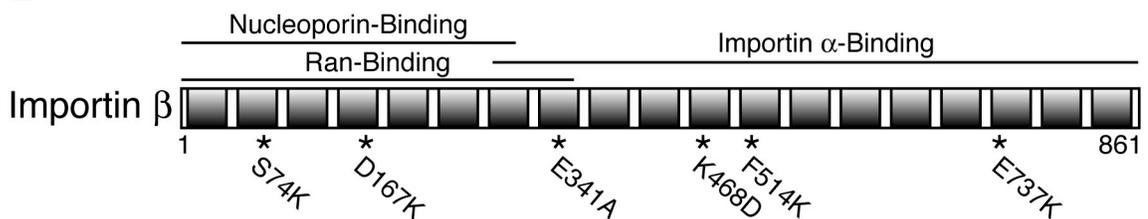


FIGURE 5.1  **Crystal structure and domain schematic of importin β.**  *A,* full-length importin β is depicted in the grey ribbon diagram.  Residues predicted to contribute to structural integrity are colored:  Red = Ser74/Glu737; Blue = Asn738/Asp167; Yellow = Tyr 793/Phe307; Green = Phe514/Glu341, Leu425/Pro3401, and Asp343/Lys468. *B,* schematic of importin β depicting known interaction domains.  HEAT repeats are denoted with shaded boxes.  Amino acid substitutions analyzed in this study are indicated and marked with asterisks.

interfaces: one involving the termini of the protein and one involving a hinge region.

Accordingly, we created amino acid substitutions in both of these regions designed to

either disrupt or enhance maintenance of the unbound structure of importin β (Table 5.2).

S74K, E737K, and D167K are located at the interface between the termini and F514K,

E341A, and K468D are located in the hinge region. The location of these residues on the

primary structure of importin β is depicted in Figure 5.1B [79, 141-143]. S74K and

F514K are predicted to enhance self-binding of importin β by creating new charge-based

interactions and E737K, D167K, E341A, and K468D are predicted to disrupt binding.

Several of the residues involved in these interactions are conserved in human importin β;

D167 is strictly conserved and E737, E341 and K468 conserve the charge of the residue.

   *Localization of importin β variants*- In *S. cerevisiae*, importin β-GFP is localized

at the nuclear rim with some nuclear and cytoplasmic signal, reflecting its constant

shuttling from one compartment to the other. It is possible that mutation of these residues

hypothesized to be involved in maintenance of the closed, unbound structure might affect

the affinity of importin β for importin α, RanGTP, or the nuclear pore, which would alter

its ability to efficiently shuttle and, therefore, change its steady-state localization. To

determine the localization of the importin β variants, wild-type and mutant importin β

were tagged with GFP and expressed in wild-type cells (Figure 5.2A). The majority of

the importin β-GFP mutants were localized to the nuclear rim in a manner similar to

wild-type importin β-GFP. In contrast, K468D importin β-GFP was localized throughout

the cell with no discernable nuclear rim signal. To verify that these results were not due

to variations in expression level, an immunoblot was performed (Figure 5.2B). Results

TABLE 5.2 **Importin β Mutations**

| Mutation Predicted to Enhance Self-Binding | Interaction | | Mutation Predicted to Disrupt Self-Binding |
|---|---|---|---|
| S74K ·············· | Ser 74 ——————— Glu 737 ·············· | | E737K |
| | Arg 696 ———— | | |
| | Asn 738 ——————— Asp 167 ·············· | | D167K |
| F514K ·············· | Phe 514 ——————— Glu 341 ·············· | | E341A |
| | His 466 ———— | | |
| | Asp 343 ——————— Lys 468 ·············· | | K468D |

The font color of the depicted interaction corresponds with the color of the involved residues on the importin β structure presented in Figure 5.1A.
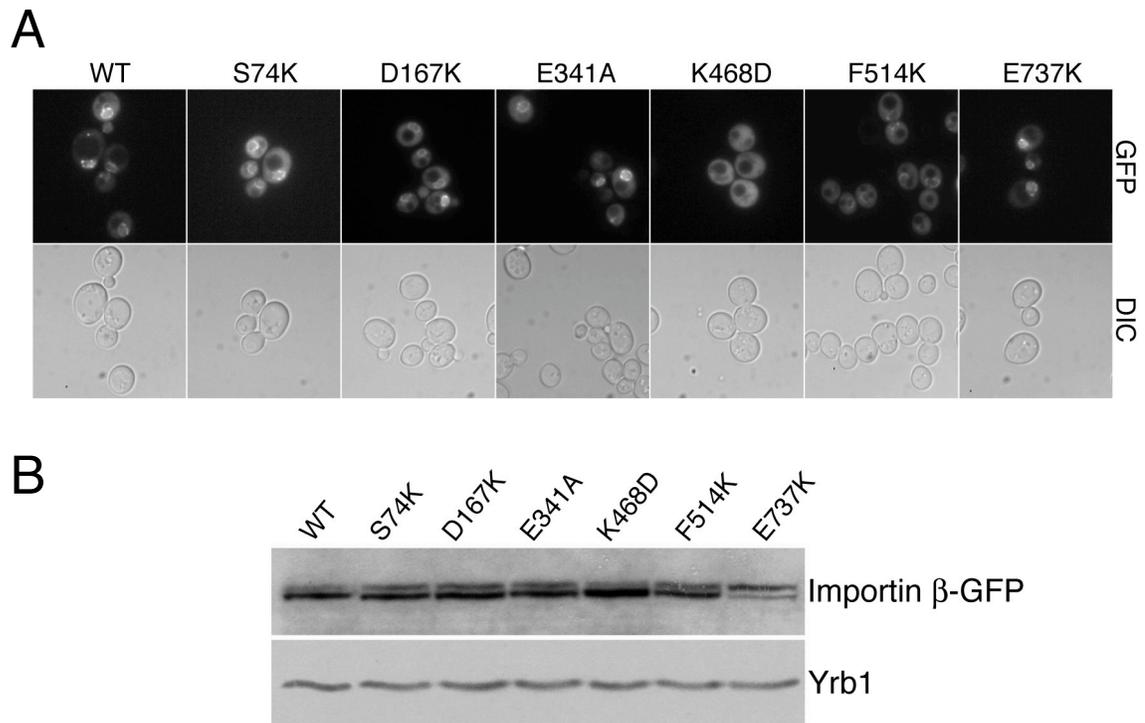
FIGURE 5.2 **Localization and expression of the importin β-GFP variants.** *A*, wild-type cells (ACY192) expressing the each of importin β-GFP variants were grown to log-phase and examined by direct fluorescence and differential interference contrast (DIC) microscopy. *B*, the level of each of the importin β-GFP variants expressed in wild-type cells (ACY192) was detected by immunoblotting with an anti-GFP antibody. Yrb1 levels were monitored with an anti-Yrb1 antibody [125] to demonstrate that an equal amount of total protein was loaded in each lane.

indicated that each of the importin β-GFP variants is expressed at approximately equal levels.

*Assessing the* in vivo *function of the importin β variants*- A plasmid shuffle assay was utilized to determine if the amino acid substitutions in importin β affect the essential function of importin β *in vivo*. *ΔRSL1* cells, which lack the gene for importin β, *RSL1*, but contain a wild-type importin β maintenance plasmid, were transformed with vector alone or a plasmid encoding either wild-type or variant importin β. Cells were then plated on control plates or on plates containing the drug 5-FOA [119], which removes the maintenance plasmid and leaves the importin β variant as the only cellular copy of importin β (Figure 5.3). All of the cells expressing variants of importin β grew like cells expressing wild-type importin β, indicating that these amino acid substitutions in importin β do not disrupt cellular viability. Serial dilution assay and growth curve analysis also failed to show a difference in growth between cells expressing the importin β variants (data not shown).

*Effect of amino acid substitutions in importin β on the nuclear import of a model cargo protein*- To assay the effect of amino acid substitutions within importin β on the import of a model classical NLS-containing cargo, SV40NLS-GFP-GFP or GFP-GFP alone was localized in cells containing the importin β variants as the only copy of importin β (Figure 5.4). The double GFP tag was used to minimize passive diffusion into the nucleus. In each of the strains, GFP-GFP alone localized throughout the cell and SV40NLS-GFP-GFP showed significant steady-state nuclear localization, indicating that these amino acid substitutions within importin β do not significantly impair its role in
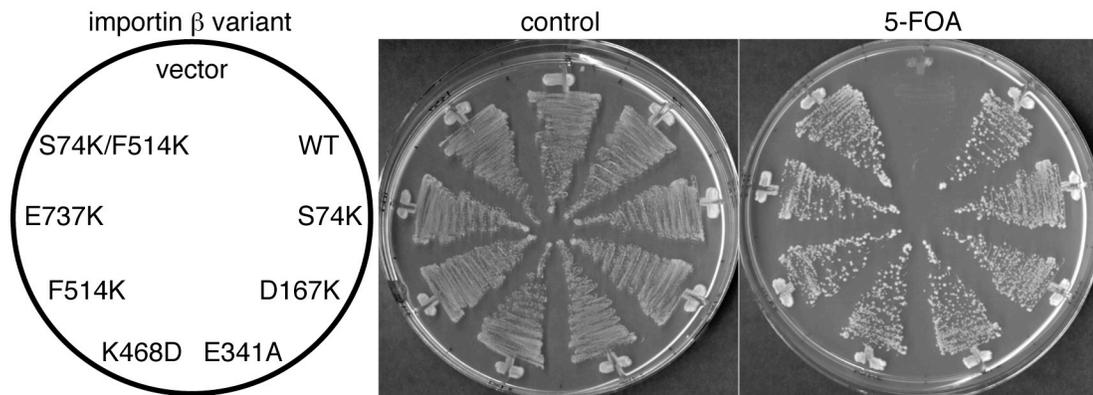
FIGURE 5.3 **Functional analysis of the importin β mutants *in vivo*.** *A*, the *RSL1* deletion strains maintained by a plasmid encoding wild-type importin β (ACY208) and expressing either wild-type or mutant importin β protein were streaked onto control or 5-FOA plates and grown at 30°C for 3 days.

ΔRSL cells expressing importin β variants
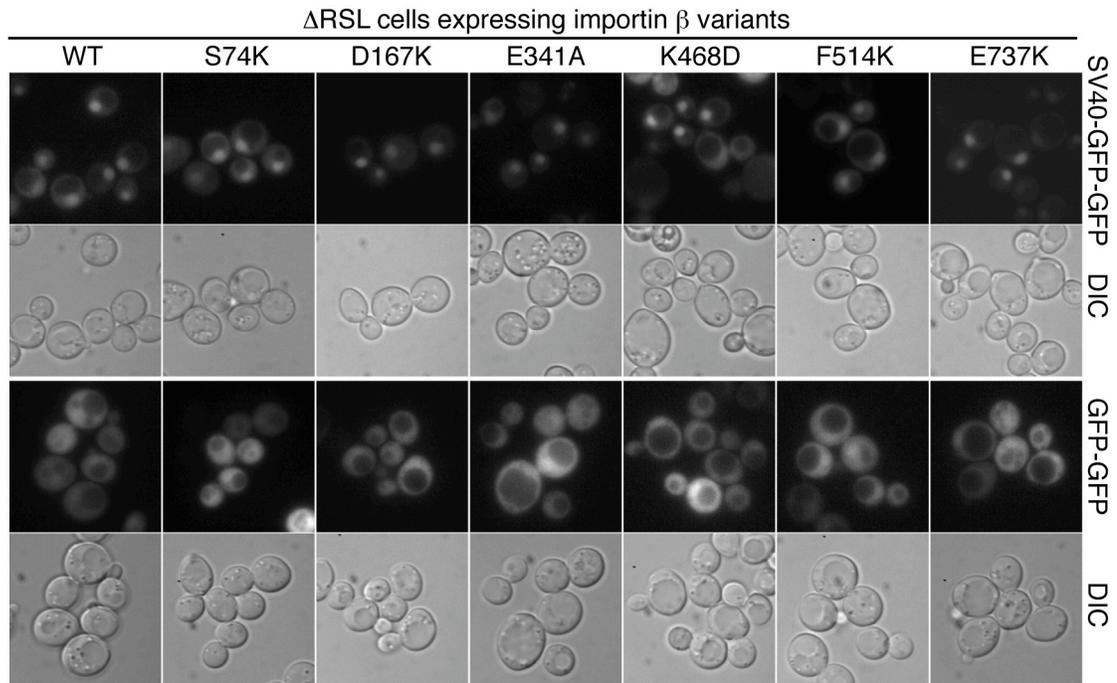
FIGURE 5.4  **Effect of importin β variants on NLS-cargo import.** *ΔRSL1* cells (ACY208 after plating on 5-FOA [119]) expressing the importin β variants were transformed with a plasmid encoding either SV40NLS-GFP-GFP (pAC1065) or GFP-GFP (pAC1069) under the control of the *MET25* promoter.  Cells were grown overnight in selective media, induced in media lacking methionine, and incubated overnight prior to localization by direct fluorescence and differential interference contrast (DIC) microscopy.

classical nuclear import. These results also suggest that the interaction between importin α and these importin β variants is not significantly altered since the heterodimeric import receptor must form in order for classical import to occur.

To determine if the initial rate of NLS-cargo import was altered despite the lack of effect on the final steady-state localization of the model classical cargo, a kinetic import assay was utilized. Cells containing the importin β variants as the only copy of importin β and expressing either GFP-GFP alone or SV40NLS-GFP-GFP were incubated with azide and 2-deoxy-glucose, which deplete the cell of energy and cause redistribution of any nuclear cargo across the cell [140]. After washing out the inhibitors, the import kinetics were measured by assessing the percentage of cells with nuclear accumulation of the reporter over time (Figure 5.5). The rate of initial import of the NLS-reporter was approximately the same in all of the cells regardless of the importin β variant expressed.

**Discussion**

In this study, we have solved the crystal structure of the classical import receptor, importin β, in its unbound state. We find that it exists in a closed, circular conformation, which is similar to the unbound conformation of the classical export receptor, Cse1 [87] (Figure 5.6). Both structures reveal a compact ring of HEAT repeats stabilized by intramolecular interactions, though importin β is more compact than Cse1, especially at the C-terminus. The interactions that mediate the ring formation are not conserved between importin β and Cse1. In Cse1, interactions between HEAT repeats 1-3 and HEAT repeats 14-16 appear to be most important, but in importin β, the interactions at

116
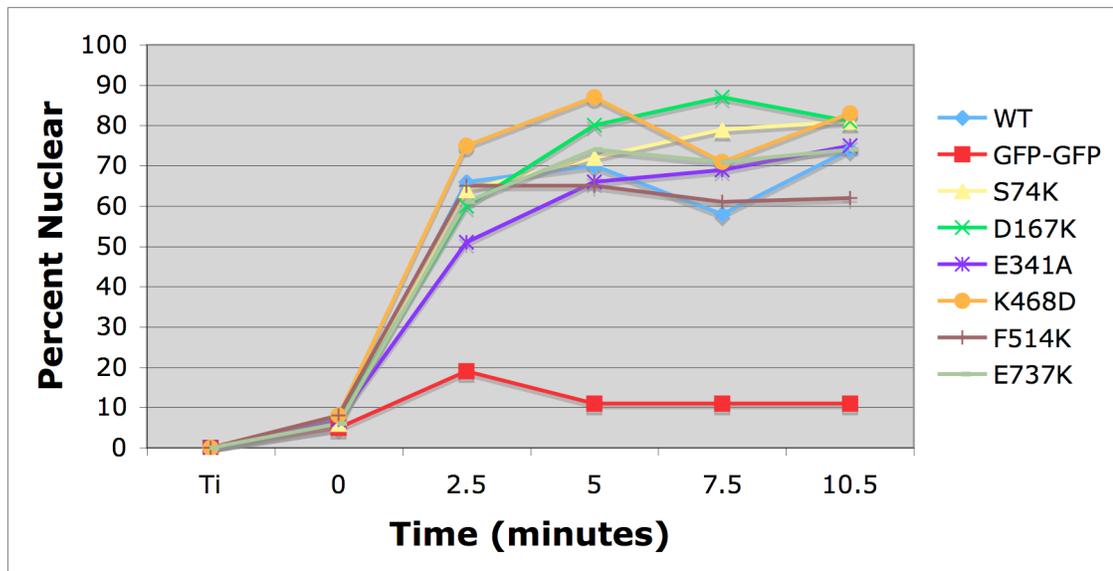
FIGURE 5.5 **Effect of importin β variants on import kinetics of an NLS-cargo.** The import kinetics of an NLS-cargo in cells expressing the importin β variant indicated as the only copy of importin β in the cell were measured by assessing the percentage of cells with nuclear accumulation of SV40NLS-GFP-GFP or GFP-GFP alone (control) every 2.5 minutes over a 10 minute time course [140]. The percentage of cells with nuclear accumulation is plotted against time with $T_i$ denoting the initial time point taken before the cells were resuspended in media.
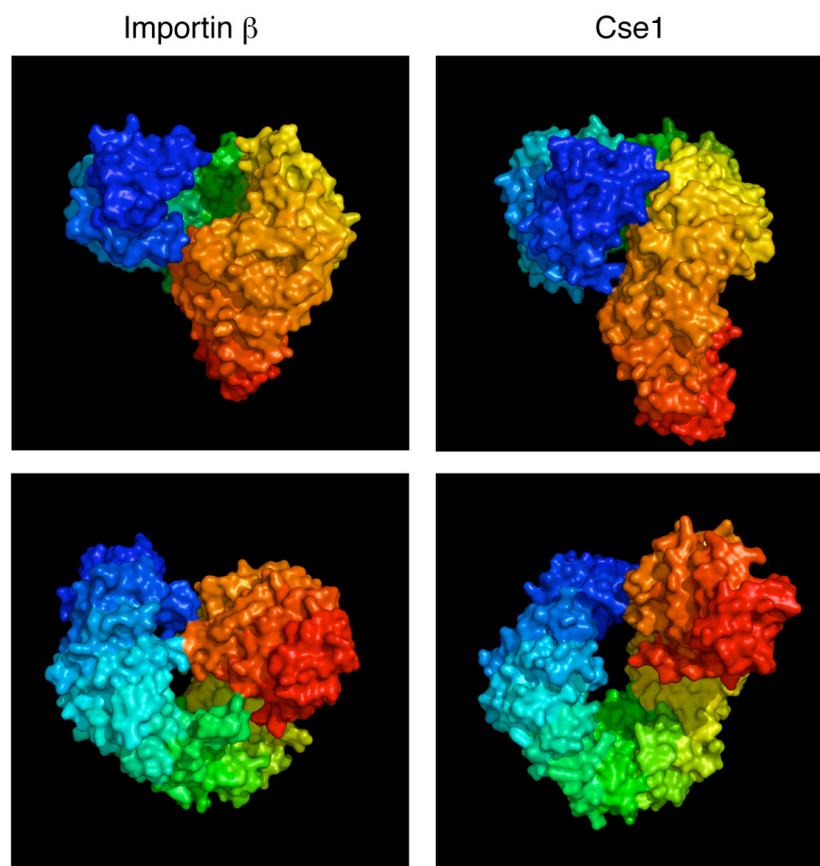
FIGURE 5.6 **Structural comparison of unbound importin β to unbound Cse1.** Both importin β and Cse1 [87] exist in a closed, ring-like conformation when in a cargo free state.

the termini between H2/H4 and H17 and at the hinge region between H8 and H11/H12 combine to bring the tail of the molecule in close proximity to the head. It is also unlikely that the physiological significance of these compacted structures is the same. For Cse1, importin α binding is dependent on RanGTP binding to open Cse1 into a conformation capable of binding importin α. In contrast, importin α binds to importin β in the cytoplasm in the absence of RanGTP. Nevertheless, it is interesting that these apparently inert states of the free receptors exist.

The structural determination of importin β in its unbound state also enabled a direct comparison to be made between the free and bound forms of the nuclear import receptor (Figure 5.7). Very large conformational changes were observed between importin β in its unbound state and importin β bound to RanGTP [37]. In the transition between unbound and RanGTP-bound importin β, the intramolecular interactions involved in holding free importin β in a compacted, coiled state are replaced with interactions binding to switch loops of RanGTP. The HEAT repeats of importin β are also uncoiled to form an S-like conformation, similar to that also observed when importin β is bound to the cargo protein SREBP-2 [83]. In contrast, importin β bound to Nup1 [81] assumes a conformation that is very similar to unbound importin β, remaining tightly coiled. A compact configuration is also observed when importin β binds to the IBB domain of the nuclear import receptor importin α [79].

Importantly, these varied conformations confirm recent models of nuclear import, which invoke the inherent flexibility of karyopherin molecules [144]. In these models, large conformational changes are achieved through the cumulative effect of extremely small relative changes within α-helices of the HEAT repeats coupled with larger
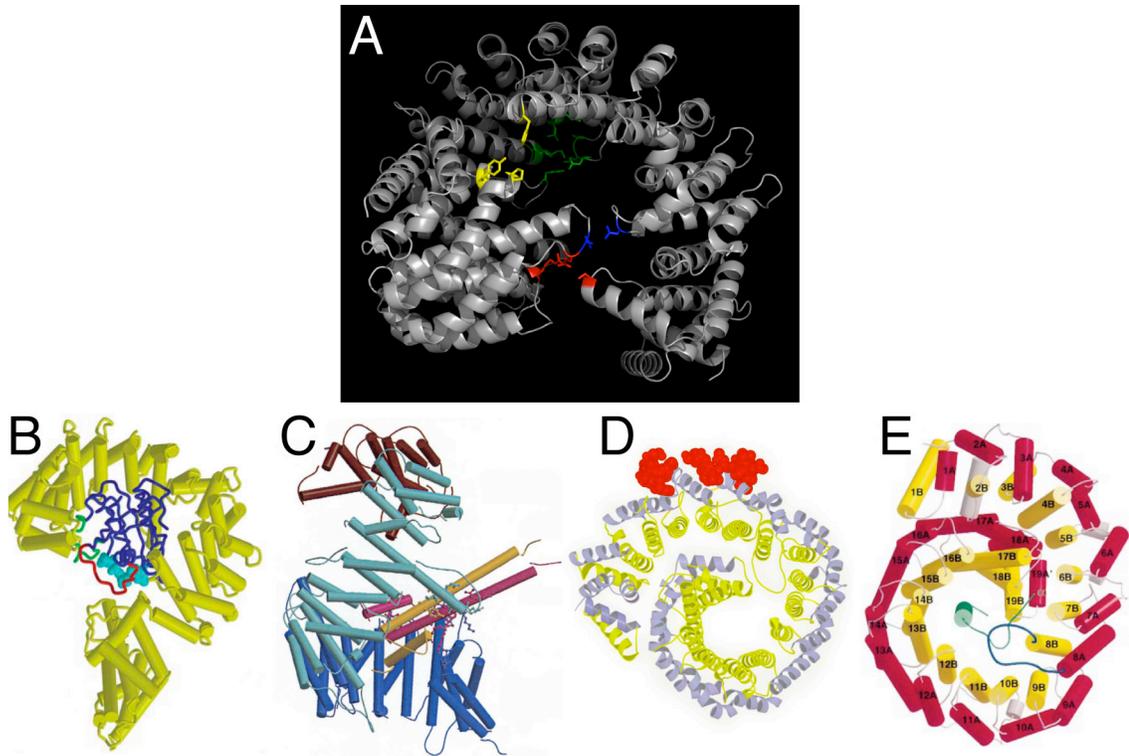
FIGURE 5.7 **Structural comparison of unbound to bound importin β.** *A,* structure of unbound importin β (grey). *B,* structure of importin β (yellow) bound to RanGTP (blue) [37]. *C,* structure of importin β (brown, teal, and blue) bound to a dimer of SREBP-2 (yellow and pink) [83]. *D,* structure of importin β (yellow and blue) bound to Nup1 (red) [81]. *E,* structure of importin β (yellow and red) bound to the IBB domain of importin α (green) [145].

movements of sections of the protein involving many HEAT repeats. This flexibility contributes to the ability of β-karyopherin proteins to coil around their interaction partners, gradually wrapping around their cargo using an extensive interaction interface. This mechanism implies that there may not be one or two key residues required for the maintenance of an interaction involving a β-karyopherin molecule; rather, a series of minor interactions may all slightly contribute to creating a functional bond.

Our *in vivo* data supports this model. While the K468D importin β-GFP variant was mislocalized from the nuclear rim, suggesting that it is possible to disrupt importin β localization with a single amino acid substitution, none of the other variants of importin β had any discernable effect on the localization or function of importin β or the localization of a model classical cargo *in vivo*. It is possible that the residues predicted by the crystal structure to enhance or disrupt the unbound configuration of importin β are simply not those that are essential for maintenance of the unbound form. However, we instead propose that the lack of a phenotype suggests that the free structure of importin β, much like the structure of importin β in complex with a variety of its interacting partners, is dependent on a series of interactions, any one of which may be lost without significantly altering the function of the protein. This fluid arrangement befits a protein with such varied interaction partners and is presumably critical for the ability of β-karyopherin receptors to specifically bind and discharge such a wide variety of cargo molecules.

**CHAPTER 6**

**Conclusions and Discussion**

Nucleocytoplasmic transport is essential to the eukaryotic cell. Without it, eukaryotic cellular processes would cease as the genome languished, separated from the structural factors that allow its organization, the repair factors that maintain its integrity, and the transcription factors that make it relevant. Understanding the mechanisms that underlie nuclear transport clarifies not only the elemental question of how proteins navigate the compartments of the cell, but also an integral step in essential processes such as cell signaling and protein translation, which require the regulated import of transcription factors and the proper export of mRNA prior to their execution. Critically, mastery of the intricacies of import and export from the nucleus also informs questions relevant to the development of cancer, which can involve improper import of carcinogenic transcription factors, and viral diseases, which often integrate the cellular transport machinery into the maintenance of their own life-cycles. By gaining insight into each step in the development of a disease, we can begin to understand the underlying etiology, which allows us to dissect disease-specific processes from those that are essential to the life of the cell and specifically target those vulnerabilities for drug development.

The major steps involved in nuclear import are well-established: 1) recognition and binding of protein cargo in the cytoplasm, 2) translocation through the nuclear pore, 3) release of cargo within the nucleus, and 4) recycling of the receptors back to the cytoplasm. However, at every step of the process, major unknowns remain. Therefore, in this final chapter we will address several of the lingering questions about the nucleocytoplasmic transport cycle, the progress that this dissertation has made towards addressing many of these questions, and which challenges remain to be solved.

In the first step of nuclear import, import receptors discriminate their cargo from the other proteins in the cytoplasm. The mechanism for this recognition has proven difficult to decipher since each receptor seems to interface with its cargo in a unique manner, quashing any early dreams of a universal nuclear localization motif. Consequently, in order to make progress towards understanding fully how the nuclear cadre of proteins is determined, the nuclear localization signal(s) for each receptor must be defined independently. Therefore, in this dissertation, our approach was to define, refine, and analyze the NLS motifs recognized by two import receptors, the classical import receptor, importin α/β, and the transportin homolog, Kap104.

One of the NLS motifs that binds to importin α/β is the bipartite cNLS. We found that the canonical consensus sequence for the bipartite cNLS was unnecessarily restrictive and determined that the two basic motifs can be separated by as few as eight amino acids and as many as twenty amino acids while still preserving the ability of the cNLS to mediate nuclear import and binding to importin α. The expansion of the parameters for the linker in a bipartite cNLS broadened our definition of this nuclear localization signal and prompted us to find a novel group of putative cargos that may be imported by the classical import system. Our next task will focus on culling this large pool of putative cargos by verifying which of these potential cNLS-cargos actually mediate binding to importin α and are necessary and sufficient for importin α/β-dependent nuclear import. By analyzing common characteristics of the *bona fide* cargos, perhaps we can further refine the consensus sequence for classical NLSs, allowing us to predict more precisely than ever the contribution of the classical nuclear import system to nuclear transport as a whole.

In Chapter 3, we extended the work of Chook and colleagues and showed that the newly proposed transportin-binding sequence, the PY-NLS, is functional *in vivo* and is conserved across species. Given our results, which showed that the C-terminal PY-core of the Hrp1 PY-NLS mediates nuclear import to a degree, we hypothesized that the consensus sequence in yeast is less strict than that for humans and consists of the $R/H/KX_{2-5}PY$ motif with upstream hydrophobic or basic regions enhancing the NLS function. Recent work examining the binding energetics of the various portions of the PY-NLS [146] corroborate these results by showing that the N-terminal region of the Hrp1 PY-NLS contributes little to its binding to Kap104, while the C-terminal PY-core of the Hrp1 PY-NLS provides a Kap104-binding hotspot in *S. cerevisiae*; however, this study also found that Kap104 only recognizes the basic flavor of PY-NLS. This disparity between yeast Kap104- and human Kapβ2-binding is thought to be due to a lack of homology between the two receptors in the interfacial region that interacts with the upstream hydrophobic portion of PY-NLSs. Interestingly, a PY-NLS in Nab2 was identified in the RGG region previously found to be required for import of the protein [112, 113]. The upstream sequence [216]**KNRRGGRGGNRGGR**[229] paired with [235]**RFNPL**[239] was found to be required for nuclear import and capable of binding to Kap104 in a RanGTP-dependent manner. The function of a PY-NLS with a terminal residue other than tyrosine led to a study of the tolerable degeneracy in a PY-NLS sequence. In sequences such as the Hrp1 PY-NLS, where the C-terminal PY-core makes major contributions to the overall binding energy, the identity of the C-terminal residue is quite fixed, with only PY-NLSs terminating in tyrosine, phenylalanine, histidine, and methionine allowing measurable binding to Kap104. In contrast, in sequences such as

the Nab2 PY-NLS where the binding energy is distributed along the entire sequence, few residues are disallowed from the final PY-NLS position. Our current studies focus on evaluating the effects of varying the terminal residue of the PY-NLS *in vivo*. Much like our work with the classical NLS, we are also currently attempting to refine our searches for putative PY-NLS-containing cargos by integrating an analysis of the predicted disorder of a sequence into our algorithm for finding potential PY-NLS cargos. A major remaining challenge will be to identify the recognition mechanisms for the many other β-karyopherin receptors.

Identifying novel nuclear localization signals for these other receptors is a challenge made all the more daunting by the fact that the new NLS motifs may be altered or disguised by various forms of regulation. Several mechanisms for modulating the binding of an NLS to its receptor [147, 148] have previously been demonstrated, such as methylating [149] or ubiquitinating [150] residues of the NLS to enhance binding or masking the NLS through intermolecular interaction [151-153] to prevent binding, though the methods for regulating the interaction between an NLS-cargo and its receptor are not fully understood. Our work on the classical bipartite NLS and the PY-NLS suggests possibly mechanisms for controlling the affinity of these two NLSs for their receptors. The presence of a longer linker region in the bipartite cNLS may provide for additional levels of regulation in the import of bipartite cNLS-containing cargo by allowing for additional post-translational modification target sites near the NLS such as negatively charged phosphorylation [122, 154], which can increase or decrease the affinity of the NLS for the receptor. The longer linker region also increases the potential conformations that the spacer region can assume. Perhaps by folding over the nuclear

localization sequence, the longer linker itself could occlude the cNLS in certain conditions. For the PY-NLS, since variation in the terminal residue causes the PY-NLS to bind to importin α with varying affinities [146], these amino acid changes could allow for different rates of import of various Kap104 cargos [55]. Additionally, having varied residues in the PY-NLS provides a method for independent regulation of individual cargos by the specific addition of a post-translation signal onto only certain of the final amino acids.

Once the NLS-cargo has been recognized and bound, it must traverse the nuclear pore complex. Since the NPC is the only entry point into the intact nucleus, it is a likely location for additional levels of regulation, providing either a physical or an energetic barrier to nuclear entry [148]. Examining the mechanism of translocation across the nuclear membrane was beyond the scope this dissertation, but understanding how import and export complexes move through the NPC is a major source of current investigation.

After the import complex reaches the nucleus, the NLS-cargo must be released. For classical nuclear import, the specifics of how RanGTP, the IBB domain of importin α, Nup2, and Cse1 individually affect cargo delivery are now quite well understood with our *in vivo* studies, biochemical studies, and the crystal structures of RanGTP, importin α, Nup2, and Cse1 providing a detailed view of steps along the dissociation pathway. Furthermore, they hint at a framework on which to order these facts. For example, since Cse1 is the export receptor for importin α, it is probably the last dissociation factor to act, remaining bound after its has played its role in dissociation, and since the IBB domain of importin α forms intimate contacts with Cse1 while already bound to the body of importin α, the auto-inhibitory region probably acts before Cse1. We favor a system

where RanGTP dissociates importin β from the IBB domain of importin α and the auto-inhibitory region folds over and effects cargo release.  Nup2 accelerates this process by binding the importin α/cNLS complex on the C-terminus of importin α and folding over to aid in freeing the cargo.  Cse1/RanGTP removes the Nup2 and secures the auto-inhibitory region in a closed state (Figure 6.1).  A looming challenge is to understand whether these dissociation factors act in a concerted or step-wise manner to effect cNLS-cargo release or even whether multiple parallel pathways of cargo delivery simultaneously occur involving all some or all of these factors [43].

Once the import receptor has released its cargo, it returns, cargo-free, to the cytoplasm to prepare for another round of import.  In Chapter 5, we determined and analyzed the conformation of the classical import receptor, importin β, in its unbound state after it has returned to the cytoplasm and released Ran.  We found that it exists in a closed, ring-like conformation, much like the circular conformation seen in the unbound state of the Cse1 export receptor [87].  In the cargo-free state of Cse1, the binding sites for RanGTP and importin α are partially blocked.  It is thought that RanGTP and importin α must both be present to take advantage of "molecular breathing" movements of the export receptor, which allow the proteins to interact with the exposed portions of their binding sites [87].  These limited interactions disrupt the ring structure and push the equilibrium of the molecule into the open state.  The requirement for RanGTP prevents Cse1 from incorrectly and unproductively interacting with importin α while in the cytoplasm.  The regulation of importin β binding to cargo is less clear-cut, since importin β seems to interact with its cargos while in both an open conformation (as in the SREBP-2 structure [83]) and in a relatively compact conformation (as in the IBB structure [145]).
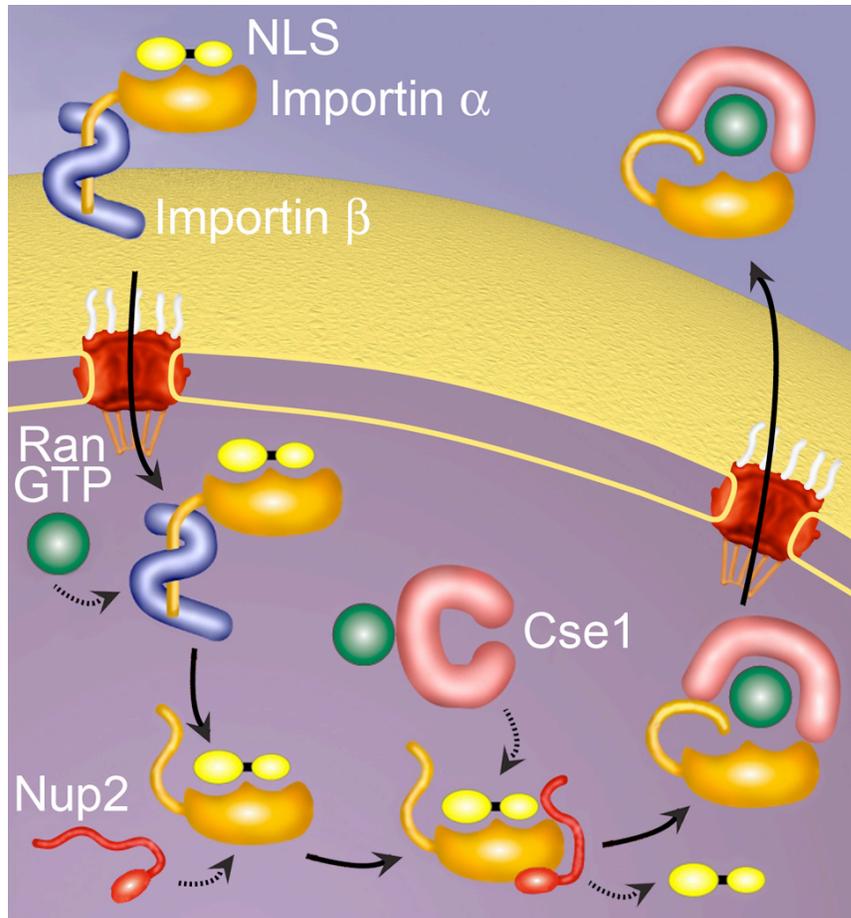
FIGURE 6.1 **Details of the classical nuclear import cycle.** In the cytoplasm, cargo containing a cNLS is bound by the heterodimeric import receptor, importin α/importin β. Importin α recognizes the cNLS and importin β mediates interactions with the nuclear pore during translocation. Once inside the nucleus, RanGTP-binding causes a conformational change in importin β, which releases the IBB region of importin α. This auto-inhibitory region folds over and contributes to cargo release. Nup2 accelerates this process by binding the importin α/cNLS complex on the C-terminus of importin α and folding over to aid in freeing the cargo. Cse1/RanGTP removes the Nup2 and secures the auto-inhibitory region in a closed state. Finally, importin α is recycled back to the cytoplasm by the export receptor, Cse1, in complex with RanGTP.

This variability in the conformation of importin β while interacting with cargo may explain why the variants of importin β failed to have an effect on the global function or localization of the protein. Our amino acid substitutions were designed to either disrupt or facilitate the closed state of the receptor. If importin β interacts with some of its cargos while open and some while closed, perhaps with each amino acid change we made, we hindered certain of the interactions that importin β must make and enhanced others, resulting in lack of an overall phenotype. To address this issue, future studies will assay the effect of the importin β variants on the interaction with specific cargos to try to clarify the role of the circular structure of importin β *in vivo*.

As we learn more about nuclear transport, it will be interesting to investigate and compare the details of the mechanisms utilized by various organisms. The processes and machinery of nuclear transport are well-conserved from yeast through humans [155-160]; not only are the primary sequences of transport factors similar, many of the resolved structures show that the protein architectures are similar as well. For example, importin α has been crystallized from yeast [58] and mouse [39, 51] and both show the same characteristic molecular shape and cNLS-binding pattern. Structural similarities between yeast and vertebrate importin β [84, 88, 144] and between yeast Nup2 [42, 43] and vertebrate Nup50/Npap60 [43] also highlight the conservation across species and underline the significance of the functional data.

Obviously, though, there are differences between yeast and higher eukaryotes with respect to both the organisms and the molecules themselves and these variations may be informative in teasing apart the details of the nucleocytoplasmic transport pathway. *S. cerevisiae* has only a single importin α, which is encoded by the *SRP1* gene,

but the human genome encodes six isoforms of the protein [161]. Each of the members of the human importin α family acts as an adaptor molecule for importin β, so there is a high degree of functional redundancy; however, the different forms show differential expression patterns throughout development and, importantly, show some variable preference for certain cNLS-cargos [161]. The mechanism for this discrimination is not well understood, as the paralogs share key residues in their cNLS-binding pockets and have identical auto-inhibitory sequences. Flanking residues do differ though, and may contribute to the ability of certain importin α isoforms to bind and import unique cNLS-cargos.

Interestingly, it is possible that the difference in sequential or structural makeup of the human importin α isoforms may affect the cargo dissociation step as well as the cargo binding step. Different isoforms may utilize different cargo release mechanisms, with some relying primarily on the traditional cooperation of RanGTP and the auto-inhibitory domain, while others utilize Nup50/Npap60 and/or CAS/RanGTP. In support of this idea, yeast Nup2 is not required for cell viability [69], though it does affect the localization of importin α [71]; however, the homologous vertebrate protein, Nup50/Npap60, is essential for mouse development [162]. One possibility for this discrepancy may be that Nup50/Npap60 is required to address the complications inherent in coordinating six importin αs. In yeast, Nup2 can accelerate the release of cargos from importin α [41, 56], but the auto-inhibitory mechanism of the one broad-specificity importin α is sufficient to release cNLS-cargo in the nucleus. Perhaps in humans, Nup50/Npap60 is essential because it is necessary for the dissociation of certain classes of cargos that bind to a particular subset of importin α proteins [74].

Throughout this dissertation, we have examined how structural studies have provided detailed snapshots of memorable moments in nuclear protein transport. Portraits of importins and their associates have been painted. Freeze frames of important milestones along the dissociation path have been captured. We have galleries of information, but while a picture is worth a thousand words, a picture also does not tell the whole story. The next challenge is to move from this static knowledge to a dynamic understanding of nucleocytoplasmic transport. Our next goals should focus on elucidating the complex choreography that culminates in rapid and selective movement of proteins into and out of the nucleus. Computer simulations and mechanistic modeling studies have started to address this question, but future experiments will be required to precisely order the events and fully depict the detailed masterpiece that is nuclear transport.

**REFERENCES**

1.      Kaffman A, O'Shea EK: **Regulation of nuclear localization: a key to a door**.
        *Annu Rev Cell Dev Biol* 1999, **15**:291-339.

2.      Johnson HM, Subramaniam PS, Olsnes S, Jans DA: **Trafficking and signaling
        pathways of nuclear localizing protein ligands and their receptors**. *Bioessays*
        2004, **26**(9):993-1004.

3.      Cyert MS: **Regulation of nuclear localization during signaling**. *J Biol Chem*
        2001, **276**(24):20805-20808.

4.      Kalderon D, Richardson WD, Markham AF, Smith AE: **Sequence requirements
        for nuclear location of simian virus 40 large-T antigen**. *Nature* 1984,
        **311**(5981):33-38.

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

6.      Fahrenkrog B, Aebi U: **The nuclear pore complex: nucleocytoplasmic
        transport and beyond**. *Nat Rev Mol Cell Biol* 2003, **4**(10):757-766.

7.      Stoffler D, Fahrenkrog B, Aebi U: **The nuclear pore complex: from molecular
        architecture to functional dynamics**. *Curr Opin Cell Biol* 1999, **11**:391-401.

8.      Allen TD, Cronshaw JM, Bagley S, Kiseleva E, Goldberg MW: **The nuclear
        pore complex: mediator of translocation between nucleus and cytoplasm**. *J
        Cell Sci* 2000, **113**(2000):1651-1659.

9.      Fahrenkrog B, Aebi U: **The vertebrate nuclear pore complex: from structure to function**. *Results Probl Cell Differ* 2002, **35**:25-48.

10.     Paine PL, Moore LC, Horowitz SB: **Nuclear envelope permeability**. *Nature* 1975, **254**(5496):109-114.

11.     Bonner WM: **Protein migration and accumulation in nuclei**, vol. 6. New York: Academic; 1978.

12.     Stewart M, Baker RP, Bayliss R, Clayton L, Grant RP, Littlewood T, Matsuura Y: **Molecular mechanism of translocation through nuclear pore complexes during nuclear protein import**. *FEBS Lett* 2001, **498**(2-3):145-149.

13.     Suntharalingam M, Wente SR: **Peering through the Pore: Nuclear pore complex structure, assembly, and function**. *Dev Cell* 2003, **4**(6):775-789.

14.     Tran EJ, Wente SR: **Dynamic nuclear pore complexes: life on the edge**. *Cell* 2006, **125**(6):1041-1053.

15.     Cordes VC, Reidenbach S, Franke WW: **High content of a nuclear pore complex protein in cytoplasmic annulate lamellae of Xenopus oocytes**. *Eur J Cell Biol* 1995, **68**(3):240-255.

16.     Ribbeck K, Gorlich D: **Kinetic analysis of translocation through nuclear pore complexes**. *Embo J* 2001, **20**(6):1320-1330.

17.     Radu A, Blobel G, Moore MS: **Identification of a protein complex that is required for nuclear protein import and mediates docking of import**

**substrate to distinct nucleoporins**. *Proc Natl Acad Sci USA* 1995, **92**:1769-1773.

18.     Gorlich D, Prehn S, Laskey RA, Hartmann E: **Isolation of a protein that is essential for the first step of nuclear protein import**. *Cell* 1994, **79**(5):767-778.

19.     Stade K, Ford CS, Guthrie C, Weis K: **Exportin 1 (Crm1p) is an essential nuclear export factor**. *Cell* 1997, **90**:1041-1050.

20.     Yoshida K, Blobel G: **The karyopherin Kap142/Msn5p mediates nuclear import and nuclear export of different cargo proteins**. *J Cell Biol* 2001, **152**:729-740.

21.     Kaffman A, Rank NM, O'Neill EM, Huang LS, O'Shea EK: **The receptor Msn5 exports the phosphorylated transcription factor Pho4 out of the nucleus**. *Nature* 1998, **396**(6710):482-486.

22.     Blondel M, Alepuz PM, Huang LS, Shaham S, Ammerer G, Peter M: **Nuclear export of Far1p in response to pheromones requires the export receptor Msn5p/Ste21p**. *Genes Dev* 1999, **13**(17):2284-2300.

23.     DeVit MJ, Johnston M: **The nuclear exportin Msn5 is required for nuclear export of the Mig1 glucose repressor of Saccharomyces cerevisiae**. *Curr Biol* 1999, **9**(21):1231-1241.

24.     Mahanty SK, Wang Y, Farley FW, Elion EA: **Nuclear shuttling of yeast scaffold Ste5 is required for its recruitment to the plasma membrane and activation of the mating MAPK cascade**. *Cell* 1999, **98**(4):501-512.

25.     Mosammaparast N, Pemberton LF: **Karyopherins: from nuclear-transport mediators to nuclear-function regulators.** *Trends in Cell Biology* 2004, **14**(10):547-556.

26.     Riddick G, Macara IG: **A systems analysis of importin-alpha-beta mediated nuclear protein import**. *J Cell Biol* 2005, **168**(7):1027-1038.

27.     Riddick G, Macara IG: **The adapter importin alpha provides flexible control of nuclear import at the expense of efficiency**. *Molecular Systems Biology* 2007, **3**(18):1-7.

28.     Quimby BB, Dasso M: **The small GTPase Ran: interpreting the signs.** *Curr Opin Cell Biol* 2003, **15**(3):338-344.

29.     Bourne HR, Sanders DA, McCormick F: **The GTPase superfamily: a conserved switch for diverse cell functions**. *Nature* 1990, **348**(6297):125-132.

30.     Corbett AH, Koepp DM, Lee MS, Schlenstedt G, Hopper AK, Silver PA: **Rna1p, a Ran/TC4 GTPase activating protein is required for nuclear import**. *J Cell Biol*, **130**(1995):1017-1026.

31.     Becker J, Melchior F, Gerke V, Bischoff FR, Ponstingl H, Wittinghofer A: **RNA1 encodes a GTPase-activating protein specific for Gsp1p, the Ran/TC4**

homologue of *Saccharomyces cerevisiae*. *J Biol Chem* 1995, **270**(20):11860-11865.

32.     Klebe C, Prinz H, Wittinghofer A, Goody RS: **The kinetic mechanism of Ran--nucleotide exchange catalyzed by RCC1**. *Biochemistry* 1995, **34**(39):12543-12552.

33.     Bischoff FR, Ponstingl H: **Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1**. *Nature*, **354**(1991):80-82.

34.     Kalab P, Weis K, Heald R: **Visualization of a Ran-GTP gradient in interphase and mitotic Xenopus egg extracts**. *Science* 2002, **295**(5564):2452-2456.

35.     Smith AE, Slepchenko BM, Schaff JC, Loew LM, Macara IG: **Systems analysis of Ran transport**. *Science* 2002, **295**(5554):488-491.

36.     Görlich D, Kostka S, Kraft R, Dingwall C, Laskey RA, Hartmann E, Prehn S: **Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope**. *Curr Biol* 1995, **5**:383-392.

37.     Lee SJ, Matsuura Y, Liu SM, Stewart MS: **Structural basis for nuclear import complex dissociation by RanGTP.** *Nature* 2005, **435**(7042):693-696.

38.     Görlich D, Panté N, Kutay U, Aebi U, Bischoff FR: **Identification of different roles for RanGDP and RanGTP in nuclear protein import**. *EMBO J* 1996, **15**:5584-5594.

39.     Kobe B: **Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha**. *Nat Struct Biol* 1999, **6**:301-304.

40.     Harreman MT, Hodel MR, Fanara P, Hodel AE, Corbett AH: **The Auto-inhibitory Function of Importin alpha Is Essential in Vivo**. *J Biol Chem* 2003, **278**(8):5854-5863.

41.     Gilchrist D, Mykytka B, Rexach M: **Accelerating the rate of disassembly of karyopherin-cargo complexes**. *J Biol Chem* 2002, **277**(20):18161-18172.

42.     Matsuura Y, Lange A, Harreman MT, Corbett AH, Stewart M: **Structural basis for Nup2p function in cargo release and karyopherin recycling in nuclear import**. *Embo J* 2003, **22**(20):5358-5369.

43.     Matsuura Y, Stewart M: **Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling**. *Embo J* 2005, **24**(21):3681-3689.

44.     Matsuura Y, Stewart M: **Structural basis for the assembly of a nuclear export complex**. *Nature* 2004, **432**(7019):872-877.

45.     Hood JK, Silver PA: **Cse1p is required for export of Srp1p/importin-alpha from the nucleus in *Saccharomyces cerevisiae***. *J Biol Chem* 1998, **273**:35142-35146.

46.    Kutay U, Bischoff FR, Kostka S, Kraft R, Görlich D: **Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor**. *Cell* 1997, **90**:1061-1071.

47.    Görlich D, Vogel F, Mills AD, Hartmann E, Laskey RA: **Distinct functions for the two importin subunits in nuclear protein import**. *Nature* 1995, **377**:246-248.

48.    Robbins J, Dilworth SM, Laskey RA, Dingwall C: **Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence**. *Cell* 1991, **64**:615-623.

49.    Dingwall C, Laskey RA: **Nuclear targeting sequences--a consensus?** *Trends Biochem Sci* 1991, **16**(12):478-481.

50.    Conti E, Kuriyan J: **Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha**. *Structure Fold Des* 2000, **8**(3):329-338.

51.    Fontes MR, Teh T, Kobe B: **Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha**. *J Mol Biol* 2000, **297**(5):1183-1194.

52.    Hodel MR, Corbett AH, Hodel AE: **Dissection of a nuclear localization signal**. *J Biol Chem* 2001, **276**(2):1317-1325.

53.  Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH: **Classical nuclear localization signals: definition, function, and interaction with importin alpha.** *J Biol Chem* 2007, **282**(8):5101-5105.

54.  Fontes MR, Teh T, Jans D, Brinkworth RI, Kobe B: **Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha**. *J Biol Chem* 2003, **278**(30):27981-27987.

55.  Hodel AE, Harreman MT, Pulliam KF, Harben ME, Holmes JS, Hodel MR, Berland KM, Corbett AH: **Nuclear localization signal receptor affinity correlates with in vivo localization in** *Saccharomyces cerevisiae*. *J Biol Chem* 2006, **281**(33):23545-23556.

56.  Gilchrist D, Rexach M: **Molecular basis for the rapid dissociation of nuclear localization signals from karyopherin alpha in the nucleoplasm**. *J Biol Chem* 2003, **278**(51):51937-51949.

57.  Timney BL, Tetenbaum-Novatt J, Agate DS, Williams R, Zhang W, Chait BT, Rout MP: **Simple kinetic relationships and nonspecific competition govern nuclear import rates in vivo.** *J Cell Biol* 2006, **175**(4):579-593.

58.  Conti E, Uy M, Leighton L, Blobel G, Kuriyan J: **Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha**. *Cell* 1998, **94**(1998):193-204.

59.  Fontes MR, Teh T, Toth G, John A, Pavo I, Jans DA, Kobe B: **Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40**

large T-antigen nuclear localization sequences by importin-alpha. *Biochem J* 2003, **375**(Pt 2):339-349.

60.     Fanara P, Hodel MR, Corbett AH, Hodel AE: **Quantitative analysis of nuclear localization signal (NLS)-importin alpha interaction through fluorescence depolarization. Evidence for auto-inhibitory regulation of NLS binding**. *J Biol Chem* 2000, **275**(28):21218-21223.

61.     Harreman MT, Cohen PE, Hodel MR, Truscott GJ, Corbett AH, Hodel AE: **Characterization of the auto-inhibitory sequence within the N-terminal domain of importin alpha**. *J Biol Chem* 2003, **278**(24):21361-21369.

62.     Gruss OJ, Carazo-Salas RE, Schatz CA, Guarguaglini G, Kast J, Wilm M, Le Bot N, Vernos I, Karsenti E, Mattaj IW: **Ran induces spindle assembly by reversing the inhibitory effect of importin alpha on TPX2 activity**. *Cell* 2001, **104**(1):83-93.

63.     Herold A, Truant R, Wiegand H, Cullen BR: **Determination of functional domain organization of the importin alpha nuclear import factor**. *J Cell Biol* 1998, **143**:309-318.

64.     Solsbacher J, Maurer P, Bischoff FR, Schlenstedt G: **Cse1p is involved in export of yeast importin alpha from the nucleus**. *Mol Cell Biol* 1998, **18**:6805-6815.

65.     Vetter IR, Arndt A, Kutay U, Görlich D, Wittinghofer A: **Structural view of the Ran-importin beta interaction at 2.3 Å resolution**. *Cell* 1999, **97**:635-646.

66. Floer M, Blobel G, Rexach M: **Disassembly of RanGTP-karyopherin beta complex, an intermediate in nuclear protein import**. *J Biol Chem* 1997, **272**:19538-19546.

67. Moroianu J, Blobel G, Radu A: **The binding site of karyopherin alpha for karyopherin beta overlaps with a nuclear localization sequence**. *Proc Natl Acad Sci U S A* 1996, **93**(13):6572-6576.

68. Denning DP, Uversky V, Patel SS, Fink AL, Rexach M: **The *Saccharomyces cerevisiae* nucleoporin Nup2p is a natively unfolded protein**. *J Biol Chem* 2002, **277**(36):33447-33455.

69. Loeb JDJ, Davis LI, Fink GR: *NUP2,* **a novel yeast nucleoporin, has functional overlap with other proteins of the nuclear pore complex**. *Mol Biol Cell* 1993, **4**:209-222.

70. Solsbacher J, Maurer P, Vogel F, Schlenstedt G: **Nup2p, a yeast nucleoporin, functions in bidirectional transport of importin alpha**. *Mol Cell Biol* 2000, **20**(22):8468-8479.

71. Hood J, Casolari JM, Silver PA: **Nup2p is located on the nuclear side of the nuclear pore complex and coordinates Srp1p/importin-alpha export**. *J Cell Sci* 2000, **113**:1471-1480.

72. Allen NPC, Huang L, Burlingame A, Rexach M: **Proteomic analysis of nucleoporin interacting proteins**. *J Biol Chem* 2001, **276**(2001):29268-29274.

73.    Lindsay ME, Plafker K, Smith AE, Clurman BE, Macara IG: **Npap60/Nup50 is a tri-stable switch that stimulates importin alpha:beta-mediated nuclear protein import**. *Cell* 2002, **110**:349-360.

74.    Moore MS: **Npap60: a new player in nuclear protein import**. *Trends in Cell Biology* 2003, **13**(2):61-64.

75.    Booth JW, Belanger KD, Sannella MI, Davis LI: **The yeast nucleoporin Nup2p is involved in nuclear export of importin alpha/Srp1p**. *J Biol Chem* 1999, **274**(45):32360-32367.

76.    Wozniak RW, Rout MP, Aitchison JD: **Karyopherins and kissing cousins**. *Trends Cell Biol* 1998, **8**:184-188.

77.    Harel A, Forbes DJ: **Importin beta:  Conducting a much larger cellular symphony**. *Molecular Cell* 2004, **16**:319-330.

78.    Strom AC, Weis K: **Importin-beta-like nuclear transport receptors**. *Genome Biol* 2001, **2**(6).

79.    Cingolani G, Petosa C, Weis K, Müller CW: **Structure of importin-β bound to the IBB domain of importin-α**. *Nature*, **399**(1999):221-229.

80.    Bayliss R, Littlewood T, Stewart M: **Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking**. *Cell*, **102**(2000):99-108.

81.     Liu SM, Stewart MS: **Structural basis for the high-affinity binding of nucleoporin Nup1p to the Saccharomyces cerevisiae importin-beta homologue, Kap95p**. *J Mol Biol* 2005, **349**(3):515-525.

82.     Cingolani G, Bednenko J, Gillespie MT, Gerace L: **Molecular basis for the recognition of a nonclassical nuclear localization signal by importin Beta**. *Mol Cell* 2002, **10**(6):1345-1353.

83.     Lee SJ, Sekimoto T, Yamashita E, Nagoshi E, Nakagawa A, Imamoto N, Yoshimura M, Sakai H, Chong KT, Tsukihara T *et al*: **The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor.** *Science* 2003, **302**(5650):1571-1575.

84.     Stewart M: **Structural Biology: Nuclear Trafficking**. *Science* 2003, **302**(5650):1513-1514.

85.     Zachariae U, Grubmuller H: **A highly strained nuclear conformation of the exportin Cse1p revealed by molecular dynamics simulations**. *Structure* 2006, **14**:1469-1478.

86.     Zachariae U, Grubmuller H: **Importin-beta: structural and dynamic determinants of a molecular spring**. *Structure* 2008, **16**(6):906-915.

87.     Cook A, Fernandez E, Lindner D, Ebert J, Schlenstedt G, E. C: **The structure of the nuclear export receptor Cse1 in its cytosolic state reveals a closed conformation incompatible with cargo binding.** *Mol Cell* 2005, **18**(3):355-367.

88.    Fukuhara N, Fernandez E, Ebert J, Conti E, Svergun D: **Conformational variability of nucleo-cytoplasmic transport factors**. *Journal of Biological Chemistry* 2004, **279**(3):2176-2181.

89.    Fischer U, Huber J, Boelens WC, Mattaj IW, Luhrmann R: **The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs**. *Cell* 1995, **82**(3):475-483.

90.    Wen W, Meinkoth JL, Tsien RY, Taylor SS: **Identification of a signal for rapid export of proteins from the nucleus**. *Cell* 1995, **82**:463-473.

91.    Damelin M, Silver PA, Corbett AH: **Nuclear protein transport**. *Methods Enzymol* 2002, **351**(2002):587-607.

92.    Lee BJ, Cansizoglu AE, Suel KE, Louis TH, Zhang Z, Chook YM: **Rules for nuclear localization sequence recognition by karyopherinbeta2**. *Cell* 2006, **126**(3):543-558.

93.    Siomi MC, Fromont M, Rain JC, Wan L, Wang F, Legrain P, Dreyfuss G: **Functional conservation of the transportin nuclear import pathway in divergent organisms**. *Mol Cell Biol* 1998, **18**(7):4141-4148.

94.    Gorlich D, Kostka S, Kraft R, Dingwall C, Laskey RA, Hartmann E, Prehn S: **Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope**. *Curr Biol* 1995, **5**(4):383-392.

95.     Xiao Z, Latek R, Lodish HF: **An extended bipartite nuclear localization signal in Smad4 is required for its nuclear import and transcriptional activity**. *Oncogene* 2003, **22**:1057-1069.

96.     Kim KH, Kanbe T, Akashi T, Mizuguchi I, Kikuchi A: **Identification of a single nuclear localization signal in the C-terminal domain of an Aspergillus DNA topoisomerase II**. *Mol Genet Genomics* 2002, **268**(3):287-297.

97.     Kenna MA, Brachmann CB, Devine SE, Boeke JD: **Invading the yeast nucleome:  a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition in vivo**. *Mol Cell Biol* 1998, **18**(1115-1124).

98.     Moore SP, Rinckel LA, Garfinkel DJ: **A Ty1 integrase nuclear localization signal required for retrotransposition**. *Molecular and Cellular Biology* 1998, **18**(2):1105-1114.

99.     Adams A, Gottschling DE, Kaiser CA, Stearns T: **Methods in Yeast Genetics**. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1997.

100.    Sambrook J, Russell D: **Molecular Cloning: A Laboratory Manual**, Third edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2001.

101.    Nakai K, Horton P: **PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization**. *Trends Biochem Sci* 1999, **24**(1):34-36.

146

102.    Gattiker A, Gasteiger E, Bairoch A: **ScanProsite: a reference implementation of a PROSITE scanning tool.** *Appl Bioinformatics* 2002, **1**(2):107-108.

103.    Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank**. *Nucleic Acids Res* 2006, **34**(Database issue):D16-20.

104.    Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK: **Global analysis of protein localization in budding yeast**. *Nature* 2003, **425**(6959):686-691.

105.    Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets**. *Nucleic Acids Res* 2006, **34**(Database issue):D535-539.

106.    Winston F, Dollard C, Ricupero-Hovasse SL: **Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C**. *Yeast* 1995, **11**(11):53-55.

107.    Yano R, Oakes ML, Tabb MM, Nomura M: **Yeast Srp1p has homology to armadillo/plakoglobin/β-catenin and participates in apparently multiple nuclear functions including the maintenance of the nucleolar structure**. *Proc Natl Acad Sci USA* 1994, **91**:6880-6884.

108.    Leung SW, Harreman MT, Hodel MR, Hodel AE, Corbett AH: **Dissection of the karyopherin alpha nuclear localization signal (NLS)-binding groove: functional requirements for NLS binding**. *J Biol Chem* 2003, **278**(43):41947-41953.

109. Bayliss R, Littlewood T, Strawn LA, Wente SR, Stewart M: **GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta**. *J Biol Chem* 2002, **277**(52):50597-50606.

110. Fagerlund R, Melen K, Kinnunen L, Julkunen I: **Arginine/lysine-rich nuclear localization signals mediate interactions between dimeric STATs and importin alpha 5**. *J Biol Chem* 2002, **277**(33):30072-30078.

111. Aitchison JD, Blobel G, Rout MP: **Kap104p: A karyopherin involved in the nuclear transport of messenger RNA binding proteins**. *Science*, **274**(1996):624-627.

112. Lee DC, Aitchison JD: **Kap104p-mediated nuclear import. Nuclear localization signals in mRNA- binding proteins and the role of Ran and Rna**. *J Biol Chem* 1999, **274**(41):29031-29037.

113. Truant R, Fridell RA, Benson RE, Bogerd H, Cullen BR: **Identification and functional characterization of a novel nuclear localization signal present in the yeast Nab2 poly(A)+ RNA binding protein**. *Mol Cell Biol* 1998, **18**(3):1449-1458.

114. Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, Moore CL: **Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast**. *Genes Dev* 1997, **11**(19):2545-2556.

115.    Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR, Wilson SM, Minvielle-Sebastia L, Swanson MS: **Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export**. *Embo J* 2002, **21**(7):1800-1810.

116.    Green DM, Marfatia KA, Crafton EB, Zhang X, Cheng X, Corbett AH: **Nab2p Is required for poly(A) RNA export in *Saccharomyces cerevisiae* and is regulated by arginine methylation via Hmt1p**. *J Biol Chem* 2002, **277**(10):7752-7760.

117.    Sikorski RS, Hieter P: **A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae***. *Genetics* 1989, **122**:19-27.

118.    Leslie DM, Zhang W, Timney BL, Chait BT, Rout MP, Wozniak RW, Aitchison JD: **Characterization of karyopherin cargoes reveals unique mechanisms of Kap121p-mediated nuclear import.** *Mol Cell Biol* 2004, **24**(19):8487-8503.

119.    Boeke JD, Truehart J, Natsoulis G, Fink G: **5-Fluoroorotic acid as a selective agent in yeast molecular genetics**. *Meth Enzymol* 1987, **154**:164-175.

120.    Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank**. *Nucleic Acids Res* 2007, **35**(Database issue):D21-D25.

121.    Weis K: **Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle**. *Cell* 2003, **112**(4):441-451.

122. Harreman MT, Kline TM, Milford HG, Harben MB, Hodel AE, Corbett AH: **Phosphorylation of nuclear import by phosphorylation adjacent to nuclear localization signals**. *J Biol Chem* 2004, **279**(20):20613-20621.

123. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al*: **Crystallography and NMR system: a new software suite for macromolecular structure determination**. *Acta Crytsallogr* 1998, **54**(Pt 5):905-921.

124. Collaborative Computational Project N: **The CCP4 suite: programs for protein crystallography**. *Acta Crystallogr D Biol Crystallogr* 1994, **50**(Pt 5):760-763.

125. Schlenstedt G, Wong DH, Koepp DM, Silver PA: **Mutants in a yeast Ran binding protein are defective in nuclear transport**. *EMBO J* 1995, **14**:5367-5378.

126. Denning D, Mykytka B, Allen NP, Huang L, Al B, Rexach M: **The nucleoporin Nup60p functions as a Gsp1p-GTP-sensitive tether for Nup2p at the nuclear pore complex**. *J Cell Biol* 2001, **154**(5):937-950.

127. Dilworth DJ, Suprapto A, Padovan JC, Chait BT, Wozniak RW, Rout MP, Aitchison JD: **Nup2p dynamically associates with the distal regions of the yeast nuclear pore complex**. *J Cell Biol* 2001, **153**(7):1465-1478.

128. Loeillet S, Palancade B, Cartron M, Thierry A, Richard GF, Dujon B, Doye V, Nicolas A: **Genetic network interactions among replication, repair and nuclear pore deficiencies in yeast.** *DNA Repair (Amst)* 2005, **4**(4):459-468.

129.	Xiao Z, McGrew JT, Schroeder AJ, Fitzgerald-Hayes M: *CSE1 and CSE2,* **two new genes required for accurate mitotic chromosome segregation in** *Saccharomyces cerevisiae*. *Mol Cell Biol* 1993, **13**:4691-4702.

130.	Inoue T, Wu L, Stuart J, Maki CG: **Control of p53 nuclear accumulation in stressed cells**. *FEBS Letters* 2005, **579**:4978-4984.

131.	Brinkmann U, Gallo M, Polymeropoulos MH, Pastan I: **The human CAS (cellular apoptosis susceptibility) gene mapping on chromosome 20q13 is amplified in BT474 breast cancer cells and part of aberrant chromosomes in breast and colon cancer cell lines**. *Genome Res*, **6**(1996):187-194.

132.	Kim IS, Kim DH, Han SM, Chin MU, Nam HJ, Cho HP, Choi SY, Song BJ, Kim ER, Bae YS *et al*: **Truncated form of importin alpha identified in breast cancer cell inhibits nuclear import of p53**. *J Biol Chem* 2000, **275**(30):23139-23145.

133.	Lee H, Ueda M, Miyamoto Y, Yoneda Y, Perry G, Smith MA, Zhu X: **Aberrant localization of importin alpha1 in hippocampal neurons in Alzheimer disease**. *Brain Research* 2006, **1124**:1-4.

134.	Wei J, Hemmings GP: **The KPNA3 gene may be a susceptibility candidate for schizophrenia**. *Neuroscience Research* 2005, **52**:342-346.

135.	Harley VR, Layfield S, Mitchell CL, Forwood JK, John AP, Briggs LJ, McDowall SG, Jans DA: **Defective importin beta recognition and nuclear**

import of the sex-determining factor SRY are associated with XY sex-reversing mutations. *PNAS* 2003, **100**(12):7045-7050.

136.    Cook A, Bono F, Jinek M, Conti E: **Structural biology of nucleocytoplasmic transport.** *Annu Rev Biochem* 2007, **76**:647-671.

137.    Koepp DM, Wong DH, Corbett AH, Silver PA: **Dynamic localization of the nuclear import receptor and its interactions with transport factors**. *J Cell Biol* 1996, **133**:1163-1176.

138.    Loeb JDL, Schlenstedt G, Pellman D, Kornitzer D, Silver PA, Fink G: **A nuclear import receptor homologue required for mitosis in yeast**. *Proc Natl Acad Sci, USA* 1995, **92**:7647-7651.

139.    Towbin H, Staehelin T, Gordon J: **Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets:  Procedures and some application**. *Proc Natl Acad Sci USA* 1979, **76**:4350-4354.

140.    Shulga N, Roberts P, Gu Z, Spitz L, Tabb MM, Nomura M, Goldfarb DS: ***In vivo* nuclear transport kinetics in *Saccharomyces cerevisiae*: a role for heat shock protein 70 during targeting and translocation**. *J Cell Biol* 1996, **135**:329-339.

141.    Kutay U, Izaurralde E, Bischoff FR, Mattaj IW, Görlich D: **Dominant-negative mutants of importin-beta block multiple pathways of import and export through the nuclear pore complex.** *EMBO J* 1997, **16**:1153-1163.

142.    Chi NC, Adam EJH, Adam SA: **Different binding domains for Ran-GTP and Ran-GDP/RanBP1 on the nuclear import factor p97**. *J Biol Chem* 1997, **272**:6818-6822.

143.    Enenkel C, Blobel G, Rexach M: **Identification of a yeast karyopherin heterodimer that targets import substrate to mammalian nuclear pore complexes**. *J Biol Chem* 1995, **270**(1995):16499-16502.

144.    Conti E, Muller CW, Stewart M: **Karyopherin flexibility in nucleocytoplasmic transport**. *Curr Opin Struct Biol* 2006, **16**(2):237-244.

145.    Cingolani G, Petosa C, Weis K, Müller CW: **Structure of importin-beta bound to the IBB domain of importin-alpha**. *Nature*, **399**(1999):221-229.

146.    Suel KE, Gu H, Chook YM: **Modular organization and combinatorial energetics of proline-tyrosine nuclear localization signals**. *PLoS* 2008, **6**(6):1253-1267.

147.    Poon IK, Jans DA: **Regulation of nuclear transport: central role in development and transformation?** *Traffic* 2005, **6**(3):173-186.

148.    Terry LJ, Shows EB, Wente SR: **Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport**. *Science* 2007, **318**(5855):1414-1416.

149.    Smith WA, Schurter BT, Wong-Staal F, David M: **Arginine methylation of RNA helicase a determines its subcellular localization**. *J Biol Chem* 2004, **279**(22):22795-22798.

150.    Trotman LC, Wang X, Alimonti A, Chen Z, Teruya-Feldstein J, Yang H, Pavletich NP, Carver BS, Cordon-Cardo C, Erdjument-Bromage H *et al*: **Ubiquitination Regulates PTEN Nuclear Import and Tumor Suppression**. *Cell* 2007, **128**(1):141-156.

151.    Stommel JM, Marchenko ND, Jimenez GS, Moll UM, Hope TJ, Wahl GM: **A leucine-rich nuclear export signal in the p53 tetramerization domain: regulation of subcellular localization and p53 activity by NES masking.** *EMBO J* 1999, **18**(6):1660-1672.

152.    Huxford T, Huang DB, Malek S, Ghosh G: **The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation**. *Cell* 1998, **95**(6):759-770.

153.    Jacobs MD, Harrison SC: **Structure of an IkappaBalpha/NF-kappaB complex**. *Cell* 1998, **95**(6):749-758.

154.    Briggs LJ, Stein D, Goltz J, Corrigan VC, Efthymiadis A, Hübner S, Jans DA: **The cAMP-dependent protein kinase site (Ser312) enhances dorsal nuclear import through facilitating nuclear localization sequence/importin interaction.** *J Biol Chem* 1998, **273**(35):22745-22752.

155. Quan Y, Ji ZL, Wang X, Tartakoff AM, Tao T: **Evolutionary and transcriptional analysis of karyopherin beta superfamily proteins**. *Mol Cell Proteomics* 2008, **7**(7):1254-1269.

156. Görlich D, Kutay U: **Transport between the cell nucleus and the cytoplasm**. *Annu Rev Cell Dev Biol* 1999, **15**:607-660.

157. Marfatia KA, Harreman MT, Fanara P, Vertino PM, Corbett AH: **Identification and characterization of the human MOG1 gene**. *Gene* 2001, **266**(1-2):45-56.

158. Corbett AH, Silver PA: **The *NTF2* gene encodes an essential, highly conserved protein that functions in nuclear transport *in vivo***. *J Biol Chem*, **271**(1996):18477-18484.

159. Melchior F, Weber K, Gerke V: **A functional homologue of the *RNA1* gene product in *Schizosaccharomyces pombe*: Purification, biochemical characterization, and identification of leucine-rich repeat motif**. *Molec Biol Cell* 1993, **4**:569-581.

160. Fleischmann M, Clark MW, Forrester W, Wickens M, Nishimoto T, Aebi M: **Analysis of yeast *prp20* mutations and functional complementation by the human homolgue RCC1, a protein involved in the control of chromosome condensation**. *Mol Gen Genet* 1991, **227**:417-423.

161. Goldfarb DS, Corbett AH, Mason DA, Harreman MT, Adam SA: **Importin alpha: a multipurpose nuclear-transport receptor**. *Trends in Cell Biology* 2004, **14**(9):505-514.

162. Smitherman M, Lee K, Swanger J, Kapur R, Clurman BE: **Characterization and targeted disruption of murine Nup50, a p27(Kip1)- interacting component of the nuclear pore complex**. *Mol Cell Biol* 2000, **20**(15):5631-5642.