RECOGNITION MECHANISMS OF NUCLEAR

LOCALIZATION AND EXPORT SIGNALS

APPROVED BY SUPERVISORY COMMITTEE

Yuh Min Chook, Ph.D.

Hongtao Yu, Ph.D.

Beatriz Fontoura, Ph.D.

Rama Ranganathan, M.D., Ph.D.

DEDICATION

I would like to thank my mentor Yuh Min Chook for her assistance and guidance throughout my graduate career. I have learned so much from her: not only about how to perform experiments, but how to approach a scientific problem and manage a laboratory. I would also like to thank all the members of the Chook lab, especially Brittany Lee. She was not only my bench mate, but is one of my best friends. She was/is always there when I need a friend and support. All of the members of the Chook lab made it an enjoyable environment including Ertugrul, Hongmei, Zichao, Xiuhua and Ani. I would especially like to thank Hongmei Gu for helping me with protein purification and the Hrp1/Nab2p project. I would like to thank Tolga Çağatay for microscopy assistance, but more importantly for his friendship. I would like to thank the members of my dissertation committee, Hongtao Yu, Beatriz Fontoura and Rama Ranganathan for their support and guidance.

I would like to thank my family, Mom, Dad and my sister Melissa for their support and for providing me with a stable foundation. I would especially like to thank my Mother. She went back to college to get her degree while raising two small children, teaching me the importance of education while being a great role model. I would also like to thank my Dad for giving me his strength. Most importantly, I would like to thank my husband Gürol. We sacrificed a lot so that I would be able to finish my degree at UTSW. Words cannot express how much I love, admire and respect you.

RECOGNITION MECHANISMS OF NUCLEAR LOCALIZATION AND EXPORT SIGNALS

by

KATHERINE ELIZABETH SÜEL

DISSERTATION /THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

January, 2009

RECOGNITION MECHANISMS OF NUCLEAR

LOCALIZATION AND EXPORT SIGNALS

Katherine Elizabeth Süel, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2009

Supervising Professor: Yuh Min Chook, Ph.D.

The transport of proteins between the nucleus and cytoplasm of cells is mediated by the Karyopherin beta family of proteins. Karyopherin betas recognize their substrates through either a nuclear localization or export signal depending on the direction of transport. Even though there are ten yeast import Karyopherin betas, for the past thirteen years there was only one well characterized nuclear localization signal, the classical nuclear localization signal. However, a second signal, the proline-tyrosine nuclear localization signal recognized by Karyopherin beta2, was recently identified through X-ray crystallography and biochemical studies of Karyopherin beta2 bound to one of its

substrates. These studies identified rules for the recognition of the proline-tyrosine nuclear localization signal by Karyopherin beta2. The signal must have overall basic charge, structural disorder and a weak consensus sequence of an amino-terminal basic or hydrophobic-enriched region followed by a carboxyl-terminal arginine residue separated from a proline and tyrosine residue by two to five residues.

The proline-tyrosine nuclear localization signal is also recognized by the *Saccharomyces cerevisiae* homolog of Karyopherin beta2, Karyopherin 104, demonstrating the generality of this import mechanism across eukaryotes. Thermodynamic analyses of the two known substrates of Karyopherin 104, Hrp1p and Nab2p, revealed physical properties governing its binding. The proline-tyrosine nuclear localization signal is an extended signal with significant sequence diversity. The signal is comprised of three binding epitopes, each of which can have varying energetic strengths in different substrates. The multivalent nature of the signal increases the diversity of the signal as well as the difficulty of identifying new substrates. A bioinformatics search identified putative proline-tyrosine nuclear localization signals which were validated through biochemical studies. Additionally, one of the proteins identified, Tfg2p, was verified as a bona fide Karyopherin 104 substrate. Analysis of Tfg2p's cellular localization revealed that its nuclear localization was not solely determined by the presence of a nuclear localization signal, but was also dependent on its retention in the nucleus. Furthermore, crystallographic studies of substrate Snurportin1 bound to the export karyopherin CRM1 revealed that its nuclear export signal has two binding

epitopes implying that the multivalent nature of targeting signals may not be limited to the proline-tyrosine nuclear localization signal.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION 1
Nucleocytoplasmic Transport 1
Direct binding versus binding through an adaptor protein 4
Classical NLS 5
Kap104p/Karyopherin β2 substrates 7
Import by Kap121p and Kap123p 8
Import by other yeast Kapβs 9
Redundancy among yeast Kapβs 10
Nuclear Import Rates 11
Conclusion 13

CHAPTER 2
RULES FOR NUCLEAR LOCALIZATION SEQUENCE RECOGNITION BY KARYOPHERINβ2[a] 22
Abstract 22
Introduction 23
Materials and Methods 26
Kapβ2Δloop cloning 26
Protein expression, purification and complex formation 26
Crystallization, data collection and structure determination 27
NLS-mapping, site directed mutagenesis and Kap β2 binding assays 30
Quantitation of binding affinity with ITC 31
Bioinformatics search for new Kapβ2 substrates 32
Results 33
Kapβ2-M9NLS complex: Structure overview 33
The Kapβ2-M9NLS binding interface 35
Distribution of binding energy along M9NLS 36
Rules for substrate recognition by Kapβ2 40
The NLS rules are predictive 46
Discussion 48

CHAPTER 3
MODULAR ORGANIZATION AND COMBINATORIAL ENERGETICS OF PY-NUCLEAR LOCALIZATION SIGNALS[a] 50
Abstract 50
Introduction 52
Materials and Methods 58
Plasmids and Strains 58

Cell culture and microscopy 59
Bioinformatics search for new Kap104p substrates 59
Protein expression and purification 60
Binding assays 67
Isothermal titration calorimetry 67
Results 68
Yeast rg-NLSs are also PY-NLSs. 68
Kap104p recognizes the basic but not hydrophobic subclass of PY-NLSs. 71
Kapβ2-NLS structures explain Kap104p subclass specificity. 75
Prediction of Kapβ2 homologs that recognize hPY-NLSs 77
Distribution of binding energy along the Hrp1p-NLS. 80
Distribution of binding energy along the Nab2p-NLS. 83
Degeneracy of tyrosine in the C-terminal PY motif. 89
Hrp1p and Nab2p mutants are mislocalized in vivo. 93
Discussion 96
Conclusion 105

CHAPTER 4
IDENTIFICATION AND VALIDATION OF PUTATIVE KAP104p NLSs REVEALS REDUNDANCY AMONG YEAST KAPβs 107
Abstract 107
Introduction 108
Materials and Methods 111
Plasmids 111
Cell culture and microscopy 111
Bioinformatics search for new Kap104p substrates 111
Protein expression and purification 112
Binding assays 112
Results 113
Identification and in vitro validation of putative Kap104p NLSs 113
Validation of putative NLSs as functional targeting signals in cells 119
Dependence on Kap104p for nuclear import 119
Recognition of NLS by other karyopherins 122
Discussion 125

CHAPTER 5
KAP104p IMPORTS THE PY-NLS-CONTAINING TRANSCRIPTION FACTOR TFG2p INTO THE NUCLEUS 130
Abstract 130
Introduction 131

Materials and Methods 134
Plasmids 134
Cell culture and microscopy 135
Protein expression and purification 135
Binding assays 136
Viability assays 136
Results 137
Tfg2p contains a PY-NLS recognized by Kap104p 137
Tfg2p PY-NLS has a strong epitope 3 141
Only Kap104p binding is affected by mutations in the PY motif 143
Nuclear import of full length Tfg2 appears unaffected by loss of Kap104p recognition 143
Other karyopherins bind Tfg2p 146
The DNA binding domain of Tfg2p is required for its nuclear accumulation 148
Tfg2 is sequestered in the nucleus via DNA binding 150
The PY-NLS is required for cell viability 153
Discussion 154

CHAPTER 6
LEUCINE-RICH NUCLEAR EXPORT SIGNAL RECOGNITION BY CRM1 159
Abstract 159
Introduction 160
Materials and Methods 165
In vitro binding assays 165
Plasmids 165
Cell culture and microscopy 166
Results 167
SPN1 1-15 is a functional NES exported by CRM1 in S. cerevisiae cells 167
L4 and L8 are necessary for efficient export of SPN1 169
Discussion 172

CHAPTER 7
CONCLUSIONS 174

APPENDIX 179
Plasmids 179
Tables of predicted Kapβ2 substrates 195

BIBLIOGRAPHY 200

PRIOR PUBLICATIONS

Süel KE, Gu H, Chook YM. *Modular organization and combinatorial energetics of proline-tyrosine nuclear localization signals.* PLoS Biol. 2008. 6(6):e137.

Lee BJ, Cansizoglu AE, Süel KE, Louis TH, Zhang Z, Chook YM. *Rules for nuclear localization sequence recognition by Karyopherinβ2*. Cell. 2006. 126(3):543-58.

Süel KE, Cansizoglu AE, Chook YM. *Atomic resolution structures in nuclear transport.* Methods. 2006. 39(4):342-55.

LIST OF FIGURES

Figure 1-1 2
Figure 1-2 3
Figure 2-1 28
Figure 2-2 28
Figure 2-3 34
Figure 2-4 37
Figure 2-5 40
Figure 2-6 43
Figure 3-1 59
Figure 3-2 61
Figure 3-3 62
Figure 3-4 63
Figure 3-5 65
Figure 3-6 66
Figure 3-7 69
Figure 3-8 70
Figure 3-9 71
Figure 3-10 73
Figure 3-11 74
Figure 3-12 76
Figure 3-13 79
Figure 3-14 82

Figure 3-15 92
Figure 3-16 94
Figure 3-17 95
Figure 4-1 111
Figure 4-2 116
Figure 4-3 116
Figure 4-4 117
Figure 4-5 117
Figure 4-6 118
Figure 4-7 120
Figure 4-8 121
Figure 4-9 122
Figure 4-10 124
Figure 5-1 138
Figure 5-2 139
Figure 5-3 140
Figure 5-4 142
Figure 5-5 144
Figure 5-6 145
Figure 5-7 147
Figure 5-8 149
Figure 5-9 151
Figure 5-10 152

Figure 5-11 153

Figure 6-1 162

Figure 6-2 164

Figure 6-3 166

Figure 6-4 167

Figure 6-5 168

Figure 6-6 169

Figure 6-7 170

Figure 6-8 171

LIST OF TABLES

Table 1-1 14

Table 1-2 15

Table 1-3 15

Table 1-4 17

Table 1-5 18

Table 1-6 19

Table 1-7 19

Table 1-8 20

Table 1-9 21

Table 1-10 21

Table 2-1 29

Table 2-2 39

Table 3-1 55

Table 3-2 81

Table 3-3 84

Table 3-4 85

Table 3-5 90

Table 3-6 91

Table 4-1 114

Table 4-2 123

Table 4-3 126

Table 7-1 179

Table 7-2 180

Table 7-3 182

Table 7-4 184

Table 7-5 187

Table 7-6 190

Table 7-7 194

Table 7-8 195

Table 7-9 198

LIST OF APPENDICES

Plasmids ……..179

Table of predicted Kapβ2 substrates …….…….195

LIST OF ABBREVIATIONS

ATP	adenosine triphosphate
B-ME	beta-mercaptoethanol
DTT	dithiothreitol
E.coli	*Eschericheria coli*
EDTA	ethylenediamine tetra-acetic acid
GAP	GTPase activating protein
GDP	guanosine 5'-diphosphate
GEF	guanine nucleotide exchange factor
GST	glutathione S-transferase
GTP	guanosine 5'-triphosphate
HEAT	Huntington, Elongation factor 3, 'A' subunit of protein phosphatase-2A, and TOR1
HEPES	4-(2-hydroxyethyl)-1-piperizine-ethanesulfonic acid
hnRNP	heterogeneous nuclear ribonucleoprotein
IPTG	isopropyl β-D-thiogalactoside
ITC	isothermal titration calorimetry
Kap	karyopherin
Kapα	Karyopherin alpha
Kapβ1	Karyopherin beta-1
Kapβ2	Karyopherin beta-2
K_D	disociation constant

kDa	kilo Dalton
MBP	maltose binding protein
NES	nuclear export signal
NLS	nuclear localization signal
NPC	nuclear pore complex
OD_{600}	optical density at 600 nm
PCR	polymerase chain reaction
PY-NLS	proline-tyrosine nuclear localization signal
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

CHAPTER 1

INTRODUCTION

Nucleocytoplasmic Transport

Eukaryotic cells are compartmentalized into the cytoplasm and nucleus by the nuclear envelope. Small molecules, ions and proteins can freely diffuse through pores in the nuclear envelope referred to as nuclear pore complexes, but larger proteins are transported through the pore by members of the Karyopherin β family of proteins. This receptor mediated process is essential to cellular viability as all proteins are translated in the cytoplasm but many have functional roles in the nucleus.

There are at least twenty mammalian Karyopherin β (Kapβ) proteins and 14 yeast Kapβs (Table 1-1). Ten of the yeast Kapβs specialize in import of proteins into the nucleus, three exclusively export proteins to the cytoplasm while one is a bidirectional transporter. Kapβs all have comparable molecular weights (95-145 kDa) and acidic isoelectric points (4.0-5.0) (Chook and Blobel 2001). They are superhelical proteins consisting of 19-20 HEAT repeats, which are two antiparallel helices connected by a short turn (Figure 1-1). The HEAT repeats are arranged into an N-terminal arch which binds Ran and a C-terminal arch which binds substrates.

Kapβ proteins recognize their substrates through either nuclear localization signals (NLSs) or nuclear export signals (NESs) in the protein and transport them through

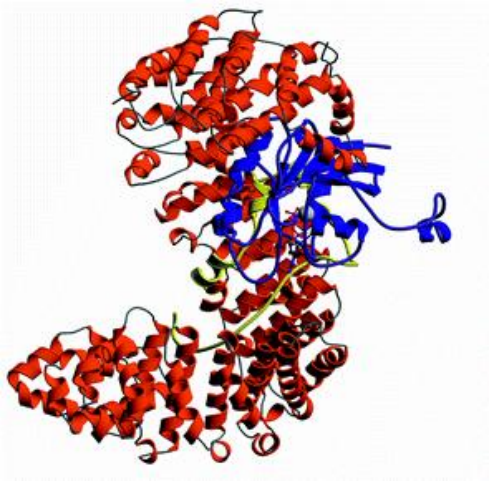


Figure 1-1: Ribbon diagram of Kapβ2 bound to RanGTP. The HEAT repeats of Kapβ2 are in red and RanGTP is colored blue (Chook and Blobel 1999).

the nuclear pore complex (NPC) (Figure 1-2). The NPC is a large macromolecular assembly comprised of about 30 different proteins called nucleoporins. Kapβs transiently bind FG-repeats in nucleoporins and translocate through the nuclear pore complex. Directionality across the NPC is regulated by a gradient of the small GTPase Ran. In the nucleus there is a 100-fold greater concentration of RanGTP than in the cytoplasm due to presence of RCC1, Ran's guanine-exchange factor. RCC1 is primarily localized in the nucleus where it binds directly to histones H2A and H2B (Bischoff and Ponstingl 1995; Nemergut, Mizzen et al. 2001). In the cytoplasm the majority of the small GTPase is in the GDP bound form. This is due to the cytoplasmic localization of the GTPase activating protein RanGAP and the RanGTP binding protein RanBP1 (Bischoff, Klebe et al. 1994). Kapβs bind preferentially to RanGTP; the binding affinity for RanGTP is subnanomolar while the affinity for RanGDP is 10 μM (Vetter, Arndt et al. 1999).

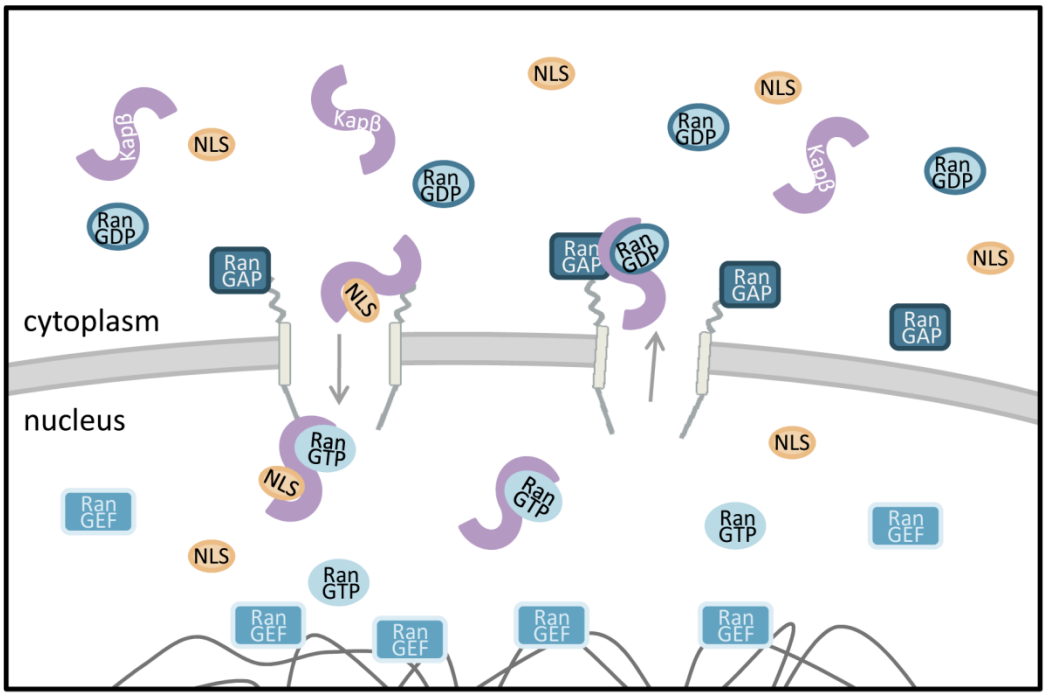


Figure 1-2: Overview of the nuclear import process. Kapβs recognize their substrates in the cytoplasm through NLSs and translocate through the nuclear pore complex into the nucleus. Once in the nucleus, Kapβ binds RanGTP and releases its substrate. There is a high concentration of RanGTP in the nucleus due to the presence of its RanGEF. The Kapβ-RanGTP complex returns to the cytoplasm where RanGTP is hydrolyzed to RanGDP and subsequently released.

Import Kapβs bind substrates in the absence of Ran and release them upon binding RanGTP in the nucleus while export Kapβs bind cooperatively with RanGTP and substrate. The RanGTP/GDP gradient across the nuclear envelope ensures the correct directionality of nuclear transport. Nachury and Weis demonstrated that an inversion of the Ran gradient causes an inversion of the directionality of transport *in vitro* (Nachury and Weis 1999). Addition of RanQ69L (which cannot efficiently hydrolyze GTP to GDP) to the cytoplasm resulted in the transport of export Kapβs Crm1 and CAS cargos into the nucleus as high concentrations of RanGTP in the cytoplasm allowed assembly of

the export complex. While the directionality of transport is determined by the distribution of RanGTP, the localization of a specific protein is predominantly determined by the presence of an NLS or NES.

Direct binding versus binding through an adaptor protein

Transport into the nucleus can either occur via direct recognition of the NLS by the Karyopherin β or through an adaptor protein. The most well studied and utilized adaptor pathway is that of Karyopherin β1/Karyopherin α (Kap95p/Kap60p in yeast). Kapα recognizes its substrates in the cytoplasm and transports them into the nucleus by forming a trimeric complex with Kapβ1.

Recently, Riddick and Macara have examined the reason for the existence of adaptor proteins. Contrary to their original hypothesis and intuition, they showed that direct transport is faster than transport mediated by an adaptor protein (Riddick and Macara 2007). Substrates imported directly by Kapβ1 had a higher steady state nuclear accumulation and a larger initial import rate than substrates imported by the Kapα/Kapβ1 pathway. However, the adaptor pathway has increased dynamic range for control of import rates and therefore has more flexible control of substrate gradients under different cellular conditions. Thus, even though this pathway requires more energy, the authors suggest that this pathway must have an evolutionary advantage, in particular in its increased robustness against environmental influences.

Several proteins are known to depend on a Kapβ for import into the nucleus, but do not bind the karyopherin directly. For example, the ββ' subunits of ribonucleotide

reductase (RNR) are mislocalized to the cytoplasm in *Kap122Δ* cells, but they do not bind Kap122p directly (Zhang, An et al. 2006). Instead, it is hypothesized that they are imported in a trimeric complex with Wtm1p, a nuclear WD40 repeat protein. Other substrates have been known to piggyback into the nucleus. Sof1p, an rRNA processing protein, can either be transported into the nucleus directly by Kap121p or use Nop1p as an adaptor protein (Leslie, Zhang et al. 2004). Nop1p is predominantly transported by Kap121p, but if necessary is imported by Kap104p. Thus, the use of an adaptor ensures that Sof1p, an essential protein, is correctly localized to the nucleus (Leslie, Zhang et al. 2004).

Classical NLS

The best characterized NLS is the classical NLS (cNLS), recognized by the Kapβ1 adaptor protein Kapα (Enenkel, Blobel et al. 1995). There are two classes of cNLS. The monopartite cNLS consists of one stretch of basic residues as in the SV40 large T antigen NLS (^{126}PKKKRKV^{132}) (Kalderon, Richardson et al. 1984). The bipartite NLS contains two clusters of basic residues separated by a 10-12 residue linker as in the nucleoplasmin NLS (^{155}KRPAATKKAGQAKKKK^{170}) (Robbins, Dilworth et al. 1991). Structural studies have shown that the monopartite and bipartite cNLSs adopt extended conformations which bind to the same sites on Kapα (Conti, Uy et al. 1998; Conti and Kuriyan 2000; Fontes, Teh et al. 2000; Fontes, Teh et al. 2003).

Thermodynamic analysis of the monopartite cNLSs of SV40 and the *c-myc* proto-oncogene (PAAKRVKLD) has confirmed the suggested consensus sequence of

K(K/R)X(K/R) (Catimel, Teh et al. 2001; Hodel, Corbett et al. 2001). Thermodynamic analysis of the bipartite cNLS is unavailable. Instead, Fontes *et al.* proposed a consensus sequence ($KRX_{10-12}KRRK$) for this cNLS class based on structural data (Fontes, Teh et al. 2003). Bipartite NLSs generally bind Kapα with higher affinity than monopartite NLSs and the addition of a bipartite-like basic cluster can rescue a non-functional SV40 monopartite NLS mutant (Fontes, Teh et al. 2000; Hodel, Corbett et al. 2001). Two peptide inhibitors of this pathway were designed using activity based profiling (Kosugi, Hasebe et al. 2008). The two peptides bind Kap60pΔIBB with affinities of 2.4 pM and 0.02 pM as measured by surface plasmon resonance. The latter is approximately 5 million times tighter than SV40.

The cNLS is considered to be the most utilized NLS in the cell. Using PSORT II to search for cNLSs in yeast proteins known to localize to the nucleus, Lange *et al.* recently predicted that 57% of the 1515 proteins localized to the nucleus at steady state contain a putative cNLS (Huh, Falvo et al. 2003; Lange, Mills et al. 2007). They also predicted that 45% of the 5850 proteins in the *S. cerevisiae* GeneBank database contain a predicted cNLS even though only 25.8% of the proteins in the yeast gfp fusion localization database are localized to the nucleus (Huh, Falvo et al. 2003; Benson, Karsch-Mizrachi et al. 2006; Lange, Mills et al. 2007). The presence of a cNLS consensus sequence in a protein does not guarantee that the protein is transported into the nucleus by the classical pathway. The literature is littered with proteins that contain a putative NLS, but are not mislocalized in *Kap60Δ* cells (Mosammaparast, Jackson et al. 2001; Ueta, Fukunaka et al. 2003; Schaper, Franke et al. 2005).

In order to classify a sequence as a nuclear localization signal, Damelin et al. suggested four criteria (Damelin, Silver et al. 2002). 1) The protein/NLS must bind the import karyopherin and be dissociated in the presence of RanGTP. 2) The NLS must target a non-nuclear protein into the nucleus. 3) The protein must be mislocalized in cells in which the specific import karyopherin has been deleted or mutated. 4) The localization of the protein must be altered if the NLS is deleted or mutated. The field is rigorously trying to identify new substrates of Kapβs. While 57% of nuclear proteins contain a cNLS, >43% of nuclear proteins must be transported by another mechanism (Lange, Mills et al. 2007). NLSs in the latter have not been characterized and are collectively termed non-classical NLSs.

Kap104p/Karyopherin β2 substrates

There are only two known substrates of Kap104p, the mRNA-binding proteins, Nab2p and Hrp1p (Table 1-2) (Aitchison, Blobel et al. 1996). The NLSs of both Nab2p and Hrp1 have been mapped to regions rich in RG repeats, thus it is thought that Kap104p recognizes an rg-NLS (Siomi, Fromont et al. 1998; Truant, Fridell et al. 1998; Lee and Aitchison 1999).

There are twenty mRNA processing proteins that have been experimentally identified as substrates of Kapβ2, the human homolog of Kap104p (Pollard, Michael et al. 1996; Bonifaci, Moroianu et al. 1997; Siomi, Eder et al. 1997; Fan and Steitz 1998; Truant, Kang et al. 1999; Kawamura, Tomozoe et al. 2002; Guttinger, Muhlhausser et al. 2004; Rebane, Aab et al. 2004; Suzuki, Iijima et al. 2005). Only the NLSs of five of

these proteins have been mapped and there is little sequence homology among them. Additionally, the mechanism of NLS recognition by Kapβ2 did not appear conserved between yeast and human homologs.

Import by Kap121p and Kap123p

Kap121p transports a wide range of proteins and by recognizing NLSs enriched in lysine residues (See Table 1-3 and references within). There is not a clear consensus sequence available due to the differences in sequence and length of the NLSs. The NLS of Ste12p was mapped to a 200 residue stretch while other NLSs are much shorter (Leslie, Grill et al. 2002). Unfortunately, mutagenesis studies have only been done on four substrates. Most of these studies identified lysines as important residues: ^{202}KPKKKR^{207}, ^{332}SKRP^{335}, and ^{352}RKPKS^{356} in Aft1p, ^{91}LVKKT^{95} in Spo12p, and ^{7}KR^{8} in Yap1p (Chaves and Blobel 2001; Isoyama, Murayama et al. 2001; Ueta, Fukunaka et al. 2003). In addition to recognizing lysine-rich NLSs, Kap121p binds an rg-NLS in the rRNA processing protein Nop1p, suggesting that there is flexibility in the NLSs recognized by Kap121p and the substrates it transports (Leslie, Zhang et al. 2004). Furthermore, Kap121p is the secondary nuclear import factor of ribosomal proteins and histones (Rout, Blobel et al. 1997; Mosammaparast, Jackson et al. 2001; Mosammaparast, Guo et al. 2002).

Kap123p imports histones and ribosomal proteins into the nucleus as well as other unrelated proteins (see Table 1-4 and references within). Mutagenic analysis suggests that positive residues, and preferentially lysines are important for H3 and H4

recognition by Kap123 (Blackwell, Wilkinson et al. 2007). An alanine substitution for K14 in H3 completely mislocalized H3 NLS to the cytoplasm, while an arginine substitution caused a less severe mislocalization (Blackwell, Wilkinson et al. 2007). In H4, there is not a specific hotspot of binding, but all three lysines (K5,8,12) cumulatively contribute to NLS recognition. If all three lysines are mutated to glutamine, the protein accumulates in the cytoplasm, but mutagenesis to arginine does not affect localization (Blackwell, Wilkinson et al. 2007). This supports their hypothesis that positive charges are important for NLS recognition in H3 and H4. There are no mutagenesis data available for the other substrates of Kap123p to deduce a trend in NLS composition. Timney et al. propose that Kap123p rapidly imports abundant substrates, while Kap121p specializes in more regulated cargos (Timney, Tetenbaum-Novatt et al. 2006).

Import by other yeast Kapβs

The remaining import-Kapβs are less studied and have much fewer known substrates. Kap114p imports the general transcription factor TFIIB as well as proteins involved in chromatin assembly: histones H2A and H2B and the chromatin assembly factor Nap1p (Table 1-5 and references within) (Mosammaparast, Jackson et al. 2001; Mosammaparast, Ewart et al. 2002). The NLSs have been mapped to large fragments of the proteins and there is currently no consensus sequence. Kap111p/Mtr10p imports mRNA binding proteins as well as tRNAs (Table 1-6 and references within). The import karyopherin Kap119p has five known substrates: three transcription factors, one kinase and a heat shock protein (Table 1-7 and references within). Interestingly, four of its

cargos are activated in response to stress or extracellular stimuli. The NLS of one of its substrates, Crz1p, has been mapped and mutated revealing that basic residues are important for its recognition by Kap119p (Polizotto and Cyert 2001). However, the fragment of Ssa4p required for recognition by Kap119p is enriched in hydrophobic and not basic residues (Quan, Rassadi et al. 2004). Kap120p, Kap108p, and Kap122p all have three or fewer cargos which are listed in Table 1-8, Table 1-9 and Table 1-10.

Redundancy among yeast Kapβs

There is significant overlap in the substrates recognized by yeast Kapβs. Since only four yeast karyopherins are essential, it makes sense that there are alternative methods for importing essential proteins into the nucleus. Included in Table 1-2 through Table 1-10 is a column noting whether a substrate is known to be imported by another pathway. Forty-five percent of the substrates listed in the tables have a secondary import pathway. For example, the rg-NLS in Nop1p is also recognized by Kap104p. In the absence of functional Kap121p, Kap104p is capable of transporting Nop1p into the nucleus (Leslie, Zhang et al. 2004). Kap121p can bind the rg-NLS in Nab2p, but does not bind the NLS in Hrp1p (personal observation). Nab2p is mislocalized in *kap104-16* temperature sensitive cells, but perhaps this essential protein utilizes the Kap121p pathway as an alternative import route.

There is redundancy among the substrates of Kap121p and Kap123p (Tables 3 and 4 and references within). Kap123p is the primary import karyopherin of many ribosomal proteins, including Rpl25 (Rout, Blobel et al. 1997). Overexpression of

Kap121p is able to support the import of Rpl25 NLS in the absence of functional Kap123p (Timney, Tetenbaum-Novatt et al. 2006). Timney et al. suggest that the import of Rpl25 by Kap121p may be a remnant of evolutionary diversification of Kapβs (Timney, Tetenbaum-Novatt et al. 2006). While redundancy among Kapβ substrates is common, it remains unknown how this redundancy arose and what are the evolutionary benefits to having multiple transport mechanisms.

Nuclear Import Rates

Many groups have investigated the limiting factor for nuclear import rates. Hodel et al. expanded on work from the Jans lab that suggested that the initial rate of nuclear import of a protein is linearly correlated with the affinity of Kapα for its NLS (Efthymiadis, Shao et al. 1997; Xiao, Jans et al. 1998; Hodel, Harreman et al. 2006). They determined the nucleus to cytoplasm ratio (N:C) for different mutants of the SV40 NLS and the affinity of these mutants for Kapα (Hodel, Harreman et al. 2006). They concluded that, for proteins with an affinity between 10 nM and 1 μM, there was a linear correlation between affinity and steady state localization. However, at affinities weaker than 1 μM, the N:C ratios were indistinguishable from GFP alone. At an affinity tighter than 10 nM, the N:C ratios were not linear, but were broadly distributed with a mean N:C value around five (Hodel, Harreman et al. 2006). Thus, they concluded that the rate of import was dependent on the affinity of the receptor for the NLS.

Timney et al. also came to a similar conclusion. They showed that an increase in K_D between Kap123p and its cargo Rpl25p gave a proportionate decrease in nuclear

import of Rpl25p. Additionally, they showed that the initial rate of import is linearly related to the cytoplasmic concentration of the import Kapβ below concentrations of 15 μM. Since physiological concentration of the two Kapβs they tested, Kap123p and Kap121p, are below this level changes in their concentrations would be linearly related to their initial import rates.

As discussed earlier, Kap123p and Kap121p are both able to import the ribosomal protein Rpl25p but Kap123p is its primary pathway. The affinities of Kap123p and Kap121p for Rpl25 NLS are 82 nM and 94 nM respectively. However, the concentration of Kap123p in the cell is over 5-fold greater than that of Kap121p and they suggest that it is this disparity that is responsible for the order of magnitude difference in their effective import rates (Timney, Tetenbaum-Novatt et al. 2006). When the cytoplasmic concentration of Kap121p was increased to the physiological concentration of Kap123p, the nucleus to cytoplasm ratio of Rpl25p was equivalent to that of Kap123p. Thus, they conclude that the fast import properties of Kap123p are a result of its high cytoplasmic concentration. Finally, they made the unexpected conclusion that the main limiting factor of nuclear import is nonspecific competition with the milieu of proteins in the cytoplasm. The addition of only 10 mg/mL bacterial cytosol, naturally void of any transport proteins, disrupted >90% of Kapβ-NLS complex formation. They suggest two possibilities for the mode of this competition: 1) Kaps may non-specifically bind to proteins in the cytoplasm or 2) basic NLSs may non-specifically bind to cytosolic proteins. They suggest that this could be the cost of certain Kapβs being able to bind NLSs with weak consensus sequences. The conclusion that the rate limiting factor for nuclear import is cytoplasmic competition and not the NPC or the Ran system is contrary

to results from other groups. A study from one group predicted that the concentrations of Kapα and RanGTP limit the rate of nuclear import, while another group concluded that RanGAP is the limiting factor (Smith, Slepchenko et al. 2002; Gorlich, Seewald et al. 2003; Riddick and Macara 2005).

Conclusion

Thousands of yeast proteins need to be imported into the nucleus of the cell by only ten Kapβs. Each Kapβ recognizes a nuclear localization signal in its cargo however there is only one well characterized NLS, the classical NLS. While several substrates have been identified for the remaining karyopherins there is no consensus sequence for any of them. This hinders the identification of new substrates and our broader understanding of nuclear import pathways. X-ray crystallography structures of Kapβs bound to substrate are needed to identify NLS residues involved in Kapβ binding. Due to the significant sequence diversity among the NLSs for each Kapβ, the signals may contain multiple Kapβ binding epitopes.

Table 1-1

Human Karyopherin β	Substrates	Yeast Karyopherin β	Substrates	Essential
Import				
Karyopherin β1	proteins with cNLS via Karyopherin α, snurportin, SREBP-2, HIV Rev, Cyclin B	Kap95p	proteins with cNLS via Kap60p, other proteins	Yes
Karyopherin β2	PY-NLS containing proteins, mRNA binding proteins, PQBP-1	Kap104p	Nab2p, Hrp1p	ts
Transportin SR1	SR proteins	Kap111p/Mtr10p	mRNA binding proteins, tRNA	ts
Transportin SR2	HuR			
Importin 4	Histones, ribosomal proteins	Kap123p	Histones, ribosomal proteins, SRP proteins	No
Importin 5	Histones, ribosomal proteins	Kap121p/Pse1p	Histones, ribosomal proteins, transcription factors, others	Yes
Importin 7	Ribosomal proteins, Glucocorticoid receptors, Smad, ERK	Kap119p/Nmd5p	TFIIS, Ssa4p, Hog1p, Crz1p	No
Importin 8	SRP19, Smad	Kap108p/Sxm1p	Ribosomal proteins, Pab1p, Lhp1p	No
Importin 9	Histones, ribosomal proteins	Kap114p	Histones, TBP, Nap1p, Sua7p	No
Importin 11	UbcM2, rpL12			
		Kap120p	Rpf1p, Ho	No
		Kap122p/Pdr6p	Rnr2p/Rnr4p, TFIIA, Wtm1p	No
Import and Export				
Importin 13	elFIA (export), UBC9 Y14 (import)			
		Msn5p/Kap142p	Pho4p, Crz1p, Cdh1p (export), Replication protein A (import)	no
Export				
Crm1	Proteins wih leucine-rich NES, snurportin, 40S ribosomal subunit, NMD3	Crm1/Xpo1p	Proteins wih leucine-rich NES, Nmd3p, Ltv1p	Yes
CAS	Karyopherin α	Cse1p/Kap109p	Kap60p/Srp1p	Yes
Exportin-t	tRNA	Los1p/Kap127p	tRNA	no
Exportin 4	elF5A			
Exportin 5	Pre-miRNA			
Exportin 6	Actin, profiling			
Exportin 7/RanBP16	P50-RhoGAP			
Uncharacterized				
RanBP6	Unknown			
RanBP17	Unknown			

Modified from (Pemberton and Paschal 2005; Tran, Bolger et al. 2007)

Table 1-2: Kap104p substrates

Substrate	NLS	Direct Binding To Kapβ	RanGTP Dissociation	Mis-localization in Δ or ts strain	NLS imported OR NLS deletion not imported	Protein Function	Viability of Null strain	Other Kapβs	Ref.[a]
Nab2p	201-251	Yes	Yes	Yes	Yes	mRNA binding	Inviable	-	1-4
	DNSQRFTQRGGGAVGKNRRGGRGGNRGGRNNNSTRFNPLAKALGMAGESNM								
Hrp1p	494-534	Yes	Yes	Yes	Yes	mRNA binding	Inviable	-	1,3
	RDRNDRDRDYNHRSGGNGRRNGRGGRGGYNRRNNGYHPYNR								

[a]References: 1) (Aitchison, Blobel et al. 1996), 2) (Truant, Fridell et al. 1998), 3) (Lee and Aitchison 1999), 4) (Siomi, Fromont et al. 1998)

Table 1-3: Kap121p/Pse1p substrates

Substrate	NLS	Direct Binding To Kapβ	RanGTP Dissociation	Mis-localization in Δ or ts strain	NLS imported OR NLS deletion not imported	Protein Function	Viability of Null strain	Other Kapβs	Ref.[a]
Aft1p	198-225 332-365	Yes	Yes	Yes	Yes	iron homeostasis	Viable	-	1
	^{198}TSSIKPKKKRCVSRFNNCPFRVRATYSL^{225} ^{332}SKRPCLPSVNNTGSINTNNVRKPKSQCKNKDTLL^{365}								
Asr1p	243-280	Yes	Yes	Yes	Yes	alcohol-responsive Ring/PHD finger protein	Viable	Kap95p Kap114p Kap123p	2
	IQMHVRRALDRYPLPLLRFKDAYKHVNKQVSRKLYRLS								
Egd1p	11-27	n.d.	n.d.	Yes	Yes	nascent polypeptide associated complex	Viable	Kap60p Kap123p	3
	QKLSANNKVGGTRRKL								
Nop1p	1-90	Yes	No (+DNA?)	Yes	Yes	RNA processing	Inviable	Kap104p	4
	MSFRPGSRGGSRGGSRGGFGGRGGSRGGARGGSRGGFGGRGGSRGGARGGSRGGFGGRGGSRGGARGGSRGGRGGAAGGARGGAKVVIEP								
Nup53p	401-448	Yes	Yes	Yes	Yes	Nucleoporin	Viable	-	5,6
	RNAEFKVSKNSSFKNPRRLEIKDGRSLFLRNRGKIHSGVLSSIESDL								

Table 1-3 continued: Kap121p/Pse1p substrates

Substrate	NLS	Direct Binding To Kapβ	RanGTP Dissociation	Mis-localization in Δ or ts strain	NLS imported OR NLS deletion not imported	Protein Function	Viability of Null strain	Other Kapβs	Ref.[a]
Pdr1p	725-769	n.d.	n.d.	Yes	Yes	Transcription factor	Viable	-	7
	IYKSWTDMNKILLDFDNDYSVYRSFAHYSISCIILVSQAFSVAE								
Pho4p	140-166	Yes	Yes	Yes	Yes	Transcription factor	Viable	-	8
	ANKVTKNKSNSSPYLNKRRGKPGPDS								
Sas2p	1-48	n.d.	n.d.	Decreased	Yes	histone acetyltransferase complex	Viable	Kap123p	9
	MARSLSQSLTATTQKLKGKKNGGKGKNKPSAKIKKTQKEMLYGILNER								
Sof1p	381-489	Yes	n.d.	Yes	Yes	RNA processing	Inviable	Kap104p via Nop1p	4
	[411]KRISRHRHVPQVIKKAQEIKNIELSSIKRREANERRRKD[450] (minimum for Kap121 recognition)								
Spo12p	76-130	Yes	n.d.	Yes	Yes	Mitosis and meiosis	Viable	-	10
	KKSTSNLKSSHTTSNLVKKTMFKRDLLKQDPKRKLQLQQRFASPTDRLVSPCSLKLNEHKVKMFGKKK								
Ste12p	494-698	Yes	Yes	Yes	Yes	Transcription factor	Viable	-	11
	NNMLYPQTATSWNVLPPQAMQPAPTYVGRPYTPNYRSTPGSAMFPYMQSSNSMQWNTAVSPYSSRAPSTTAKNYPPSTFYSQNIN QYPRRRTVGMKSSQGNVPTGNKQSVGKSAKISKPLHIKTSAYQKQYKINLETKARPSAGDEDSAHPDKNKEISMPTPDSNTLVVQS EEGGAHSLEVDTNRRSDKNLPDAT								
Yap1p	5-59	Yes	Yes	Yes	Yes	Transcription factor	Viable	Kap123p	12
	TAKRSLDVVSPGSLAEFEGSKSRHDEIENEHRRTGTRDGEDSEQPKKKGSKTSKK								
Yra1p	1-77	Yes	Yes	Mostly	Yes	RNA binding	Inviable	Kap123p	13
	MSANLDKSLDEIIGSNKAGSNRARVGGTRGNGPRRVGKQVGSQRRSLPNRRGPIRKNTRAPPNAVARVAKLLDTTRE								
Secondary import karyopherin of histones (H2A, H2B, H3, H4) (Mosammaparast, Jackson et al. 2001; Mosammaparast, Guo et al. 2002) and ribosomal proteins (Rp10a, Rpl25p…) (Rout, Blobel et al. 1997)									

[a]References: 1) (Ueta, Fukunaka et al. 2003), 2) (Fries, Betz et al. 2007), 3) (Franke, Reimann et al. 2001), 4) (Leslie, Zhang et al. 2004), 5) (Marelli, Lusk et al. 2001), 6) (Lusk, Makhnevych et al. 2002), 7) (Delahodde, Pandjaitan et al. 2001), 8) (Kaffman, Rank et al. 1998), 9) (Schaper, Franke et al. 2005), 10) (Chaves and Blobel 2001), 11) (Leslie, Grill et al. 2002), 12) (Isoyama, Murayama et al. 2001), 13) (Zenklusen, Vinciguerra et al. 2001)

Table 1-4: Kap123p substrates

Substrate	NLS	Direct Binding To Kapβ	RanGTP Dissociation	Mis-localization in Δ or ts strain	NLS imported OR NLS deletion not imported	Protein Function	Viability of Null strain	Other Kapβs	Ref.[a]
Egd1p	11-27	n.d.	n.d.	Yes	Yes	nascent polypeptide associated complex	Viable	Kap60p Kap121p	1
	QKLSANNKVGGTRRKL								
H3 (Hht2p)	1-28	Yes	n.d.	Yes	Yes	Histone protein	Viable	Kap121p Kap119p Kap108p	2-3
	MARTKQTARKSTGGKAPRKQLASKAARK								
H4 (Hhf2p)	1-42	Yes	n.d.	Decreased	Yes	Histone protein	Viable	Kap121p Kap119p Kap108p	2-3
	MSGRGKGGKGLGKGGAKRHRKILRDNIQGITKPAIRRLARRG								
Ho	286-302	Yes	Yes	Yes	Yes	yeast mating switch endonuclease	Viable	Kap121p Kap108p	4
	RKNNPFWKAVILKFKR								
Rpl25p	1-41	Yes	n.d.	Yes	Yes	Ribosomal protein	Viable	Kap121p	5
	MAPSAKATAAKKAVVKGTNGKKALKVRTSATFRLPKTLKLARAPKYASKAVPHYNRLDSYKV								
Sas2p	1-48	n.d.	n.d.	Decreased	Yes	histone acetyltransferase complex	Viable	Kap121p	6
	MARSLSQSLTATTQKLKGKKNGGKGKNKPSAKIKKTQKEMLYGILNER								
SRP (srp14p, srp21p, srp68p, srp72p)	n.d.	n.d.	n.d.	Only in 121/123 double mutant	n.d.	targeting of secretory and membrane proteins to the ER membrane	Viable	Kap121p	7
Other ribosomal proteins which are substrates, but whose NLS have not been mapped (Rp10a, Rps1p, Rpl4p, Rpl15p, Rpl16p, Rpl18p,Rpl25p, and Rpl41p) (Rout, Blobel et al. 1997)									

[a]References: 1) (Franke, Reimann et al. 2001), 2) (Mosammaparast, Guo et al. 2002), 3) (Blackwell, Wilkinson et al. 2007), 4) (Bakhrat, Baranes et al. 2006), 5) (Rout, Blobel et al. 1997), 6) (Schaper, Franke et al. 2005), 7) (Grosshans, Deinert et al. 2001)

Table 1-5: Kap114p substrates

Substrate	NLS	Direct Binding To Kapβ	RanGTP Dissociation	Mis-localization in Δ or ts strain	NLS imported OR NLS deletion not imported	Protein Function	Viability of Null strain	Other Kapβs	Ref.[a]
H2A (Hta1p)	1-46	Yes	Yes	Decreased	Yes	Histone Protein	Viable	Kap121p Kap123p Kap114p Kap104p	1,2
	MSGGKGGKAGSAAKASQSRSAKAGLTFPVGRVHRLLRRGNYAQRIG								
H2B (Htb1p)	1-52 (minimal: 21-33)	Yes	Yes	Decreased	Yes	Histone protein	Inviable	Kap121p	1
	MSAKAEKKPASKAPAEKKPAAKKTSTSTDGKKRSKARKETYSSYIYKVLKQT								
Nap1	200-362	Yes	NO	Yes	Yes	chromatin assembly factors	Viable	-	3,4
	DTITDRDAEVLEYLQDIGLEYLTDGRPGFKLLFRFDSSANPFFTNDILCKTYFYQKELGYSGDFIYDHAEGCEISWKDNAHNVTVDL EMRKQRNKTTKQVRTIEKITPIESFFNFFDPPKIQNEDQDEELEEDLEERLALDYSIGEQLKDKLIPRAVDWFTG								
Spt15p (TBP)	n.d.	Yes	Yes (DNA increases dissociation)	Yes	n.d.	TATA binding protein	Inviable	Kap121p Kap123p Kap95p	5,6
Sua7p (TFIIB)	101-345	Yes	Yes	Yes	Yes	Preinitiation complex	Inviable	-	7
	TTDMRFTKELNKAQGKNVMDKKDNEVQAAFAKITMLCDAAELPKIVKDCAKEAYKLCHDEKTLKGKSMESIMAASILIGCRRAEV ARTFKEIQSLIHVKTKEFGKTLNIMKNILRGKSEDGFLKIDTDNMSGAQNLTYIPRFCSHLGLPMQVTTSAEYTAKKCKEIKEIAGKSP ITIAVVSIYLNILLFQIPITAAKVGQTLQVTEGTIKSGYKILYEHRDKLVDPQLIANGVVSLDNLPGVEKK								

[a]References: 1) (Mosammaparast, Jackson et al. 2001), 2) (Greiner, Caesar et al. 2004), 3) (Mosammaparast, Del Rosario et al. 2005), 4) (Mosammaparast, Ewart et al. 2002), 5) (Morehouse, Buratowski et al. 1999), 6) (Pemberton, Rosenblum et al. 1999), 7) (Hodges, Leslie et al. 2005)

Table 1-6: Kap111p/Mtr10p substrates

Substrate	NLS	Direct Binding To Kapβ	RanGTP Dissociation	Mis-localization in Δ or ts strain	NLS imported OR NLS deletion not imported	Protein Function	Viability of Null strain	Other Kapβs	Ref.[a]
Gbp2p	N-terminus	n.d.	n.d.	Yes	n.d.	mRNA binding	Viable	-	1
Hrb1p	n.d.	n.d.	n.d.	Yes	n.d.	mRNA binding	Viable	-	2
Npl3p	283-414 (397-414 necessary)	Yes	Yes (+RNA)	Yes	Yes	mRNA binding	Viable	-	3,4
	NRGGFRGRGGFRGGFRGGFRGGFSRGGFGGPRGGFGGPRGGYGGYSRGGYGGYSRGGYGGSRGGYDSPRGGYDSPRGGYSRGGYGGPRNDYGPPRGSYGGSRGGYDGPRGDYGPPRDAYRTRDAPRERSPTR								
tRNAs (Shaheen and Hopper 2005)									

[a]References: 1) (Windgassen and Krebber 2003), 2) (Hacker and Krebber 2004), 3) (Senger, Simos et al. 1998), 4) (Pemberton, Rosenblum et al. 1997)

Table 1-7: Kap120p substrates

Substrate	NLS	Direct Binding To Kapβ	RanGTP Dissociation	Mis-localization in Δ or ts strain	NLS imported OR NLS deletion not imported	Protein Function	Viability of Null strain	Other Kapβs	Ref.[a]
Ho	480-492	n.d.	n.d.	Yes	Yes	yeast mating switch endonuclease	Viable	-	1
	KKNLKHCVACPRK								
Rpf1p	n.d.	Yes	Yes	Only in triple mutant	n.d.	Ribosome biogenesis	Inviable	Kap119p Kap114p	2

[a]References: 1) (Bakhrat, Baranes et al. 2006), 2) (Caesar, Greiner et al. 2006)

Table 1-8: Kap119p/Nmd5p substrates

Kap119p/Nmd5p									
Substrate	**NLS**	**Direct Binding To Kapβ**	**RanGTP Dissociation**	**Mis-localization in Δ or ts strain**	**NLS imported OR NLS deletion not imported**	**Protein Function**	**Viability of Null strain**	**Other Kapβs**	**Ref.[a]**
Crz1p	394-422	Yes	Yes	Yes	Yes	Transcription Factor in response to stress	Viable	-	1
	IINGRKLKLKKSRRRSSQTSNNSFTSRRS								
Gal4p	1-147	n.d.	n.d.	Yes	Yes	Transcription factor	Viable	Kap95p	2,3
	MKLLSSIEQACDICRLKKLKCSKEKPKCAKCLKNNWECRYSPKTKRSPLTRAHLTEVESRLERLEQLFLLIFPREDLDMILKMDSLQ DIKALLTGLFVQDNVNKDAVTDRLASVETDMPLTLRQHRISATSSSEESSNKGQRQLTV								
Hog1p	n.d.	n.d.	n.d.	Yes	n.d.	MAP kinase	Viable	-	4
Ssa4p	1-236	n.d.	n.d.	Yes	Yes	Heat shock protein	Viable	-	5
	[162]IAGLNVLRII[171] required								
TFIIS (Dst1p)	n.d.	Yes	Yes	Yes	n.d.	Transcription elongation factor	Viable	-	6

[a]References: 1) (Polizotto and Cyert 2001), 2) (Nikolaev, Cochet et al. 2003), 3) (Chan, Hubner et al. 1998), 4) (Ferrigno, Posas et al. 1998), 5) (Quan, Rassadi et al. 2004), 6) (Albertini, Pemberton et al. 1998)

Table 1-9: Kap108p/Sxm1p substrates

Substrate	NLS	Direct Binding To Kapβ	RanGTP Dissociation	Mis-localization in Δ or ts strain	NLS imported OR NLS deletion not imported	Protein Function	Viability of Null strain	Other Kapβs	Ref.[a]
Lhp1p	112-224	Yes	n.d.	Yes	n.d.	maturation of pre-tRNA	Viable	-	1,2
	AARNARIEQNQRTLAVMNFPHEDVEASQIPELQENLEAFFKKLGEINQVRLRRDHRNKKFNGTVLVEFKTIPECEAFLKSYSNDDESNEILSYEGKKLSVLTKKQFDLQREAS								
Pab1p	281-338	Yes	Yes	Yes	Yes	major poly(A)-binding protein	Inviable	-	3
	DSELNGEKLYVGRAQKKNERMHVLKKQYEAYRLEKMAKYQGVNLFVKNLDDSVDDEKL								
Ribosomal proteins (Rpl16p, Rpl25p, Rpl34p) (Rosenblum, Pemberton et al. 1997)									

[a]References: 1) (Rosenblum, Pemberton et al. 1997), 2) (Rosenblum, Pemberton et al. 1998), 3) (Brune, Munchel et al. 2005)

Table 1-10: Kap122p/Pdr6p substrates

Kap122p/Pdr6p									
Substrate	**NLS**	**Direct Binding To Kapβ**	**RanGTP Dissociation**	**Mis-localization in Δ or ts strain**	**NLS imported OR NLS deletion not imported**	**Protein Function**	**Viability of Null strain**	**Other Kapβs**	**Ref.[a]**
Rnr2p/ Rnr4p	n.d.	NO (Wtm1 may act as adaptor)	n.d.	Yes	n.d.	Ribonucleotide reductase	Viable	-	1
Toa1p/ Toa2p (TFIIA)	n.d.	Yes	Yes	Yes	n.d.	Transcription factor	Inviable	-	2
Wtm1p	n.d.	Yes (In vivo)	n.d.	Yes	n.d.	Transcriptional modulator	Viable	-	1

[a]References: 1) (Zhang, An et al. 2006), 2) (Titov and Blobel 1999)

CHAPTER 2

RULES FOR NUCLEAR LOCALIZATION SEQUENCE RECOGNITION BY KARYOPHERIN β2[a]

Abstract

Karyopherin β (Kapβ) proteins bind nuclear localization and export signals (NLSs and NESs) to mediate nucleocytoplasmic trafficking, a process regulated by Ran GTPase through its nucleotide cycle. Diversity and complexity of signals recognized by Kapβs have prevented prediction of new Kapβ substrates. The structure of Kapβ2 bound to one of its substrates, the NLS of hnRNP A1, that we report here explains the mechanism of substrate displacement by Ran GTPase. Further analyses reveal three rules for NLS recognition by Kapβ2: NLSs are structurally disordered in free substrates, have overall basic character, and possess a central hydrophobic or basic motif followed by a C-terminal R/H/K$X_{(2-5)}$PY consensus sequence. We demonstrate the predictive nature of these rules by identifying NLSs in seven previously known Kapβ2 substrates and uncovering 81 new candidate substrates, confirming five experimentally. These studies define and validate a new NLS that could not be predicted by primary sequence analysis alone.

[a]Originally published in Cell **126**(3): 543-58. Copyright by Elsevier. License to reproduce: 2103720076988.

Introduction

Karyopherin β proteins (Kapβs; also known as Importins and Exportins) are responsible for the majority of nucleocytoplasmic transport in the cell. At least 20 members of the Kapβ family have been identified in humans. Kapβs bind specific sets of transport substrates and target them to the nuclear pore complex. The Ran GTPase regulates Kapβ-substrate interactions and transport directionality through its nucleotide cycle (Gorlich and Kutay 1999; Chook and Blobel 2001; Conti and Izaurralde 2001; Weis 2003). RanGTP is concentrated in the nucleus, while RanGDP is concentrated in the cytoplasm. In import pathways, RanGTP and substrates bind Kapβs competitively, allowing substrate binding in the cytoplasm and RanGTP-mediated release in the nucleus. In contrast, in export pathways, RanGTP, substrates and Kapβs bind cooperatively, resulting in substrate binding in the nucleus and release in the cytoplasm as Ran-bound nucleotide is hydrolyzed.

In humans, 10 import-Kapβs have been shown to carry a diverse set of macromolecular substrates into the nucleus (Mosammaparast and Pemberton 2004). Despite significant efforts, only a few substrates have been identified for most import-Kapβs, and large panels of substrates have been identified for only two pathways, those of Kapβ1 and Kapβ2 (see below). Each import-Kapβ appears to bind distinct sets of substrates, suggesting that each Kapβ recognizes a different nuclear localization signal(s) or NLS(s). However, large sequence diversity among various substrates has prevented

identification of NLSs for most Kapβs, and it remains extremely difficult to predict NLSs in candidate import substrates.

The classical-NLSs are short, lysine-rich sequences that bind the adaptor protein Kapα, which forms a heterodimer with Kapβ1, which in turn mediates nuclear import (Conti and Izaurralde 2001). Most other proteins imported into the nucleus do not utilize such an adaptor, but rather bind directly to a Kapβ. The few characterized NLSs that bind directly to Kapβs are diverse, encompassing both structural domains and linear epitopes. For example, crystal structures of three Kapβ1-substrate complexes show structurally diverse substrates binding at different sites on the karyopherin (Cingolani, Petosa et al. 1999; Cingolani, Bednenko et al. 2002; Lee, Sekimoto et al. 2003). Furthermore, most proteins that bind Kapβ1 show little sequence or structural homology and thus general features among substrates in this pathway cannot be inferred at this time.

In another import pathway, more than 20 mRNA processing proteins (including hnRNPs A1, D, F, M, HuR, DDX3, Y-box binding protein 1 and TAP) have been identified as import substrates of Kapβ2 (Pollard, Michael et al. 1996; Bonifaci, Moroianu et al. 1997; Siomi, Eder et al. 1997; Fan and Steitz 1998; Truant, Kang et al. 1999; Kawamura, Tomozoe et al. 2002; Guttinger, Muhlhausser et al. 2004; Rebane, Aab et al. 2004; Suzuki, Iijima et al. 2005). Kapβ2 binds its best-characterized substrate, splicing factor hnRNP A1, through the 38-residue M9 sequence (Pollard, Michael et al. 1996; Bonifaci, Moroianu et al. 1997) that we will refer to as M9NLS. Many studies have shown that the M9NLS peptide is both necessary and sufficient for nuclear import mediated by Kapβ2 (Siomi and Dreyfuss 1995; Weighardt, Biamonti et al. 1995). Other

than hnRNP A1, only NLSs in HuR (Fan and Steitz 1998), TAP (Truant, Kang et al. 1999), hnRNP D and its homologs, the JKTPB proteins (Kawamura, Tomozoe et al. 2002; Suzuki, Iijima et al. 2005), have been characterized. The NLSs of hnRNP D and HuR show marginal sequence homology to M9NLS, that of TAP shares no sequence homology with M9NLS and none of the other Kapβ2 substrates contain obvious M9NLS-like sequences. Like the Kapβ1 system, the diversity of substrates and known NLSs in Kapβ2 has also prevented prediction of NLSs in this pathway.

In order to understand the mechanism of substrate recognition and distill the critical elements for NLS recognition by Kapβ2 we have determined the structure of Kapβ2 bound to the M9NLS of hnRNP A1. The structure and complementary biochemical studies reveal a set of rules for NLS recognition by Kapβ2: NLSs imported by Kapβ2 should occur within large (>30-residue) structurally disordered elements, have overall basic character and contain a set of consensus sequences. These rules are predictive and have allowed us to identify and biochemically confirm NLSs in seven known Kapβ2 substrates. Most importantly, we used these NLS rules in a bioinformatics approach, and identified 81 new candidate import substrates for Kapβ2. Finally, we have confirmed that five of these bind Kapβ2 through the predicted NLS in a Ran-dependent manner.

Materials and Methods

Kapβ2Δloop cloning

In the crystallographic studies Kapβ2 residues 337-367 were replaced with a GGSGGSG linker. The deletion of the loop was accomplished through two rounds of polymerase chain reaction (PCR). The first round of PCR used an extended 3' primer to clone the N-terminal half of Kapβ2 with a GGSGGGSG addition after residue 338. The second round of PCR added GGSGGSG 5' of residue 368 to the C-terminal half of Kapβ2. In the coding region of GGSGGSG is a BamHI site. This site was used to ligate the two pieces together. The mutated Kapβ2 gene was then ligated into the pGexTev vector.

Protein expression, purification and complex formation

Protein expression, purification, and complex formation were done by Brittany Lee in the Chook lab. Kapβ2 protein was expressed in *E. coli* BL21 (DE3) as a GST-fusion from pGEX-Tev vector and purified as previously reported (Chook and Blobel 1999; Chook, Jung et al. 2002). M9NLS was expressed in *E. coli* as a GST-fusion of hnRNP A1 residues 257-305, and purified as previously described (Chook, Jung et al. 2002). Two-fold molar excess of GST-M9NLS was added to purified Kapβ2, cleaved with Tev protease and the complex purified by gel filtration chromatography. Selenomethionine-Kapβ2 and selenomethionine-M9NLS were purified and assembled as for the native proteins. All complexes were concentrated to 25 mg/ml for crystallization.

Crystallization, data collection and structure determination

Native Kapβ2-M9NLS complex was crystallized by vapor diffusion (reservoir solution: 40 mM MES pH 6.5, 3M potassium formate and 10% glycerol) and flash frozen in liquid propane. These crystals diffracted at best to 3.5 Å. However, soaking the crystals in crystallization solution containing 0.7 mM of a 12-residue FXFG-peptide (sequence: TGGFTFGTAKTA) improved diffraction to 3.05 Å (Figure 2-1 and Figure 2-2). Data from an FXFG-soaked crystal was collected on the X-ray Operations and Research beamline 19-ID at the Advanced Photon Source, Argonne National Laboratory and processed using HKL2000 (Table 2-1) (Otwinowski and Minor 1997). Crystals of the selenomethionine complex were also obtained by vapor diffusion (reservoir solution: 0.1M Tris 8.0, 3M potassium formate and 15% glycerol), soaked in FXFG-peptide and diffracted to 3.3 Å by Brittany Lee. Single-wavelength anomalous dispersion (SAD) data was collected on SBC-19-ID (Table S1) and processed with HKL2000 (Otwinowski and Minor 1997).

The structure of Kapβ2-M9NLS was solved by Yuh Min Chook. Native Kapβ2-M9NLS crystals (space group C2, unit cell parameters a=152.0 Å, b=154.1 Å, c=141.7 Å and β=91.7°) contain two complexes in the asymmetric unit. Selenomethionine Kapβ2-M9NLS also crystallized space group C2, but has a significantly different unit cell length in its a axis (unit cell parameters: a=155.6 Å, b=154.6 Å, c=141.6 Å and β=91.6°; Table S1). Native Patterson maps indicate that the two complexes in the asymmetric unit are related by pseudo-translation along the crystallographic c axis. Molecular replacement

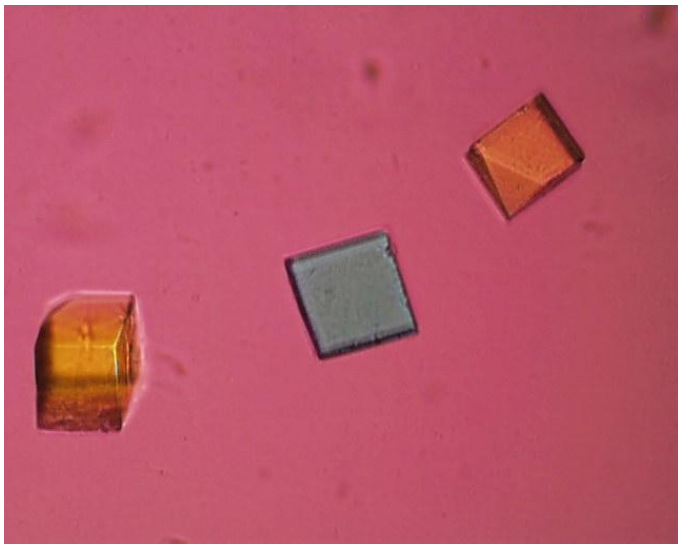


Figure 2-1: Crystals of Kapβ2-M9 NLS soaked in FxFG peptide.

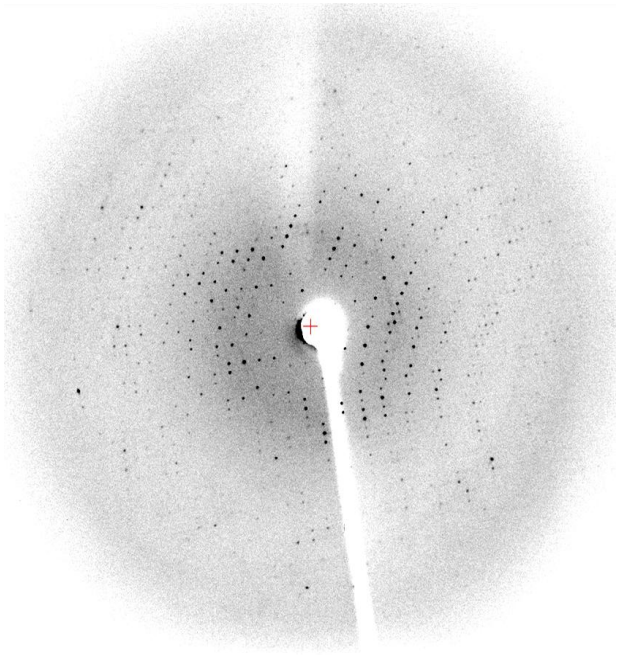


Figure 2-2: Diffraction pattern of a Kapβ2-M9 NLS crystal soaked in FxFG peptide. Data collected on R-Axis IV at UT Southwestern.

Table 2-1

Kap β2-M9NLS Complex

Data Collection:

Native:		Selenomethionine:	
Resolution 100.00 - 3.05 Å		Resolution 100.00 – 3.30 Å	
Space group C2		Space group C2	
a=152.01 Å, b=154.09 Å, c=141.67 Å, β=91.75°		a=155.65 Å, b=154.59 Å, c=141.56 Å, β=91.56°	
[a]R_{sym}	0.055 (0.429)[b]	[a]R_{sym}	0.103 (0.500)[b]
I/σ	24.7 (2.0)[b]	I/σ	21.5 (2.1)[b]
Redundancy	4.6 (4.1)[b]	Redundancy	4.9 (4.7)[b]
Completeness	99.0% (92.8%)[b]	Completeness	98.5% (91.5%)[b]

Refinement:

Resolution 100.00 - 3.05 Å

[c]R_{factor} = 0.242 R_{free} = 0.272

rmsd from ideal bond lengths 0.0074 Å

rmsd from ideal bond angles 1.136°

Ramachandran Plot: 90.4% in most favored regions, 9.6% in allowed regions

Model:

	Residues	Average B factor
Complex 1:		
Kap β2 Chain A	6-36, 44-77, 80-319, 368-890	72.7 Å^2
M9NLS Chain C	263-289	81.9 Å^2
Complex 2:		
Kap β2 Chain B	6-36, 44-55, 59-75, 80-319, 368-890	74.4 Å^2
M9NLS Chain D	266-289	77.6 Å^2

[a]Rsym = $\sum_h \sum_i | (I_i(h) — <I(h)> | / \sum_h \sum_i I_i(h)$; $I_i(h)$ is the i-th measurement of reflection h and <I(h)> is the weighted mean of all measurements of h.

[b]Values in parentheses are calculated for data in the highest resolution shell

[c]R factor = $\sum_h | |F_{obs}(h)| - |F_{calc}(h)| | / \sum_h F_{obs}(h) |$; R_{free} is calculated with 10% of the data.

trials using the Kapβ2-Ran structure were unsuccessful but SAD phasing followed by solvent flipping, both using the program CNS produced interpretable electron density maps (Brunger, D. et al. 1998). A model comprising 90% of Kapβ2 was built using O (Jones, Cowan et al. 1991) but electron density for the substrate remained uninterpretable even though M9NLS residue M276 could be clearly placed using a selenium site. The partial SAD-phased model was used as a search model for molecular replacement using the program Phaser with the higher resolution native dataset (McCoy, Grosse-Kunstleve et al. 2005). Positional refinement using REFMAC5 (CCP4 1994) followed by solvent flipping using CNS (Brunger, D. et al. 1998) yielded electron density maps that allowed 97% of Kapβ2 to be built. The density was further improved by rigid body, positional and simulated annealing refinement of Kapβ2 alone, using the programs CNS (Brunger, D. et al. 1998). The Fo-Fc map plotted at 2.5 sigma clearly showed strong density for M9NLS residues 267-289 in the complex I, and residues 263-289 in complex II. Even though soaking the crystals in FXFG peptide improved diffraction, no density was observed for the FXFG peptide. The final refined model shows good stereochemistry with R_{factor} of 24.2% and R_{free} of 27.2%.

NLS-mapping, site directed mutagenesis and Kap β2 binding assays

Brittany Lee, a research associate in the Chook lab, performed most of the following experiments. cDNA for hnRNPs F, M, PQBP-1, EWS, SAM68, HMBA-inducible protein, YBP1, FUS, DDX3, Clk3, Sox14 and WBS16 were obtained from Open Biosystems. cDNA for HCC1 and RB15B were obtain by PCR from a human fetal thymus cDNA library (Clontech). The full-length proteins as well as fragments were sub-

cloned using PCR into pGEX-Tev vector. Expression constructs for NLSs of cyclin T1 and CPSF6 were generated using synthetic complementary oligonucleotides coding for the 28-mer peptides. Single, double and triple mutations to alanine residues were performed using the Quickchange method (Stratagene), and all constructs were confirmed by nucleotide sequencing. Substrate proteins were expressed in *E. coli* BL21 (DE3) cells. GST-M9NLS was expressed at 37°C, GST- Kapβ2 was expressed at 30°C and the other substrates were expressed at 25°C, and all were purified using glutathione sepharose (GE Healthcare).

In each binding reaction involving new NLSs, mutant NLSs and new Kapβ2 substrates, approximately 18 μg of Kapβ2 were added to 5-10 μg of GST-substrate immobilized on glutathione sepharose followed by extensive washing of the beads with buffer containing 20 mM Hepes pH 7.3, 110 mM potassium acetate, 2 mM DTT, 1 mM EGTA, 2 mM Magnesium acetate and 20% glycerol. Immobilized proteins were visualized using SDS-PAGE and Coomassie Blue staining. 3-5 fold molar excess of RanGTP (compared to Kapβ2) is also used in some binding assays.

Binding assays involving mutants of Kapβ2 were performed similarly, with each reaction using approximately 10 μg of MBP-M9NLS added to 5-10 μg of GST-Kapβ2.

Quantitation of binding affinity with ITC

ITC experiments were done in tandem with Ertugrul Cansizoglu, a graduate student in the Chook lab. Binding affinities of wild type and mutant MBP-M9NLS to Kapβ2 were quantitated using ITC. The ITC experiments were done using a MicroCal Omega VP-ITC calorimeter (MicroCal Inc., Northampton, MA). Proteins were dialyzed

against buffer containing 20 mM Tris pH 7.5, 100 mM NaCl and 2 mM β-mercaptoethanol. 100-500 μM Wild type and mutant MBP-M9NLS proteins were titrated into a sample cell containing 10-100 μM full length Kapβ2. Most ITC experiments were done at 20°C with 35 rounds of 8 μl injections. ITC experiments involving wild type M9NLS were similar, but with 56 rounds of 5 μl injections. Data was plotted and analyzed using MicroCal Origin software version 7.0, with a single binding site model.

Bioinformatics search for new Kapβ2 substrates

The search for new substrates was done by Yuh Min Chook. Candidate Kapβ2 substrates were identified by the program ScanProsite (Gattiker, Gasteiger et al. 2002) using motifs ϕ_1-G/A/S-ϕ_3-ϕ_4-X_{7-12}-R/K/H-X_{2-5}-P-Y (where ϕ_1 is strictly hydrophobic, ϕ_3 and ϕ_4 are hydrophobic and also includes long aliphatic sidechains R and K), K/R-X_{0-2}-K/R-K/R-X_{3-10}-R/K/H-X_{1-5}-P-Y and human proteins in the UniProtKB/Swiss-Prot protein database (Bairoch, Boeckmann et al. 2004). All resulting entries were filtered for structural disorder using the program DisEMBL (Linding, Jensen et al. 2003) and for positively charged NLS segments of 50 amino acids (beginning 40 residues N-terminus of the PY to 10 residues C-terminus of that motif). Proteins with potential PY-NLSs that are found in transmembrane proteins and those that occur within identified domains were eliminated from the list even though some NLSs may occur in long loops within folded domains.

Results

Kapβ2-M9NLS complex: Structure overview

Kapβ2 is a superhelical protein with 20 HEAT repeats. It is almost exclusively α–helical except for a 62-residue loop in repeat 8 (H8 loop; Figure 2-3A). Each repeat consists of two antiparallel helices A and B, each lining the convex and concave side of the superhelix respectively (Chook and Blobel 1999; Chook, Jung et al. 2002). The Kapβ2-M9NLS crystals contain a Kapβ2 mutant with a truncated H8 loop bound to residues 257 to 305 of hnRNP A1 (Figure 2-3B). Biochemical studies showed that the loop neither hinders nor is necessary for substrate binding. However, it is sensitive to proteolytic degradation in substrate-bound Kapβ2, suggesting structurally flexibility (Chook, Jung et al. 2002). In the final Kapβ2 construct, the H8 loop was truncated (a GGSGGSG linker replaces residues 337-367) to minimize disorder in the crystal. The Kapβ2-M9NLS crystal structure was solved to 3.05 Å resolution (Table 2-1, PDB ID code 2H4M).

The asymmetric unit of the crystal contains two Kapβ2-M9NLS complexes (I and II). All residues in both Kapβ2s are modeled except for three short loops at the N-termini, H8 loop residues 320-337 and the engineered GGSGGSG H8 loop linker (disordered regions are indicated by dashes in Figure 2-3A). Substrate residues 267-289 are observed in complex I, while additional substrate residues 263-266 are modeled in complex II (Figure 2-3C). Thus, the latter is used in structural analysis and discussion

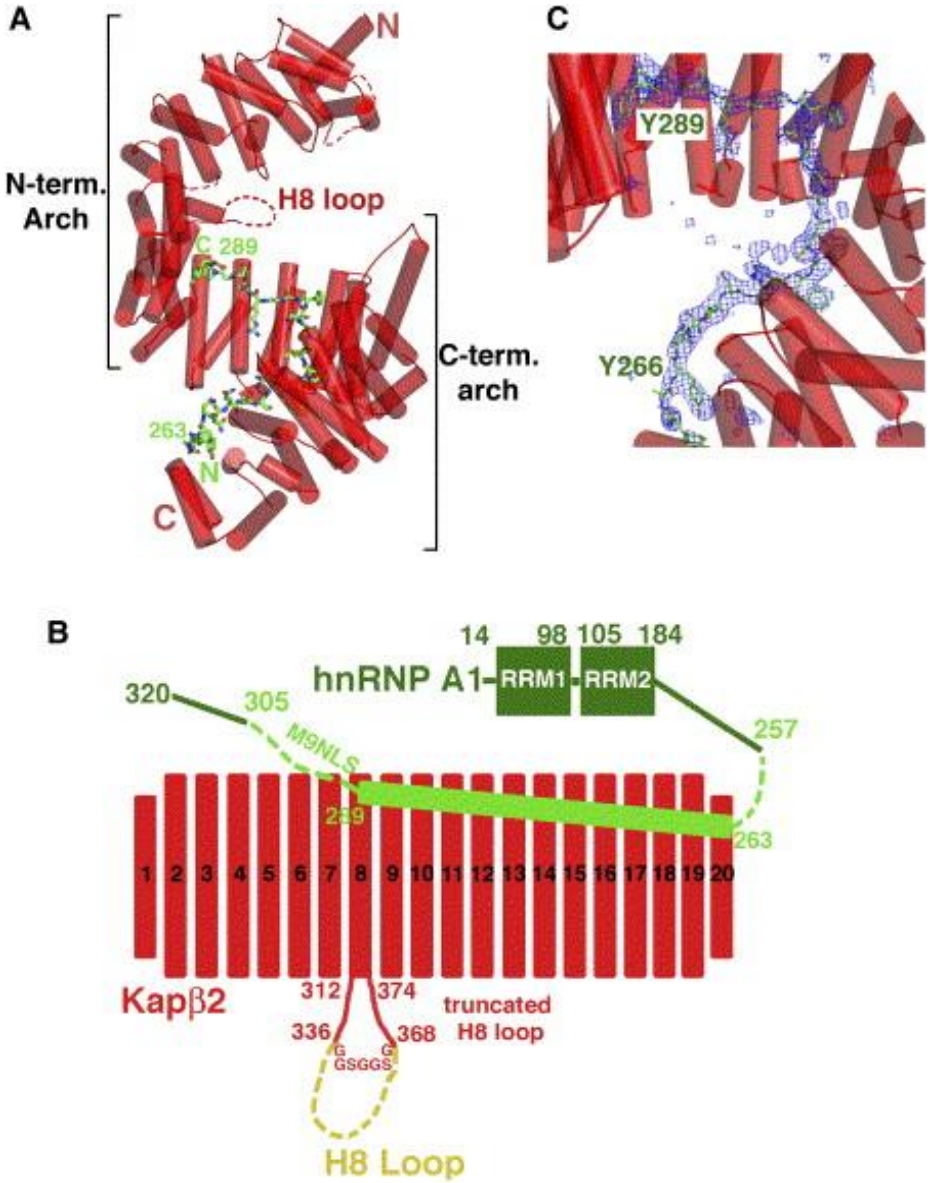


Figure 2-3: Crystal structure of the Kapβ2-M9NLS complex. (A) Ribbon diagram of the Kapβ2-M9NLS complex with Kapβ2 in red (α-helices represented as cylinders and structurally disordered loops as red dashes) and M9NLS shown as a stick figure (carbon - green, oxygen - red, nitrogen - blue and sulfur - orange). (B) The 20 HEAT repeats and H8 loop of Kapβ2 used in structural analyses (red), and M9NLS (light green) within hnRNP A1 (green). The deleted portion of the H8 loop is in yellow. (C) The M9NLS binding site with Fo-Fc map (2.5σ) calculated using Kapβ2 alone (blue mesh), drawn with PYMOL (DeLano 2002).

below. HEAT repeats 5-20 share similar conformations in both complexes (rmsd 1.7 Å). In contrast, HEAT repeats 1-4 diverge to a distance of 7 Å at their N-termini with high average B-factors (93 $Å^2$ for complex I and 118 $Å^2$ for complex II), suggesting inherent conformational flexibility in this region of Kapβ2.

The 20 HEAT repeats of the Kapβ2-M9NLS complex form an almost perfect superhelix (pitch ~72 Å, diameter ~60 Å and length ~111 Å; Figure 2-3A). The superhelix can also be described as two overlapping arches, with the N-terminal arch spanning HEAT repeats 1-13 and the C-terminal arch spanning repeats 8-20. In the Kapβ2-Ran complex, RanGTP binds in the N-terminal arch (Chook and Blobel 1999). Here, we observe that M9NLS binds in the C-terminal arch (Figure 2-3 A and C).

The Kapβ2-M9NLS binding interface

M9NLS binds in extended conformation to line the concave surface of C-terminal arch of Kapβ2 (Figure 2-3A). Its peptide direction is antiparallel to that of the karyopherin superhelix, and substrate buries 3432 $Å^2$ of surface area in both binding partners. Tracing M9NLS from N- to C-terminus, residues 263-266 interact with helices H18A, H19A and H20B of Kapβ2 while residues 267-269 drape over the intra-HEAT 18 loop into the C-terminal arch of the karyopherin. The rest of M9NLS follows the curvature of the C-terminal arch to contact B helices of repeats 8-17 (Figure 2-3A and Figure 2-4A). The substrate interface on Kapβ2 comprises ~30% of the concave surface of the C-terminal arch, which is relatively flat and devoid of deep pockets or grooves.

Most of this surface, which includes the M9NLS interface, is also highly acidic (Figure 2-4B).

M9NLS forms an extensive network of polar and hydrophobic interactions with Kapβ2, involving both the main chain and sidechains of the substrate (Figure 2-4A). Most of the substrate interface on Kapβ2 is acidic with the exception of several scattered hydrophobic patches. At the N-terminus of M9NLS, residues 263-266 contact a hydrophobic patch on Kapβ2 helices H19A and H20B (Figure 2-4B, left). In the central region, a hydrophobic stretch ^{273}FGPM^{276} contacts hydrophobic Kapβ2 residues I773 and W730 (Figure 2-4 B and C). Farther C-terminus, F281 binds near a hydrophobic patch formed by Kapβ2 residues F584 and V643 (Figure 2-4B, center) and finally, the C-terminal ^{288}PY^{289} residues bind a large hydrophobic swath that includes Kapβ2 residues A380, A381, L419, I457 and W460 (Figure 2-4B, right and D). Despite the extensive acidic interface on Kapβ2, there are only two basic residues in M9NLS. R284 forms salt links with Kapβ2 residues E509 and D543, and the sidechain of K277 is not observed.

Distribution of binding energy along M9NLS

In order to understand the distribution of binding energy along M9NLS, we measured dissociation constants (K_Ds) of a series of M9NLS mutants binding to Kapβ2 using isothermal titration calorimetry (ITC). The results of the binding studies using MBP-fusion proteins of M9NLS residues 257-305 and wild type Kapβ2 are summarized

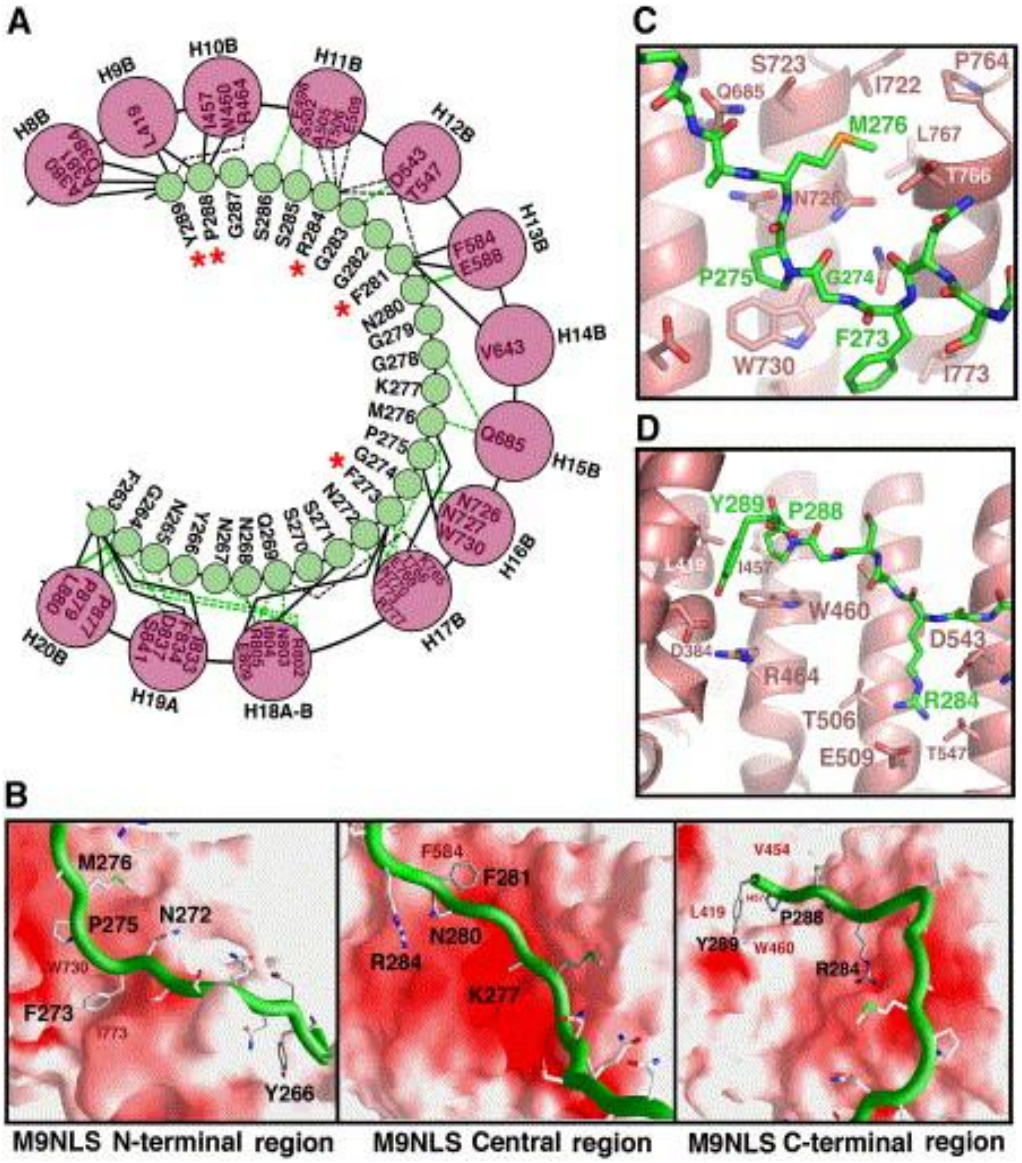


Figure 2-4: Kapβ2-M9NLS interactions. (A) Kapβ2-M9NLS contacts (< 4.0 Å) with M9NLS residues in green circles and Kapβ2 helices as pink circles. Contacts involving main chain atoms of M9NLS are shown with green lines. Contacts involving M9NLS sidechains are shown with black lines. Solid lines are hydrophobic contacts and dashed lines are polar contacts. Red asterisks label M9NLS residues that make two or more sidechain contacts in both complexes in the asymmetric unit. (B) The Kapβ2-M9NLS interface. The N-terminal third (left), the central region (middle) and the C-terminal third (right) of M9NLS. Substrate is shown as a green ribbon and the Kapβ2 electrostatic potential is mapped onto its surface, all drawn using GRASP (Nicholls, Sharp et al. 1991). Red indicates negative electrostatic potential, white neutral and blue positive.

Residues in the hydrophobic patches of Kapβ2 are labeled in red and M9NLS residues labeled in black. (C) Interactions between Kapβ2 (red) and substrate at M9NLS (green) residues ^{273}FGPM^{276}, drawn using PYMOL (DeLano 2002).
(D) Interactions between Kapβ2 (red) and M9NLS (green) at the C-terminus of the substrate, drawn using PYMOL (DeLano 2002).

in Table 2-2. Wild type M9NLS binds Kapβ2 with a KD of 42 nM. This ITC-measured affinity is somewhat lower than the previous K_D of 2 nM measured by fluorescence titration, but may be explained by the presence of both a covalently attached aromatic fluorophore and a significantly longer M9NLS spanning residues 238-320 in the earlier studies (Chook, Jung et al. 2002). Substrate residues that make two or more sidechain contacts with Kapβ2 (F273, F281, R284, P288 and Y289) were systematically mutated to alanines. Additional residues G274, P275 and M276 were also mutated given their implied importance in yeast-two-hybrid studies (Bogerd, Benson et al. 1999).

G274A is the only single mutant that shows significant (18-fold) decrease in Kapβ2 binding (Table 2-2). Single mutants of C-terminal residues P288 and Y289 follow with modest decreases of 3-4 fold. Thus, it appears that M9NLS binds Kapβ2 in a mostly distributive fashion, with a strict requirement for glycine at position 274 and modest though possibly important energetic contributions from C-terminal residues P288 and Y289. The importance of the PY motif is suggested in the R284/P288/Y289 and G274/P288/Y289 triple mutants where 10-fold and 140-fold decreases were observed, respectively. Both triple mutants show non-additivity in their binding energies when compared with single G274A, R284A and the double PY mutants, suggesting cooperativity between the C-terminal PY motif and both upstream binding sites at R284 and G274. The significance of the G274A mutation had previously been reported in both

Table 2-2: Kapβ2 binding to M9 NLS mutants: dissociation constants by ITC

MBP-M9NLS(257-305) proteins	K_D
Wild type	42 ± 2 nM
F273A	61 ± 10 nM
G274A	746 ± 63 nM
P275A	74 ± 5 nM
M276A	83 ± 17 nM
F281A	56 ± 11 nM
R284A	92 ± 9 nM
P288A	158 ± 20 nM
Y289A	133 ± 21 nM
P288A/Y289A	136 ± 8 nM
R284A/P288A/Y289A	461 ± 27 nM
G274A/P288A/Y289A	5.9 ± 0.7 μM

Kapβ2-binding and nuclear import assays (Nakielny, Siomi et al. 1996; Fridell, Truant et al. 1997). The alpha carbon of G274 is in close proximity to neighboring substrate sidechains F273 and P275 as well as Kapβ2 residue W730, such that a sidechain in position 274 may result in a steric clash (Figure 2-4C).

The important energetic contributions of the substrate's C-terminal PY motif and its central G274 residue are also supported by mutations of interacting residues in Kapβ2. Double and triple Kapβ2 mutants, W460A/W730A and I457A/W460A/W730A, both show significant decreases in Kapβ2 binding (Figure 2-5). I457 and W460 interact with the substrate PY motif while W730 makes a hydrophobic contact with substrate P275 and is also close to G274 (Figure 2-4 C and D).

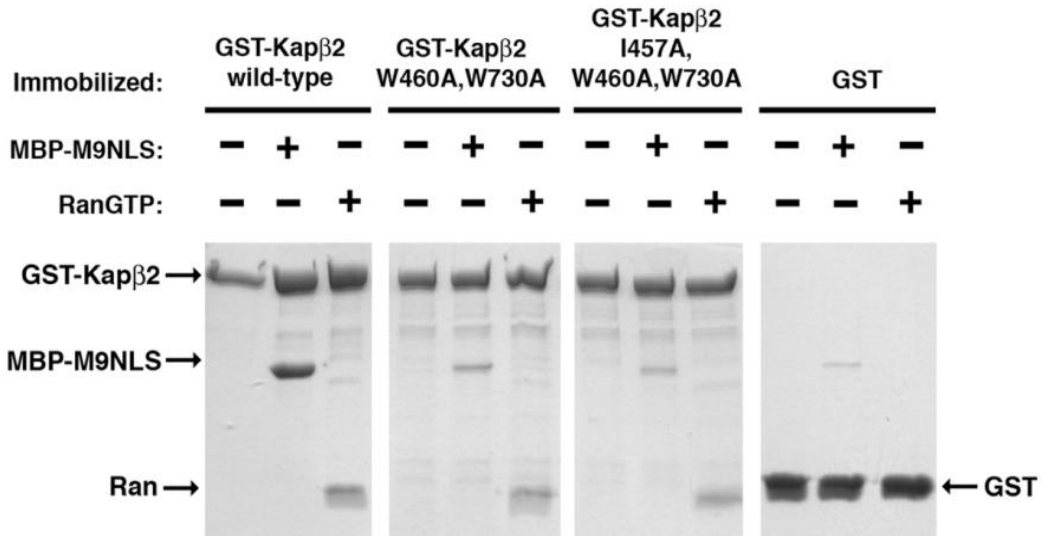


Figure 2-5: Binding studies of MBP-M9NLS and immobilized Kapβ2 mutants. Control experiments were also performed using immobilized Kapβ2 proteins and RanGTP.

Rules for substrate recognition by Kapβ2

Prior to this study, among more than 20 known Kapβ2 substrates, only NLSs from hnRNP A1, D, HuR, TAP and their homologs had been identified (Siomi and Dreyfuss 1995; Fan and Steitz 1998; Truant, Kang et al. 1999; Kawamura, Tomozoe et al. 2002; Suzuki, Iijima et al. 2005). All four NLSs span 30-40 residues, are rich in glycine and serine residues, have overall basic character, but share little sequence homology. To aid in assessment of the rules for NLS recognition by Kapβ2 suggested below, we constructed a series of deletion mutants to map three additional NLSs from hnRNP F, M and PQBP-1. The results of *in vitro* binding assays map the NLSs to residues 151-190 in PQBP-1, residues 41-70 in hnRNP M and residues 190-245 in hnRNP F. Structural and mutagenesis analysis of the Kapβ2-M9NLS complex combined

with sequence comparison and analysis of all seven NLSs reveals three rules for NLS recognition by Kapβ2.

Rule 1: NLS is structurally disordered in substrate

The extended conformation of the 26-residue M9NLS results in a linear epitope that traces a path of ~ 110 Å. The structure of the bound substrate suggests that an NLS recognized by Kapβ2 should exist within a stretch of at least 30 residues that lacks secondary structure in its native, unbound state. Thus, the NLS is most likely structurally disordered in the free substrate. The prediction of this NLS requirement is further supported by the fact that all seven known NLSs in Kapβ2 substrates occur within sequences with high probability of structural disorder (> 0.7) calculated by the program DISEMBL (Linding, Jensen et al. 2003). All seven NLSs are found either in loop regions between the RNA binding or other folded domains or at the termini of the substrates.

Rule 2: Overall positive charge for NLS is preferred

A second requirement for an NLS recognized by Kapβ2 emerges from the observation that Kapβ2's substrate interface is highly negatively charged. An acidic peptide would likely not bind due to electrostatic repulsion, while an NLS with overall positive charge would most likely be favored. Examination of all known Kapβ2 NLSs indicates overall basic character spanning at least 30 residues in six of seven cases (Figure 2-6A). In addition, regions that flank the NLSs most likely also contribute favorably to electrostatics. For example, although the TAP-NLS sequence delineated in Figure 2-6A has slightly more acidic than basic residues, flanking regions are highly basic and may ultimately contribute to overall basic character to promote Kapβ2 binding.

The importance of basic flanking regions is also observed in hnRNP A1. Here, the entire 135-residue C-terminal tail of the substrate has overall positive charge. A recent study showed that following osmotic shock stress in cells, four serine residues C-terminally adjacent to the M9NLS are phosphorylated, resulting in decreased binding to Kapβ2 and accumulation of hnRNP A1 in the cytoplasm (Allemand, Guil et al. 2005). Phosphorylation of the M9NLS-flanking serines may decrease the basic character of M9NLS and thus modulate interactions with Kapβ2.

Rule 3: Consensus sequences for the NLS

All seven characterized NLSs recognized by Kapβ2 exist in structurally disordered regions suggesting that this class of NLS is represented by linear epitopes and not folded domains. However, apparent sequence diversity among previously characterized NLSs from hnRNP A1, HuR, TAP and JKTBP homologs had prevented delineation of a consensus sequence that could be used to identify new NLSs or substrates. However, despite apparent NLS diversity, mutagenesis, structural and sequence analysis have resulted in identification of two regions of conservation within the sequences.

The first region of conservation is found at the C-terminus of the NLSs. Mutagenesis of M9NLS suggested the importance of its C-terminal PY motif (Table 2-2). Sequence examination of previously characterized NLSs from hnRNP D, HuR and TAP as well as the newly characterized NLSs of hnRNP F, M and PQBP-1, identified consecutive PY residues in six of the seven sequences (Figure 2-6A). Mutations of the PY residues in PQBP-1 and hnRNP M also decreased Kapβ2 binding suggesting that they make energetically important contacts (Figure 2-6B). Mutations of the PY motif in

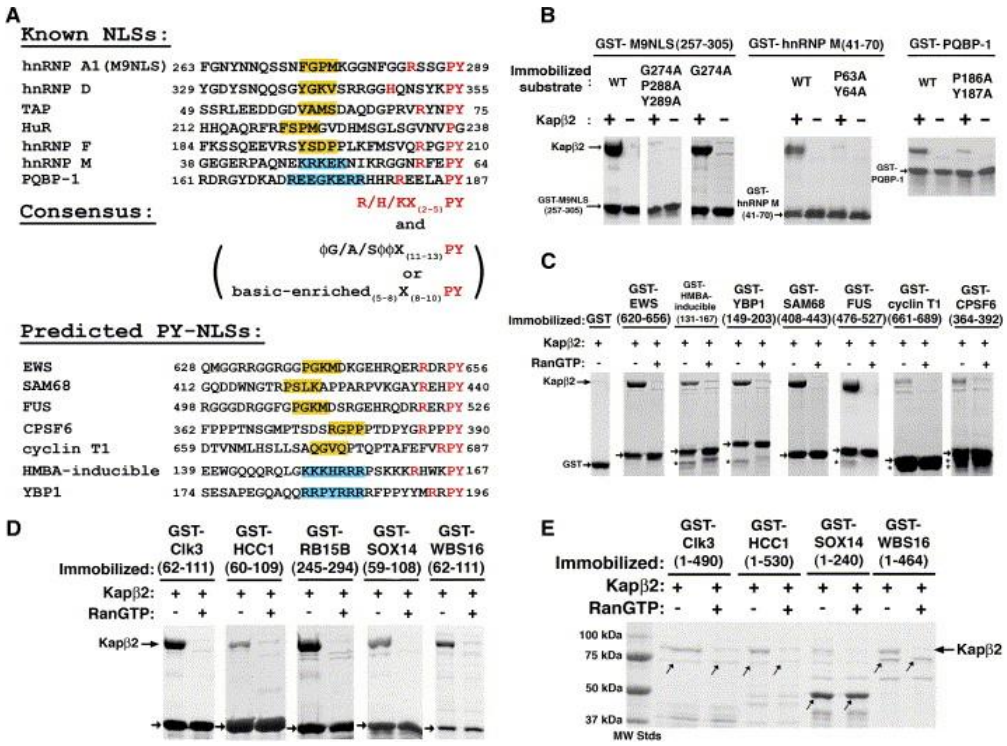


Figure 2-6: Consensus sequences of NLSs recognized by Kapβ2. (A) Alignment of all known (top) and predicted NLSs (bottom) recognized by Kapβ2, at conserved PY residues. NLSs in known Kapβ2 substrates are predicted by the presence of the R/K/H-$X_{(2-5)}$-P-Y C-terminal motifs (red) within structurally disordered and positively charged regions of 30 amino acids. Central hydrophobic motifs ϕG/A/Sϕϕ (ϕ is a hydrophobic sidechain) are shaded yellow. Central basic motifs are shaded blue. (B) Binding assays of Kapβ2 and immobilized alanine mutants of M9NLS, PQBP-1 and NLS-containing fragments of hnRNP M. Bound proteins are visualized with Coomassie Blue. (C) Binding assays of predicted NLSs from known Kapβ2 substrates EWS, HMBA-inducible protein, YBP1, SAM68, FUS, cyclin T1 and CPSF6. Kapβ2 is added to immobilized GST-NLSs (arrows) in the presence and absence of excess RanGTP, and bound proteins visualized with Coomassie Blue. Asterisks label degraded fragments of substrates. (D) Five predicted Kapβ2 substrates (Clk3, HCC1, RB15B, Sox14 and WBS16) are validated experimentally. GST-NLSs (arrows) are immobilized on glutathione sepharose. (E) Binding assays of full-length substrates Clk3, HCC1, Sox14 and WBS16 to Kapβ2. Expression of recombinant full-length RB15B was not successful. Coomasie-stained bands at the size of the GST-substrates are labeled with arrows. Lower molecular weight proteins are likely degraded substrates.

JKTBP proteins and M9NLS were also previously shown to inhibit nuclear import (Suzuki, Iijima et al. 2005; Iijima, Suzuki et al. 2006). In addition, we observe that a basic residue is always found several residues N-terminal of the PY sequence, consistent with an adjacent acidic surface on Kapβ2 (Figure 2-4B and D and Figure 2-6A). Based on these observations, we propose a C-terminal consensus sequence R/K/H-$X_{(2-5)}$-P-Y (where X is any residue) for NLSs recognized by Kapβ2. We refer to this class of NLSs as PY-NLSs.

A second region of conservation within the PY-NLSs is found in the central region of the peptides. Examination of the central region divides the seven PY-NLSs into two sub-classes. The first sub-class includes M9NLS and NLSs of hnRNP D, F, TAP and HuR, where four consecutive predominantly hydrophobic residues are located 11-13 residues N-terminal to the PY residues (Figure 2-6A). We refer to this sub-class of sequences as hydrophobic PY-NLSs or hPY-NLSs. In contrast, the central regions of NLSs from hnRNP M and PQBP-1 are virtually devoid of hydrophobic residues but are instead enriched in basic residues. They appear to represent a distinct sub-class of PY-NLSs that we call the basic PY-NLSs or bPY-NLSs.

The central hydrophobic motif in M9NLS spans residues ^{273}FGPM^{276} previously found in yeast two-hybrid and mutagenesis analysis to be important for import by Kapβ2, and a consensus sequence of Z-G-P/K-M/L/V-K/R (where Z is a hydrophobic residue) was previously suggested (Bogerd, Benson et al. 1999). The mutagenesis-derived consensus holds in the context of the M9NLS sequence, but does not describe NLSs in other Kapβ2 substrates. A loose consensus of ϕ-G/A/S-ϕ-ϕ (where ϕ is a hydrophobic sidechain) seems more appropriate upon comparison of the five central hydrophobic

motifs in hnRNPs A1, D, F, TAP and HuR (Figure 2-6). The Kapβ2-M9NLS structure explains preferences for hydrophobic sidechains in positions 1, 3 and 4 as well as small or no sidechain in position 2. Position 1 in M9NLS is F273, which occupies a hydrophobic pocket formed by Kapβ2 residues W730 and I773 (Figure 2C). Position 3 is occupied by P275, which stacks on top of the indole ring of Kapβ2 W730, and M276 in position 4 binds a small hydrophobic patch on Kapβ2 formed by I722, P764, L766 and the Cβ of S767. Thus, hydrophobic or long aliphatic sidechains at positions 1, 3 and 4 in other hydrophobic hPY-NLSs would provide energetically favorable hydrophobic contacts with Kapβ2. Mutagenesis of M9NLS suggests a strict requirement for glycine at position 2 (residue G274 in M9NLS) of the central hydrophobic motif. G274 is surrounded by adjacent substrate residues F273, P275 and Kapβ2 residue W730, suggesting that the strict requirement for glycine is likely heavily dependent on the identity of adjacent substrate residues. Nevertheless, hydrophobic neighbors, even those not as bulky as F273 and P275 in M9NLS, will likely still not accommodate large sidechains in position 2.

The Kapβ2-M9NLS structure provides some suggestion for the how the central basic motif in the bPY-NLSs could be accommodated. In the structure, the M9NLS hydrophobic motif interacts with Kapβ2 hydrophobic residues that are surrounded by numerous acidic residues (Figure 2-4B and C). Thus, the highly acidic substrate interface on Kapβ2 that contacts the central region of an NLS should also be able to interact favorably with numerous basic sidechains. It is possible that the central basic and hydrophobic motifs in the two sub-classes of PY-NLSs may take slightly different paths

on Kapβ2. Structures of Kapβ2 bound to bPY-NLSs will be necessary to understand the difference between the two subclasses of PY-NLSs.

The NLS rules are predictive

We have examined the sequences of eight recently identified Kapβ2 substrates: Ewing Sarcoma protein (EWS), HMBA-inducible protein, Y-box binding protein 1 (YBP1), SAM68, FUS, DDX3, CPSF6 and cyclin T1 (Guttinger, Muhlhausser et al. 2004), and found the C-terminal R/K/H-$X_{(2-5)}$-P-Y consensus within structurally disordered and positively charged regions of seven of them. The predicted NLSs for EWS, HMBA-inducible protein, YBP1, SAM68, FUS, CPSF6 and cyclin T1 are listed in the bottom half of Figure 2-6A. The predicted signals in EWS, SAM68, FUS, CPSF6 and Cyclin T1 are hPY-NLSs and those from HMBA-inducible protein and YBP1 are bPY-NLSs. The easily-detected PY motif is absent from DDX3, and we have not been able to show direct binding of DDX3 to Kapβ2 (data not shown). Thus, DDX3 may not be a substrate of Kapβ2, but may enter the nucleus by binding to a bona-fide Kapβ2 substrate. All seven predicted NLSs bind Kapβ2 and are dissociated from the karyopherin by RanGTP, consistent with NLSs imported by Kapβ2 (Figure 2-6C). The NLSs of cyclin T1 and CPSF6 bind Kapβ2, but more weakly than other substrates. It is not clear if this is due to proteolytic degradation of the substrates or to poor central hydrophobic motifs (Figure 2-6A and C). Confirmation of these seven NLSs indicates that the three rules for NLS recognition by Kapβ2 described above are predictive.

We have also applied the NLS rules to human proteins in the SwissProt protein database (Bairoch, Boeckmann et al. 2004) to identify potential Kapβ2 substrates. A search for proteins containing NLS sequence motifs (Figure 2-6A) using the program ScanProsite (Gattiker, Gasteiger et al. 2002), followed by filtering for structural disorder (DisEMBL) (Linding, Jensen et al. 2003) and for overall positive charge in the NLS resulted in 81 new candidate Kapβ2 substrates (Table 7-8 and Table 7-9). We chose five of these at random - protein kinase Clk3 (P49761), transcription factors HCC1 (Q14498), mRNA processing protein RB15B (Q8NDT2) and Sox14 (O95416), and the Williams-Beuren syndrome chromosome region 16 protein/WBS16 (Q96I51) and showed that both their predicted NLSs and the full length proteins (except for RB15B, which could not be expressed in bacteria) bind Kapβ2 and can be dissociated by RanGTP (Figure 2-6D and E). Thus, the rules not only identify NLSs in known substrates, but also are highly effective in predicting entirely new substrates.

Of the 81 candidate Kapβ2 substrates, 48 contain hPY-NLSs (Table 7-8), 28 contain bPY-NLSs (Table 7-9) and 5 contain PY-NLSs with both basic and hydrophobic central motifs. 49 of the new substrates (~60%) are involved in transcription or RNA processing, 18 have unknown cellular activity and the rest are involved in signal transduction (8), cell cycle regulation (3) and the cytoskeleton (3). Interestingly, information on subcellular localization is available for 62 of the predicted substrates, of which 57 (92%) are annotated to have nuclear localization. The SwissProt database used in the search is the most highly annotated and non-redundant protein database, but it is still incomplete for human proteins (Apweiler, Bairoch et al. 2004). Thus, the number of

new Kapβ2 substrates listed in Table 7-8 and Table 7-9 is a lower limit of the complete set of Kapβ2 import substrates. The large number of Kapβ2 substrates currently predicted by our NLS rules already implies the generality and prevalence of PY-NLSs. Kapβ1 and Crm1 are also involved in mitosis and centrosome duplication ((Arnaoutov, Azuma et al. 2005) and reviewed in (Harel and Forbes 2004; Mosammaparast and Pemberton 2004; Budhu and Wang 2005)), suggesting that many other Kapβs may be similarly involved in multiple cellular functions in addition to nuclecytoplasmic transport. Thus, Kapβ2 substrates will likely include ligands responsible for other still unknown cellular functions of Kapβ2 as well as large numbers of cargoes for nuclear import.

Discussion

The crystal structure of Kapβ2 bound to its substrate M9NLS has revealed a set of rules that describe the recognition of a large class of nuclear import substrates. M9NLS adopts an extended conformation for 26 residues when bound to Kapβ2, leading to the first rule, that NLSs recognized by Kapβ2 are structurally disordered in the free substrates. The structure also shows that the substrate binding site on Kapβ2 is highly acidic, leading to the second rule, that NLSs will have an overall positive charge. Finally, biochemical analyses of Kapβ2-M9NLS interactions have mapped M9NLS residues that are important for Kapβ2 binding, and examination of other Kapβ2 substrates has revealed consensus motifs at these regions. The consensus motifs include

a central hydrophobic or basic motif followed by a C-terminal R/K/HX$_{(2-5)}$PY motif, leading to the name PY-NLSs for this class of signals. Although these rules are not strong filters individually or in pairs (not shown), together they provide substantial restrictions in sequence space. The three rules have been used to identify NLSs in seven previously identified Kapβ2 substrates and more importantly to predict 81 new candidate Kapβ2 substrates in our initial bioinformatics endeavor. Of the members of this predicted group with annotated subcellular localization, >90% are reported to be nuclear localized. We have experimentally validated all seven new NLSs of known Kapβ2 substrates and five new bioinformatics-predicted substrates for Kapβ2 recognition as well as Ran-mediated dissociation, demonstrating the predictive nature of the rules. The large number of predicted Kapβ2 substrates further suggests the prevalence of PY-NLSs in the genome. Finally, the fact that all 81 proteins likely use Kapβ2 suggests potential functional linkages in the group that may be revealed by comparison with other genome-wide analyses.

CHAPTER 3

MODULAR ORGANIZATION AND COMBINATORIAL ENERGETICS OF PY-NUCLEAR LOCALIZATION SIGNALS[a]

Abstract

Proline-tyrosine nuclear localization signals (PY-NLSs) are recognized and transported into the nucleus by human Karyopherin(Kap)β2/Transportin and yeast Kap104p. Multipartite PY-NLSs are highly diverse in sequence and structure, share a common C-terminal R/H/K$X_{(2-5)}$PY motif and can be subdivided into hydrophobic and basic subclasses based on loose N-terminal sequence motifs. PY-NLS variability is consistent with weak consensus motifs, but such diversity potentially renders comprehensive genome-scale searches intractable. Here, we use yeast Kap104p as a model system to understand the energetic organization of this NLS. First, we show that Kap104p substrates contain PY-NLSs, demonstrating their generality across eukaryotes. Previously reported Kapβ2-NLS structures explain Kap104p specificity for basic PY-NLS. More importantly, thermodynamic analyses revealed physical properties that govern PY-NLS binding affinity: 1) PY-NLSs contain three energetically significant linear epitopes, 2) each epitope accommodates substantial sequence diversity, within defined limits, 3) the epitopes are energetically quasi-independent and 4) a given linear epitope can contribute differently to total binding energy in different PY-NLSs, amplifying signal diversity through combinatorial mixing of energetically weak and strong motifs. The modular organization of PY-NLS coupled with its combinatorial

energetics lays a path to decode this diverse and evolvable signal for future comprehensive genome-scale identification of nuclear import substrates.

[a]Originally published in PLoS Biology (PLoS Biol 6(6): e137).

Introduction

Karyopherinβ proteins (Kapβs; Importins/Exportins) mediate the majority of nucleocytoplasmic protein transport. There are 19 known Kapβs in human and 14 in yeast (Fried and Kutay 2003; Mosammaparast and Pemberton 2004). Kapβs bind substrates through nuclear localization or export signals (NLSs or NESs), transport them through the nuclear pore complex and the Ran GTPase regulates Kapβ-substrate interactions (Gorlich and Kutay 1999; Chook and Blobel 2001; Conti and Izaurralde 2001; Weis 2003). Ten Kapβs are known to function in nuclear import, each recognizing at least one distinct NLS.

The best-known NLS is the short basic classical-NLS, which is recognized by Kapα/Kapβ1 (Conti and Izaurralde 2001), and this pathway is functionally conserved from human to yeast (Enenkel, Blobel et al. 1995; Conti, Uy et al. 1998). Classical-NLSs can be divided into monopartite and bipartite NLSs. Monopartite NLSs contain a single cluster of basic residues whereas bipartite sequences contain two clusters of basic residues separated by a 10-12 amino acid linker. Thermodynamic dissection by scanning alanine mutagenesis of monopartite NLSs from the SV40 large T antigen (PKKKRKV) and the *c-myc* proto-oncogene (PAAKRVKLD) (Catimel, Teh et al. 2001; Hodel, Corbett et al. 2001; Lange, Mills et al. 2007) confirmed a previously determined consensus sequence of K(K/R)X(K/R) (Kalderon, Richardson et al. 1984; Conti, Uy et al. 1998). Binding energies of these small signals are dominated by a single lysine residue, in the third position of the SV40 large T antigen and in the fourth position of c-myc, which makes numerous interactions with Kapα (Hodel, Corbett et al. 2001). Thus, in the

monopartite classical-NLS, it is well-known that a relatively small motif is recognized, and binding energy is concentrated in stereotypical fashion across small sequences. Although numerous structures are available for bipartite NLSs (Conti and Kuriyan 2000; Fontes, Teh et al. 2000; Fontes, Teh et al. 2003), thorough thermodynamic analysis of this subclass is not available and its consensus is less well-defined (one example is $KRX_{10-12}KRRK$) than for the monopartite NLS. Furthermore, a non-functional SV40 NLS mutant was rescued by a bipartite-like addition of a 2-residue N-terminal basic cluster (Hodel, Corbett et al. 2001), suggesting that bipartite classical-NLSs can accommodate larger sequence diversity than their monopartite counterparts.

Recently, structural and biochemical analyses of human Kapβ2 (Transportin) bound to the hnRNP A1-NLS revealed physical rules that describe Kapβ2's recognition of a diverse set of 20-30 residue long NLSs that we termed PY-NLS (Lee, Cansizoglu et al. 2006). These rules are structural disorder of a 30-residue or larger peptide segment, overall basic character, and weakly conserved sequence motifs composed of a loose N-terminal hydrophobic or basic motif and a C-terminal $RX_{2-5}PY$ motif. The composition of N-terminal motifs divides PY-NLSs into hydrophobic and basic subclasses (hPY- and bPY-NLSs). The former contains four consecutive predominantly hydrophobic residues, while the equivalent region in bPY-NLSs is enriched in basic residues.

Approximately one hundred different human proteins have been identified as potential Kapβ2 substrates (Siomi and Dreyfuss 1995; Weighardt, Biamonti et al. 1995; Pollard, Michael et al. 1996; Bonifaci, Moroianu et al. 1997; Fan and Steitz 1998; Truant, Kang et al. 1999; Kawamura, Tomozoe et al. 2002; Guttinger, Muhlhausser et al. 2004; Suzuki, Iijima et al. 2005; Lee, Cansizoglu et al. 2006). Table 3-1 summarizes previously

reported validated and potential PY-NLSs. Although many of these potential substrates were predicted by bioinformatics (Lee, Cansizoglu et al. 2006) and still need experimental testing, >20 have been validated for Kapβ2 binding (Table 3-1) (Siomi and Dreyfuss 1995; Weighardt, Biamonti et al. 1995; Pollard, Michael et al. 1996; Bonifaci, Moroianu et al. 1997; Fan and Steitz 1998; Truant, Kang et al. 1999; Kawamura, Tomozoe et al. 2002; Guttinger, Muhlhausser et al. 2004; Suzuki, Iijima et al. 2005; Lee, Cansizoglu et al. 2006). Comparison of *in vivo* and *in vitro* validated PY-NLSs shows large sequence diversity, which is reflected in weak consensus sequences (Lee, Cansizoglu et al. 2006). Structures of five different Kapβ2-bound PY-NLSs also show substantial variability, with structurally diverse linkers separating the convergent consensus regions (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007; Imasaki, Shimizu et al. 2007). The PY-NLS is significantly larger than the short monopartite classical-NLS. The well-defined consensus and concentrated binding energy of the latter may reflect compactness of the signal. In contrast, binding energy of the PY-NLS is spread over a much larger sequence. Physical properties of the multipartite PY-NLS may be more similar to those of the less-studied, larger and sequentially more diverse bipartite classical-NLS. Diverse PY-NLSs are necessarily described by weak consensus motifs. Therefore, instead of the traditional way of describing a linear recognition motif with a strongly restrictive consensus sequence, PY-NLSs were described by a collection of individually weak physical rules that together were able to provide substantial limits in

Table 3-1. Summary of validated and potential PY-NLSs

Human proteins	**PY-NLSs: validated *in vivo* and *in vitro*** [a, d, e]	**References**[h]
PQBP-1[b, g]	161 RDRGYDKADREEGKERRHHRREELAPY 187	1[e], 2[b,e]
YBP1[b, c, g]	176 SAPEGQAQQRRPYRRRRFPPYYMRRPY 196	1[c,e], 2[b], 4[d]
PABP2[b]	280 FYSGFNSRPRGRVYRGRARATSWYSPY 306	5[d]
EWS[b, c]	630 GGRRGGRGGPGKMDKGEHRQERRDRPY 656	1[c,e], 2[b], 6[d]
FUS[b, c, f]	500 GGDRGGFGPGKMDSRGEHRQDRRERPY 526	1[c,e], 2[b], 6[f]
$TAF_{II}68$[f]	566 YGGDRGGYGGKMGGRNDYRNDQRNRPY 592	6[f]
SAM68[b, c]	414 DDWNGTRPSLKAPPARPVKGAYREHPY 440	1[c,e], 2[b], 8[d], 7[d]
ETLE[b]	317 GQEEWTNSRHKAPSARTAKGVYRDQPY 343	9[d]
hnRNP M[b]	38 GEGERPAQNEKRKEKNIKRGGNRFEPY 64	1[e], 2, 28[d,e]
hnRNP A1[b]	263 FGNYNNQSSNFGPMKGGNFGGRSSGPY 289	10-12 [e,d], 13[d]
hnRNP A2[f]	293 NYNQQPSNYGPMKSGNFGGSRNMGGPY 319	12
hnRNP A3[f]	334 NYSGQQQSNYGPMKGGSFGGRSSGSPY 360	14
hnRNP A0[f]	264 SYGPMKSGGGGGGGSSWGGRSNSGPY 290	
hnRNP D[b]	329 YGDYSNQQSGYGKVSRRGGHQNSYKPY 355	15[e]
JKTBP-1[b]	394 ADYSGQQSTYGKASRGGGNHQNNYQPY 420	17[d,e]
TAP[b]	49 SSRLEEDDGDVAMSDAQDGPRVRYNPY 75	18, 19, 20
HuR[b]	212 HHQAQRFRFSPMGVDHMSGLSGVNVPG 238	22[d], 2[e], 21[e]
Human proteins	**Potential PY-NLSs: validated *in vitro*** [a, e]	**References**
HMBA-ind[b, c, g]	141 WGQQQRQLGKKKHRRRPSKKKRHWKPY 167	1[c,e], 2[b]
RB15B[c, g]	258 GLPKPWEERRKRRSLSSDRGRTTHSPY 284	1[c,e]
HCC1[c, g]	73 RSRSKERRRSRSRSRDRRFRGRYRSPY 99	1[c,e]
SOX14[c, g]	72 HPDYKYRPRRKPKNLLKKDRYVFPLPY 98	1[c,e]
Clk3[c, g]	75 RSPSFGEDYYGPSRSRHRRRSRERGPY 101	1[c,e]
hnRNP F[b]	184 FKSSQEEVRSYSDPPLKFMSVQRPGPY 210	1[e], 2[b], 23[b,e]
hnRNP H1[b,f]	184 FKSSRAEVRSYSDPPLKFMSVQRPGPY 210	2[b]
CPSF6[b, c]	364 PPTNSGMPTSDSRGPPPTDPYGRPPPY 390	1[c,e], 2[b]
Cyclin T1[b, c, g]	661 VNMLHSLLSAQGVQPTQPTAFEFVRPY 687	1[c,e], 2[b]
WBS16[c]	75 PSFVVPSSGPGPRAGARPRRRIQPVPY 101	1[c,e]
Yeast proteins	**PY-NLSs: validated *in vivo* and *in vitro*** [a, d, e]	**References**
Hrp1p[b]	506 RSGGNHRRNGRGGRGGYNRRNNGYHPY 532	24-27[b,d,e]
Nab2p[b]	214 AVGKNRRGGRGGNRGGRNNNSTRFNPL 241	24-27[b,d,e]

[a] Basic and hydrophobic motifs are shaded in black and grey, respectively. R/K/H and PY/L of the R/K/Hx_{2-5}PY/L motif are underlined and in bold font.

[b] Proteins were identified experimentally as Kapβ2 or Kap104p substrates.

[c] PY-NLSs (validated and potential) were identified by bioinformatics.

[d] Experimentally validated as NLS *in vivo* (includes targeting reporter to nucleus and subcellular mislocalization upon NLS mutation).

[e] Binds recombinant Kapβ2 or Kap104p and is dissociated from the Karyopherin by RanGTP.

[f] FUS and $TAF_{II}68$ (full length and NLSs) are highly homologous to EWS; hnRNPs A2, S3 and A0 are highly homologous to hnRNP A1; NLS of hnRNPs H1 and F are almost identical.

[g] Predicted to have a classical NLS within this region by WoLF PSORT (Horton, Park et al. 2007).

[h] References: 1) (Lee, Cansizoglu et al. 2006), 2) (Guttinger, Muhlhausser et al. 2004), 3) (Waragai, Lammers et al. 1999), 4) (Bader and Vogt 2005), 5) (Calado, Kutay et al. 2000), 6) (Zakaryan and Gehring 2006), 7) (Ishidate, Yoshihara et al. 1997), 8) (Lukong, Larocque et al. 2005), 9) (Wu, Zhou et al. 1999), 10) (Bonifaci, Moroianu et al. 1997), 11) (Pollard, Michael et al. 1996), 12) (Siomi and Dreyfuss 1995), 13) (Weighardt, Biamonti et al. 1995), 14) (Ma, Moran-Jones et al. 2002), 15) (Suzuki, Iijima et al. 2005), 16) (Siomi, Fromont et al. 1998), 17) (Kawamura, Tomozoe et al. 2002), 18) (Truant, Kang et al. 1999), 19) (Bear, Tan et al. 1999), 20) (Katahira, Strasser et al. 1999), 21) (Rebane, Aab et al. 2004), 22) (Fan and Steitz 1998), 23) (Siomi, Eder et al. 1997), 24) (Aitchison, Blobel et al. 1996), 25) (Truant, Fridell et al. 1998), 26) (Lee and Aitchison 1999), 27) (Siomi, Fromont et al. 1998), 28) (Cansizoglu, Lee et al. 2007)

sequence space for reasonable predictions of new Kapβ2 substrates (Lee, Cansizoglu et al. 2006). However, the currently predicted substrates are most likely only a fraction of all PY-NLS-containing proteins since narrow sequence patterns were used in the initial search to achieve optimal accuracy. In fact, the sequence patterns used (Lee, Cansizoglu et al. 2006) were too narrow to predict PY-NLSs in known substrates HuR, TAP, hnRNP F and JKTBP-1. The coverage of conventional sequence-based bioinformatics searches is expected to be severely limited due to PY-NLS diversity. Although sequence patterns obviously need to be expanded, we do not yet understand the limits of sequence diversity within motifs or how the different motifs may be combined. Knowledge of how binding energy is parsed in PY-NLSs will shape future efforts to decode these highly degenerate signals. Furthermore, physical understanding of how diverse PY-NLS sequences can achieve common biological function will also provide unique insights into many biological recognition processes that involve linear recognition motifs with weak and obscure consensus sequences, such as vesicular cargo sorting and protein targeting to the mitochondria and the peroxisome (Rapaport 2003; Mancias and Goldberg 2005; Brocard and Hartig 2006; Hegde and Bernstein 2006; Swanton and High 2006; Van Ael and Fransen 2006).

The yeast homolog of Kapβ2 is Kap104p (32% sequence identity) (Aitchison, Blobel et al. 1996). Only two Kap104p substrates, the mRNA processing proteins Nab2p and Hrp1p, are known. Several groups have mapped and validated NLSs of these substrates using both *in vivo* and *in vitro* methods to arginine-glycine (RG)-rich regions that were termed rg-NLSs (Siomi, Fromont et al. 1998; Truant, Fridell et al. 1998; Lee and Aitchison 1999). Little sequence homology was detected between NLSs recognized

by Kapβ2 and Kap104p. Furthermore, substrate recognition by the two Karyopherins appears non-analogous, as Kap104p does not recognize human substrate hnRNP A1 (Siomi, Fromont et al. 1998; Truant, Fridell et al. 1998). Given the recent physical understanding of Kapβ2-NLS interactions, we seek to examine the evolutionary conservation and energetic organization of signals in this pathway through studies of Kap104p-NLS interactions.

First, we present biochemical and biophysical analyses showing that RG-rich substrates of yeast Kap104p share similar physical characteristics as human PY-NLSs. Kap104p recognizes the basic but not hydrophobic PY-NLS subclass, and structural analyses of Kapβ2-NLS complexes suggested the origin of this specificity, enabling prediction of PY-NLS subclass specificity for all eukaryotic Kapβ2s. Thermodynamic analyses of Kap104p-NLS interactions revealed biophysical properties that govern binding affinity of PY-NLSs. These signals contain at least three energetically significant binding epitopes that are also linear motifs. Each linear epitope accommodates significant sequence diversity, and we have characterized some of the limits of this diversity. The linear epitopes are also energetically quasi-independent, a property that is probably due to intrinsic disorder of the free signals. Finally, in different PY-NLSs, a given epitope can vary significantly in its contribution to total binding energy. When combined with multivalency, this energetic variability can amplify signal diversity through combinatorial mixing of energetically weak and strong motifs.

Materials and Methods

Plasmids and Strains

The *Kap104* gene (gift from J.Aitchison) was subcloned into a modified pGex4T3 vector (GE Healthcare) with a Tev protease cleavage site (pGexTev) using BamHI and NotI restriction sites. Yeast substrate genes were obtained by PCR from a *S. cerevisiae* genomic DNA library (Novagen) and subcloned into BamHI and NotI sites in the pGexTev and/or pMalTev vector. Site directed mutagenesis of Nab2p 201-251 and Hrp1p 494-534 were performed using the QuikChange method (Stratagene) and confirmed by nucleotide sequencing. A complete list of plasmids can be found in Table 7-2 and Table 7-3 in the appendix.

Full-length *Nab2* and *Hrp1* wild type and mutant genes were cloned into a pRS415 shuttle vector (New England Biosystems). Both plasmids contain a constitutive ADH1 plasmid cloned into SacI and XbaI restriction sites (Figure 3-1). *Nab2* was inserted via SpeI and SmaI sites and *Hrp1* was inserted using SpeI and XhoI sites. A GFP gene was cloned 3' of *Nab2* into PstI and XhoI sites (Figure 3-1A). The NLS in Hrp1p is at its C-terminus therefore GFP was cloned 5' to *Hrp1*. GFP and a ggsgg linker were inserted using XbaI and SpeI sites (Figure 3-1B). The vector had to be treated with calf intestinal protease prior to ligation because XbaI and SpeI have compatible sticky ends. A complete list of shuttle plasmids can be found in Table 7-4.

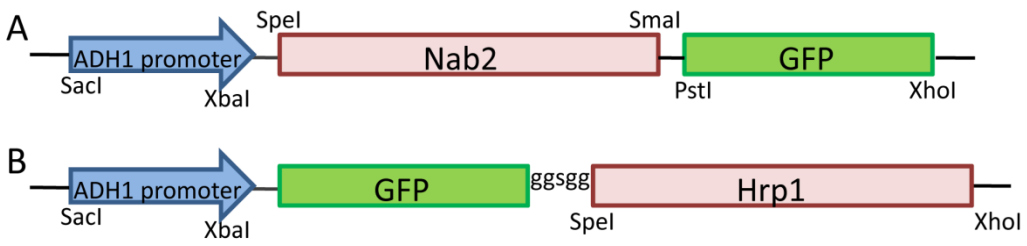


Figure 3-1: Full length *Nab2* (A) and *Hrp1* (B) in pRS415 shuttle vectors.

Cell culture and microscopy

Wild type BY4741 (*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) cells harboring pRS415 plasmids were grown at 30°C in SC-leu media to mid-logarithmic phase (Brachmann, Davies et al. 1998). Cells were transferred to a 1.5% low melting agarose pad made with SC-leu in a coverslip bottom Wilco dish. Cells were observed on an Olympus IX-81 inverted microscope (60X objective) and images were acquired with a Hamamatsu ORCA-ER camera. All images were analyzed in Image-Pro Plus software (Media Cybernetics). To obtain N:C ratio, mean fluorescence intensity in a 36-pixel box was measured in the nucleus and cytoplasm for at least 50 cells of each mutant.

Bioinformatics search for new Kap104p substrates

Potential Kap104p substrates were identified from *S. cerevisiae* proteins in the UniProtKB/Swiss-Prot database by the program ScanProsite using the sequence patterns Φ_1-G/A/S-Φ_3-Φ_4-X_{7-12}-R/K/H-X_{2-5}-P-Y/L (Φ_1 is a hydrophobic residue and Φ_3 and Φ_4 are hydrophobic residues or R or K) and K/R-X_{0-6}-K/R-X_{0-6}-K/R-X_{0-6}-K/R-X_{2-5}-R/K/H-X_{1-5}-PY (Gattiker, Gasteiger et al. 2002; Bairoch, Boeckmann et al. 2004). The resulting proteins were filtered for structural disorder using the program DisEMBL (Linding,

Jensen et al. 2003). Finally, only positively charged 50-residue sections were kept. We also eliminated transmembrane proteins and proteins found in domains.

Protein expression and purification

Twelve liters of GST-Kap104p protein was expressed in *Escherichia coli* Rosetta(DE3)pLysS cells. The Rosetta cells had better expression than a normal BL21 strain. Rosetta cells contain an extra plasmid which supplies tRNAs for rarely used codons and the pLysS plasmid minimizes basal expression of toxic genes. Cells were grown at 37ºC for 3 hours until they reached an OD600=0.7. They were induced with 0.5 mM IPTG for 6 hours at 30ºC. After centrifugation the pellet was resuspended in Tris buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 2 mM DTT, 20% glycerol) plus protease inhibitors.

Cells were lysed three times using an EmulsiFlex-C5 homogenizer (Avestin). After the lysate was centrifuged for 30 minutes at 12000 rpm, the supernatant was applied to 15 mL glutathione sepharose (GE Healthcare) at 4ºC. The beads were washed 12 times with 12 mL Tris buffer followed by 10 washes with ATP buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EGTA, 10 mM MgAc, 2 mM DTT, 20% glycerol, 5 mM ATP, protease inhibitors) at room temperature. The beads were then washed three more times with Tris buffer. GST-Kap104p protein was eluted from the beads with Tris buffer plus 20 mM glutathione pH 8.1 (Figure 3-2). The eluted protein was concentrated to 10 mL using an Amicon Ultra Centrifugal Device (Millipore). Tev protease was added to the protein overnight at room temperature to cleave GST from Kap104p (Figure 3-2).

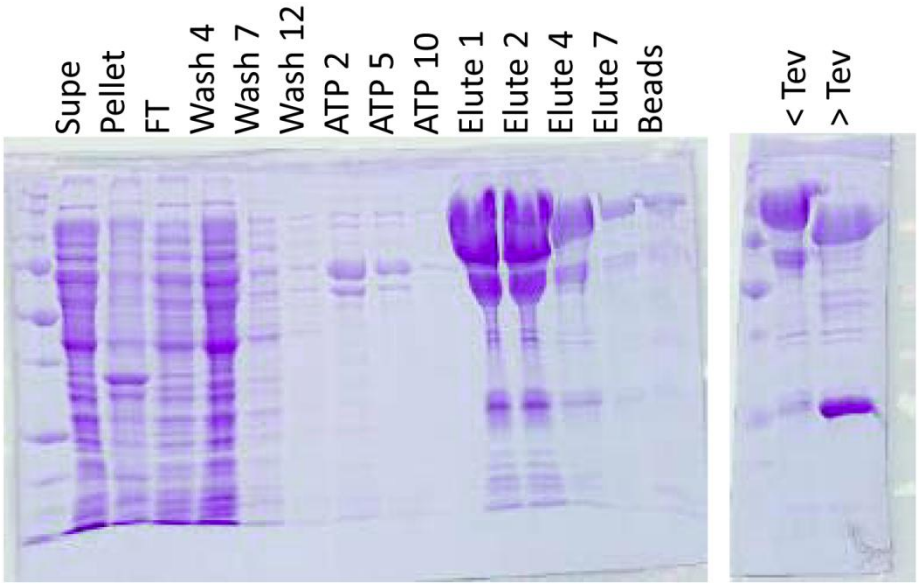


Figure 3-2. Coomassie stained gel of 10µL samples of washes from a Kap104p affinity purification.

The next morning, cleaved protein was diluted to 50 mL in no salt Q buffer (20 mM Imidazole pH 6.5, 1 mM EDTA, 2 mM DTT, 20% glycerol) and injected onto an anion exchange column (GE Healthcare). Protein elutes in two main peaks. The first peak is GST (Figure 3-3). It is more positively charged than Kap104p and does not bind the column as tightly. The second, broader peak is Kap104p. The selected fractions were pooled and concentrated to 8 mL. Kap104p is purified further by gel filtration chromatography (Figure 3-4). One milliliter aliquots were injected onto the column in TB buffer (20 mM Hepes pH 7.3, 110 mM KAc, 2 mM MgAc, 1 mM EGTA, 2 mM DTT, 20% glycerol). After the cleanest fractions were pooled, Kap104p was applied to 1 mL glutathione sepharose to remove any remaining GST protein. The yield

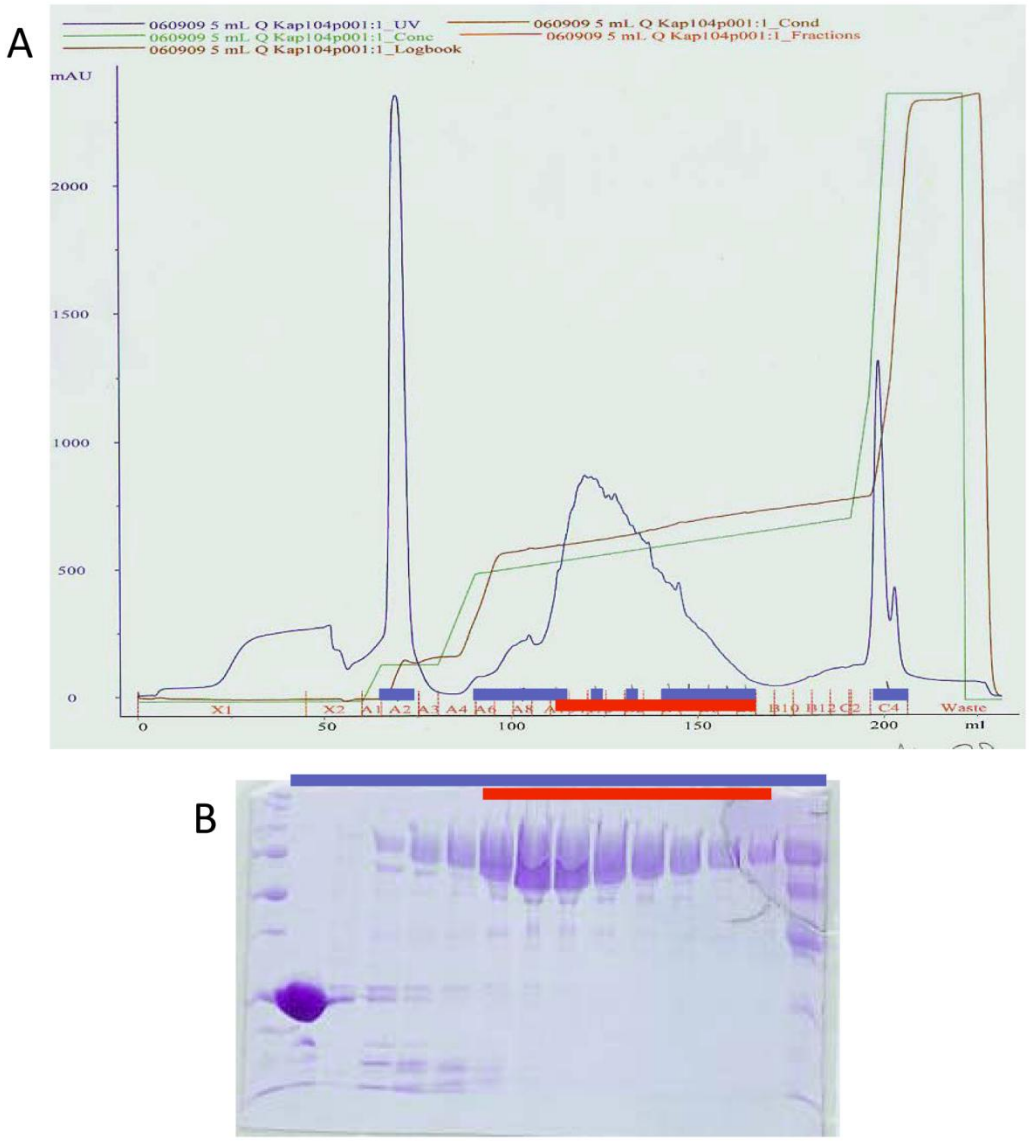


Figure 3-3. Anion exchange purification of Kap104p. The fractions marked by a blue line on the chromatogram (A) were the fractions run on the gel (B). Fractions designated with a red line were pooled for further purication.

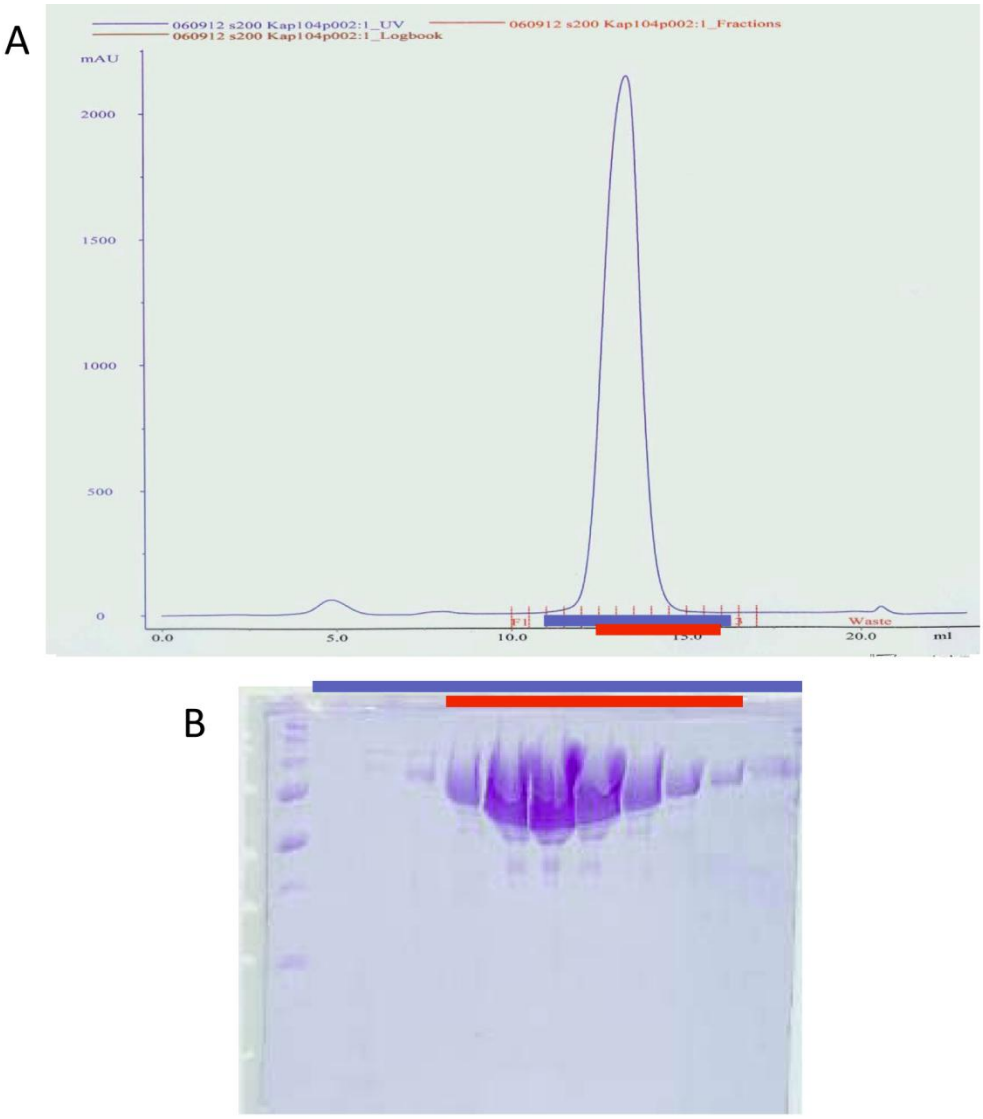


Figure 3-4. Purification of Kap104p by gel filtration chromatography. The fractions marked by a blue line on the chromatogram (A) were the fractions run on the gel (B). Fractions designated with a red line were pooled.

was consistently 4 mg/mL. Small aliquots of purified protein was flash frozen in liquid nitrogen and stored at -80ºC.

Three liters of BL21(DE3) cells expressing MBP-Hrp1p NLS or MBP-Nab2p NLS wild type or mutant protein were grown at 37°C for 3 hours until OD600=0.7. Cells were induced with 0.5 mM IPTG at 30°C for 3 hours and then harvested. After centrifugation the pellet was resuspended in MBP Tris buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 2 mM β-ME, 10% glycerol) plus protease inhibitors. Cells were lysed three times using an EmulsiFlex-C5 homogenizer. After the lysate was centrifuged for 30 minutes at 12000 rpm, the supernatant was applied to 10 mL amylose resin (New England Biosystems) at 4ºC. The beads were washed 20 times with 10 mL MBP Tris buffer and the protein was eluted with MBP Tris buffer plus 10 mM maltose (Figure 3-5).

Elutions 1-6 were pooled and concentrated to 20 mL. The protein was diluted to 50 mL with Tris no salt buffer (Tris pH 7.5, 2 mM DTT, 10% glycerol) and injected onto a cation exchange column (GE Healthcare). The NLSs are rich in basic residues and thus are very positively charged. A lot of proteins did not bind to the column and were in the flow through fractions. The majority of the MBP-NLS protein was eluted in the second peak (Figure 3-6). The final yield was between 20 and 40 mg/mL.

GST-substrates for *in vitro* binding assays were grown as 50 mL cultures in BL21(DE3) cells under the same conditions as the MBP-NLSs. The cells were centrifuged and the pellet was resuspended in 4 mL TB buffer. Two milliliters of cells were lysed by sonication and immobilized on 0.15 mL glutathione sepharose. The

protein was then washed three times with TB buffer and left on the beads for binding assays.

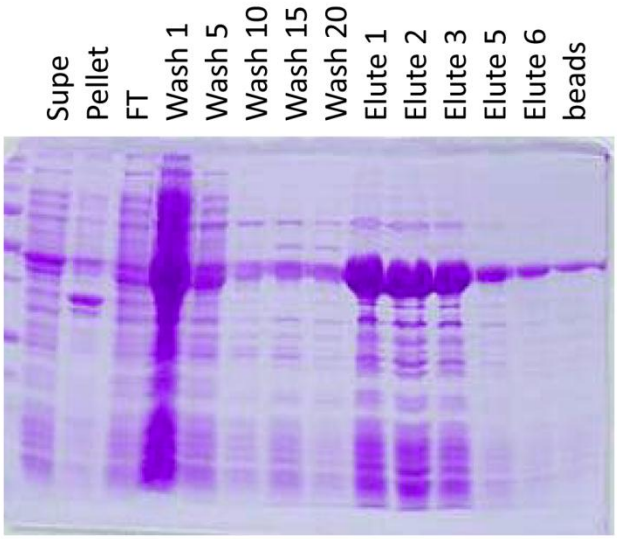


Figure 3-5: Gel from affinity purification of MBP-NLS proteins. Each lane is a 10 μL sample of the wash/elution. Gel is stained with Coomassie blue.

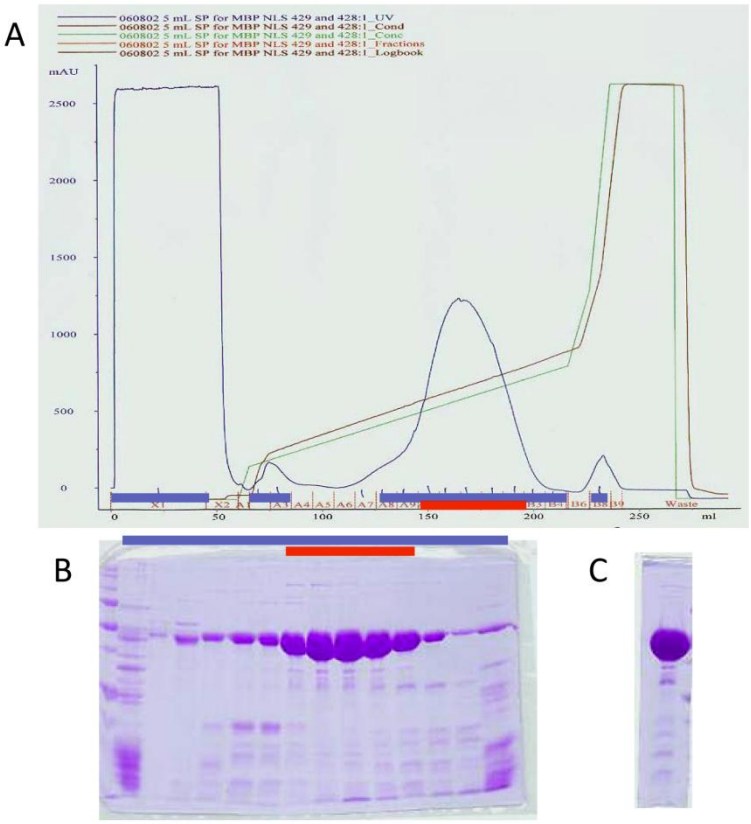


Figure 3-6. Cation exchange purification of MBP-NLS protein. Fractions marked with the blue line in the chromatogram (A) were run on a gel (B). Only fractions marked with a red line were pooled and concentrated for the final product (C).

Binding assays

Approximately 30 μg of Kap104p was added to ~10 μg of GST protein immobilized on 20 μL of glutathione sepharose followed by extensive washes with TB buffer and a second incubation with either buffer or RanGTP (5-fold molar excess). Immobilized proteins were visualized with SDS-PAGE and Coomassie staining.

Isothermal titration calorimetry

Affinities of wild type and mutant MBP-Nab2p-NLS and MBP-Hrp1p-NLS binding to Kap104p were determined by ITC using a MicroCal Omega VP-ITC calorimeter (MicroCal Inc., Northampton, MA). Proteins were dialyzed overnight against buffer containing 20 mM Tris pH 7.5, 100 mM sodium chloride, 2 mM β-mercaptoethanol and 10% glycerol. 90-350 μM MBP-NLS proteins were titrated into a sample cell containing 9-35 μM Kap104p. All ITC experiments were done at 20°C with 35 rounds of 8 μL injections. Data were plotted and analyzed with a single binding site model using MicroCal Origin software version 7.0.

Results

Yeast rg-NLSs are also PY-NLSs.

In vivo validated RG-rich NLSs of Hrp1p and Nab2p (or rg-NLSs) are located at residues 494-534 and 201-250, respectively (Figure 3-7A) (Siomi, Fromont et al. 1998; Truant, Fridell et al. 1998; Lee and Aitchison 1999; Marfatia, Crafton et al. 2003). Examination of their sequences revealed physical characteristics similar to those of human PY-NLSs. Hrp1p- and Nab2p-NLSs are located within structurally disordered segments of 120-190 residues (DisEMBL structural disorder probabilities 0.72 and 0.63 for Hrp1p and Nab2p, respectively (Linding, Jensen et al. 2003)) in the full length proteins (Figure 3-7A). ^{506}**R**SGGN**HRR**NG**R**GG**R**^{519} of Hrp1p and ^{216}**K**N**RR**GG**R**GGN**R**GG**R**^{229} of Nab2p contain many basic residues, like basic N-terminal motifs in human Kapβ2 substrates hnRNP M, PQBP-1 and YB-1 (Table 3-1) (Lee, Cansizoglu et al. 2006). Further C-terminally, the Hrp1p ^{525}**R**NNGYH**PY**^{532} and the Nab2p ^{235}**R**FN**PL**^{239} segments either match or are homologous to the C-terminal $Rx_{2-5}PY$ consensus.

Immobilized full length Hrp1p, Nab2p and their NLSs bound Kap104p in stoichiometric proportions in pull-down binding assays (Figure 3-7B). Although it was previously reported that Ran could not dissociate substrate from Kap104p (Lee and Aitchison 1999), we observed efficient dissociation of both full length substrates and NLSs by RanGTP, possibly due to higher activity and GTP loading of the recombinant Ran. Our results suggest that Kap104p-NLS interactions and regulation by Ran are

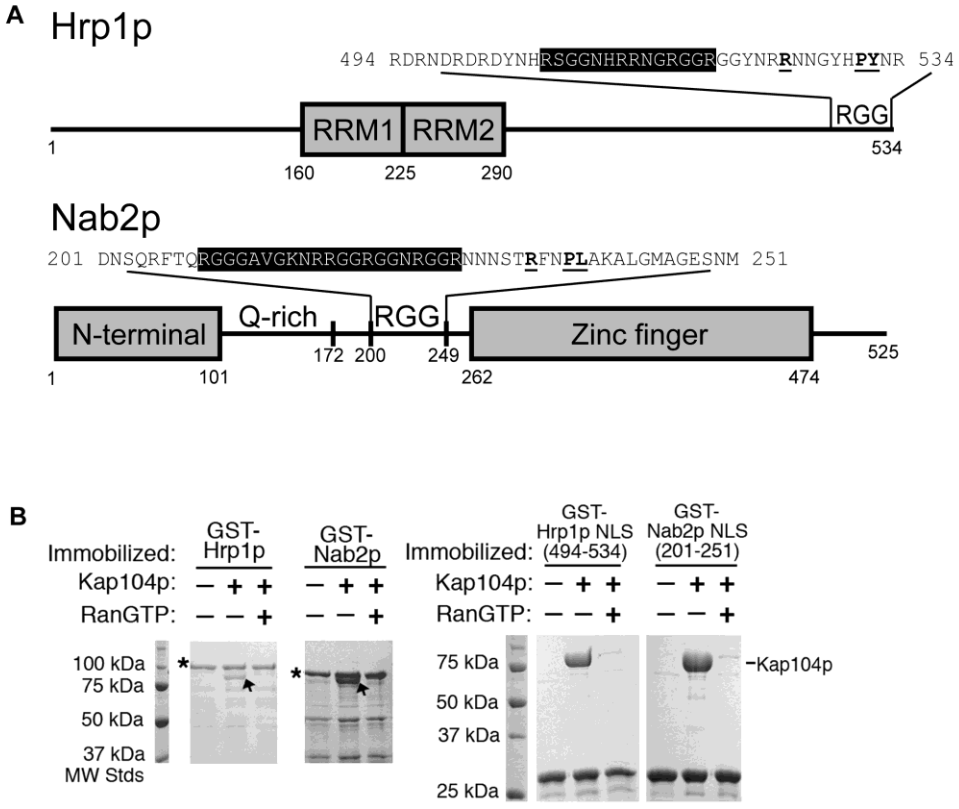


Figure 3-7. Interactions between Kap104p and import substrates Hrp1p and Nab2p. (A) Domain organization of Hrp1p and Nab2p. Domains are indicated by grey boxes and the RGG-regions and the glutamine-rich region are labeled. The sequences of the two NLSs are shown, with the basic motif highlighted in black and the $Rx_{2\text{-}5}PY(L)$ motif in bold and underlined. (B) Binding assays of Kap104p (arrow) with immobilized full length Hrp1p and Nab2p (asterisks) or Hrp1p- and Nab2p-NLSs, in the presence and absence of RanGTP. Bound proteins are Coomassie blue-stained.

similar to other characterized Kapβ-mediated nuclear import processes in human (Gorlich and Kutay 1999; Chook and Blobel 2001; Conti and Izaurralde 2001; Weis 2003). Thermodynamic parameters for Kap104p binding to Hrp1p and Nab2p-NLSs were obtained by isothermal titration calorimetry (ITC) (Figure 3-8). Both NLSs bound Kap104p with high affinity (K_D 32 nM for Hrp1p and 37 nM for Nab2p (Table 3-2 and Table 3-3)) and extensive mutagenesis of NLSs is discussed below. Thus, based on their sequence characteristics, high affinity for Karyopherin and dissociation by RanGTP, yeast NLSs recognized by Kap104p resemble PY-NLSs.

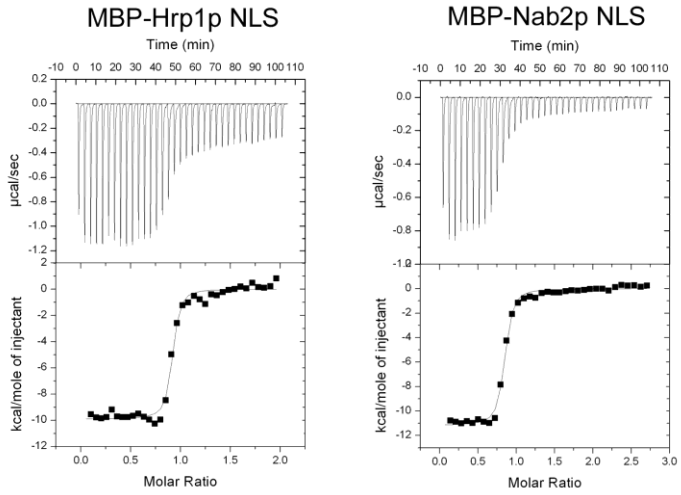


Figure 3-8. Isothermal titration calorimetry (ITC) measurements of Kap104p binding to MBP-Hrp1p-NLS and MBP-Nab2p-NLS.

Kap104p recognizes the basic but not hydrophobic subclass of PY-NLSs.

To investigate the PY-NLS subclass specificity of Kap104p, we examined its interaction with several human hPY- and bPY-NLSs as well as several predicted (see below) yeast hPY and bPY-NLSs. Splicing factor hnRNP A1 and mRNA transport factor TAP/NXF1 contain hPY-NLSs and splicing factor hnRNP M and FUS contain bPY-NLSs (Figure 3-10A). All four human PY-NLSs interacted with Kapβ2 (Lee, Cansizoglu et al. 2006), but only bPY-NLSs from hnRNP M and FUS bound yeast Kap104p in GST pull-down assays (Figure 3-10B). Both yeast Hrp1p- and Nab2p-NLSs bound equally well to Kap104p and Kapβ2 (Figure 3-9).

Hrp1p and Nab2p are the only two known Kap104p substrates (Aitchison, Blobel et al. 1996; Truant, Fridell et al. 1998; Lee and Aitchison 1999). We needed to identify additional yeast sequences to test the preference of Kap104p for bPY-NLS. Since Nab2p has a C-terminal PL instead of PY motif, suggesting that PL motifs may also be present in other functional PY-NLSs, we used the program ScanProsite (Gattiker, Gasteiger et al. 2002) and sequence patterns Φ_1-G/A/S-Φ_3-Φ_4-$X_{7\text{-}12}$-R/K/H-$X_{2\text{-}5}$-P-Y/L (where Φ_1 is a

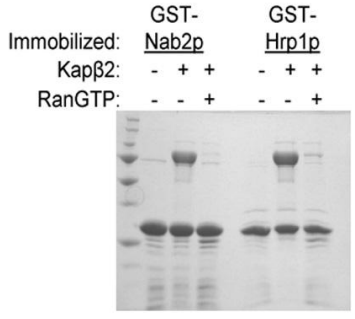


Figure 3-9. Binding assays of Kapβ2 with immobilized Nab2p-NLS and Hrp1p-NLS in the presence and absence of RanGTP. Bound proteins are Coomassie blue-stained.

hydrophobic residue and Φ_3 and Φ_4 are hydrophobic residues or R or K) (Lee, Cansizoglu et al. 2006) to search for potential hPY-NLSs within *S. cerevisiae* proteins in the UniProtKB/Swiss-Prot protein database (Bairoch, Boeckmann et al. 2004). A consensus sequence for the N-terminal motif of bPY-NLSs is not available due to lack of an apparent specific pattern. As a result, we modified a previously used sequence pattern that is consistent with the basic motifs of hnRNP M and PQBP-1 (Lee, Cansizoglu et al. 2006) to accommodate additional validated human bPY-NLSs and NLSs in Nab2p and Hrp1p (Table 3-1). The resulting sequence pattern K/R-X_{0-6}-K/R-X_{0-6}-K/R-X_{0-6}-K/R-X_{2-5}-R/K/H-X_{1-5}-PY/L is used to search for potential yeast bPY-NLSs. The resulting lists were filtered for structural disorder (Linding, Jensen et al. 2003) and overall basic character. Six hPY/L-containing fragments were tested but none bound Kap104p (Figure 3-10A and C). However, 11 of 20 bPY/L-containing fragments tested bound Kap104p and were dissociated by RanGTP (Figure 3-10A,D and Figure 3-11a,b). Two bPY/L-containing full length substrates, Tfg2p and Rml2p, were tested and both bound Kap104p and were dissociated by RanGTP (Figure 3-10D).

Of the eleven bPY/L-containing proteins in yeast that bound Kap104p, seven (or 64%) have been shown to be predominantly nuclear or show both nuclear and cytoplasmic localization. Thus, recognition of the basic subclass of PY-NLS is conserved between human and yeast. However, human Kapβ2 has evolved to recognize an additional hydrophobic PY-NLS subclass, enabling it to transport a broader range of substrates. Alternatively, Kap104p may have evolved to be more specific and lost its ability recognize hPY-NLS.

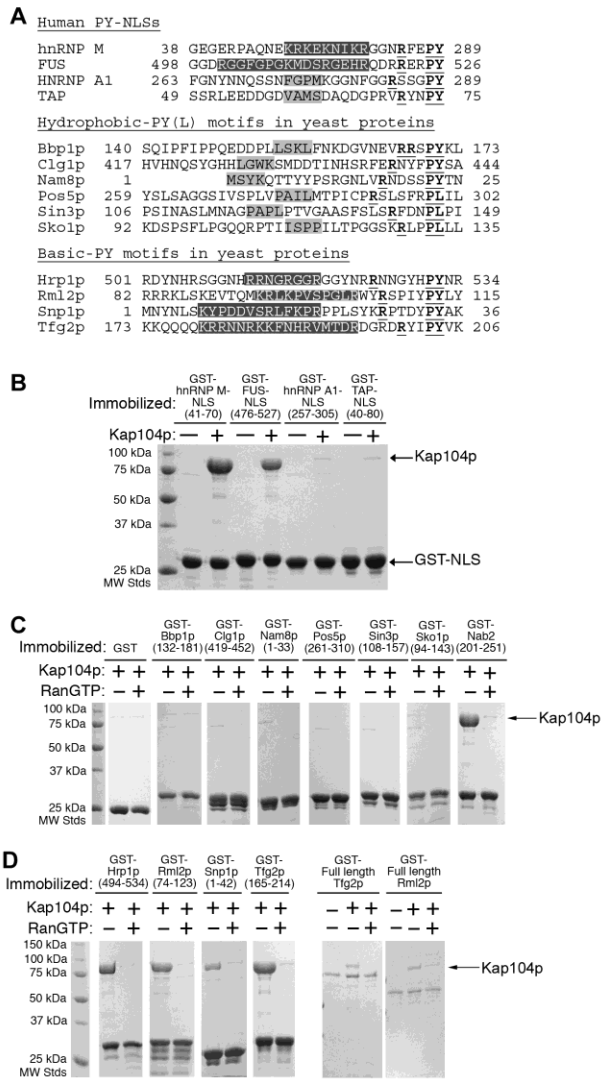


Figure 3-10. Kap104 recognizes basic-PY-NLSs, but not hydrophobic-PY-NLSs. (A) Sequences of known PY-NLSs in human proteins and predicted hydrophobic-PY(L) (shaded in light grey) and basic-PY motifs (shaded in dark grey) in yeast proteins. The $Rx_{2\text{-}5}PY(L)$ motif is underlined. (B) Binding assays of Kap104p and immobilized human bPY NLSs (hnRNP M and FUS) or human hPY NLSs (hnRNP A1 and TAP). (C) Binding assays of Kap104p with six immobilized predicted yeast hPY(L)-NLSs. GST and GST-Nab2p are included as controls. Faint bands at ~ 70kD are likely heat shock protein contaminants. (D) Binding assays of Kap104p and immobilized predicted bPY-NLSs or full length proteins. GST-Hrp1p is included as a control. Bound proteins are Coomassie blue-stained.

a

Basic-PL motifs in yeast proteins

```
Naf1p       269 QQRQRKKRDNRKLANDSDNVKVKRARQPKANSLPKLVPPLGMSSNAPMQH 318
Prb1p       125 GGCHENKVEEKKMKGKKVKGKKHHEKTLEKGFHHNRLAPLVSTAQFNPDA 174
Enp1p         1                          MARASSTKARKQRHDPLLKDLDAAQGT 27
Cdc25p      774 IDLKASSAASGSVFTPFNRPSHNRTFSRARVSKRKKKYPLTVDTLNTMKK 823
Sbp1p       138 GFRGGFRGGYRGGFRGRGNFRGRGGARGGFNGQKREKIPLDQMERSKDTL 187
Dig1p        66 SGDKEADHEDSETATAKKRKAQPLKNPKKSLKRGRVPAPLNLSDSNTNTH 115
Arp8p       114 NELGSSRDKRAPPAVQTSKRYKKYPKLDPAKAPPGKKVPLHLLEKRRLGR 163
Ste20p      226 SHSLSNPKHKQHKPKVKPSKPEAKSKPVSVKKSFPSKNPLKNSSPPKKQT 275
```

b

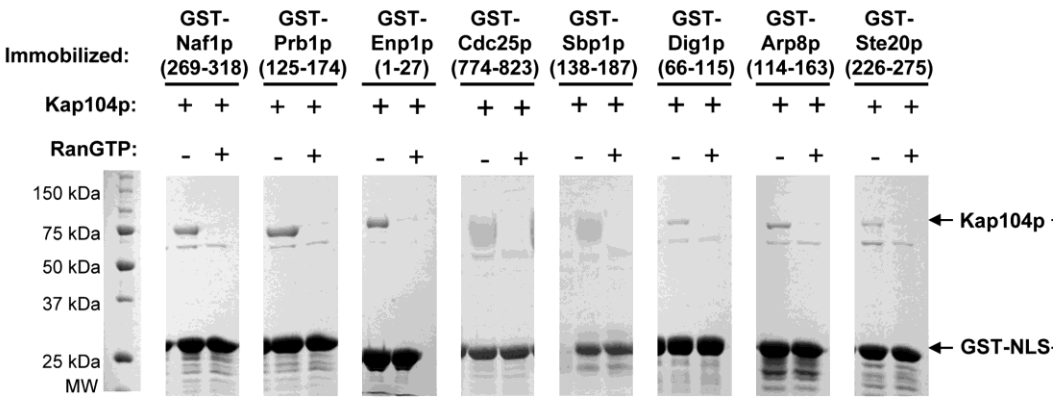


Figure 3-11. Kap104 recognizes potential basic-PL-NLSs. *(a)* The sequences of predicted basic-PL motifs in yeast proteins. Basic motifs are shaded in dark grey and the $Rx_{2\text{-}5}PY(L)$ motif is in bold and underlined. *(b)* Experimental testing of predicted bPL substrates. Kap104p is added to immobilized GST-NLSs in the presence or absence of RanGTP. Bound proteins are visualized with Coomassie blue. Faint bands at ~ 70kD are likely heat shock protein contaminants.

Kapβ2-NLS structures explain Kap104p subclass specificity.

Kapβ2 and Kap104p sequences were aligned and examined in the context of crystal structures of Kapβ2 bound to NLSs of hnRNPs A1 (hPY-NLS) and M (bPY-NLS) (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007). Kapβ2 has 20 HEAT repeats, each consisting of two antiparallel helices A and B. Both PY-NLSs bind the Kapβ2 interface lined with B helices of HEAT repeats 8-18 (abbreviated H8B-H18B), converging structurally at three spatially distinct binding sites: 1) overlapping portions of the N-terminal hydrophobic and the larger basic motifs, 2) the arginine residue and 3) the PY residues, both of the C-terminal $Rx_{2\text{-}5}PY$ motifs (Cansizoglu, Lee et al. 2007). Correspondingly, both structures share many common Kapβ2 interface residues, especially those that contact the conserved C-terminal $Rx_{2\text{-}5}PY$ motif (Figure 3-12A).

Nineteen of forty residues on the Kapβ2-NLS interface are different in yeast and human. Most of these residues are located at interfaces for the N-terminal motifs (H15-H17) and in the linkers between the N- and C-terminal motifs (H12-H14). Interestingly, in the N-terminal motif interfaces, all residues that differ between yeast and human contact the hydrophobic motif in hnRNP A1 but not the basic residues in hnRNP M (Figure 3-12A, Figure 3-13A and B). These include Kapβ2 residues I722, S723, N726, E734, T766 and I773 that are replaced with T, P, I, L, S and V, respectively, in yeast. Kapβ2 residues I722, S723 and T766 contact substrate M276 in the FGPM hydrophobic motif of hnRNP A1 (Figure 3-13A), and changes in yeast may compromise interaction with a hydrophobic residue. Kapβ2 residue N726 makes a polar contact with the backbone carbonyl of G274 in hnRNP A1, and the loss of backbone constraints upon the

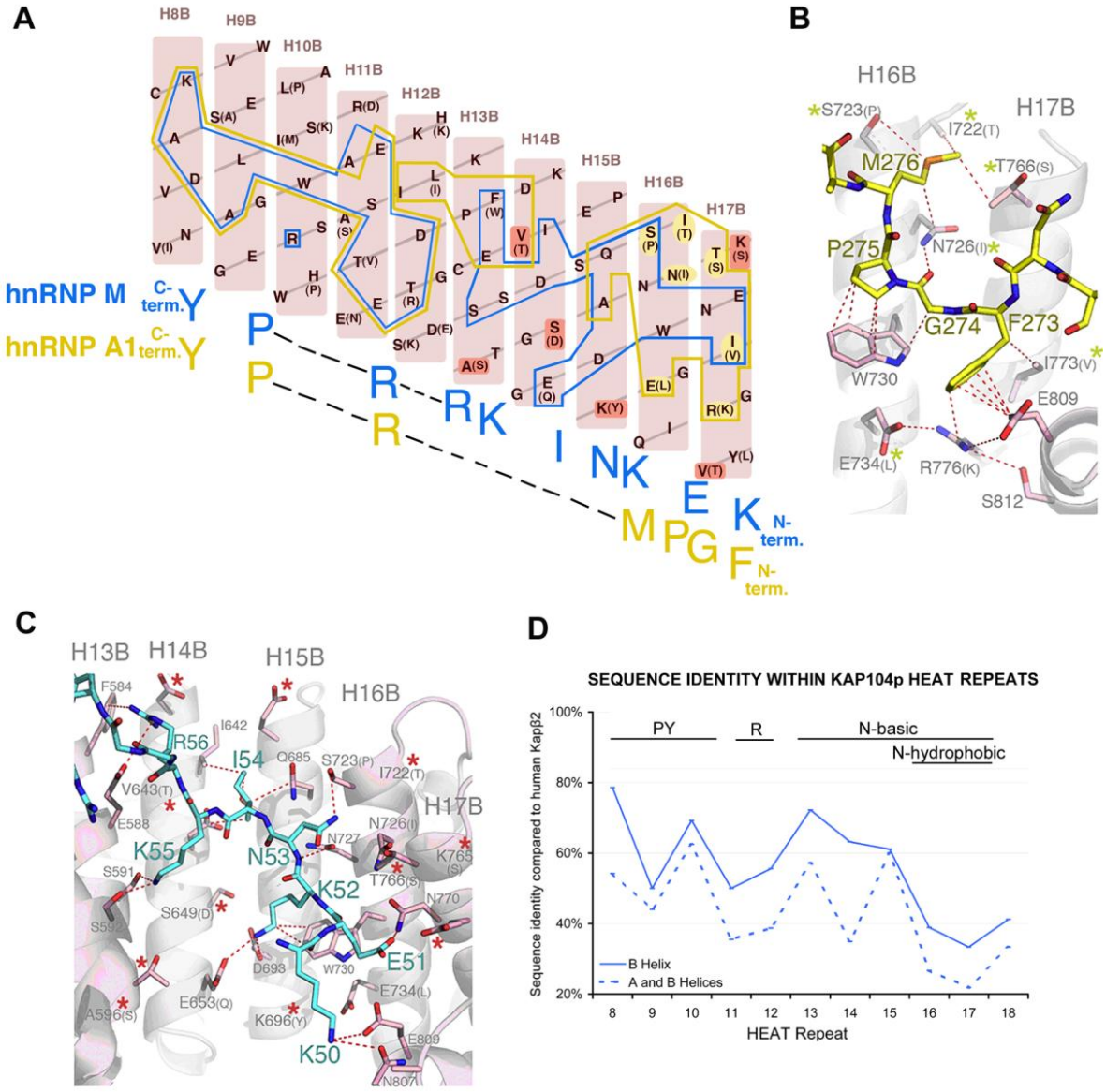


Figure 3-12. Kapβ2-NLS structures and Kap104p specificity. (A) Schematic representation of Kapβ2-NLS interface showing B helices of Kapβ2 H8-H17 (pink). Residues that are different in Kap104p are in parentheses. Kapβ2 residues that contact hnRNP A1-NLS (hPY-NLS) and hnRNP M-NLS (bPY-NLS) are outlined in yellow and blue, respectively. Residues contacting the N-terminal FGPM hydrophobic motif in hnRNP A1, which are also different in Kap104p, are highlighted in yellow. Residues that

increase electronegativity of the Kap104p surface are highlighted in red. (B) Interactions between Kapβ2 (pink) and the N-terminal hydrophobic motif of hnRNP A1 (yellow) (2H4M), drawn with PYMOL (DeLano 2002). Residues that are different in Kap104p are in parentheses (yellow asterisks label residues that may affect interactions with hPY-NLSs). (C) Interactions between Kapβ2 (pink) and the N-terminal basic motif of hnRNP M (blue) (2OT8). Residues that are different in Kap104p are in parentheses. Red asterisks label Kap104p substitutions that increase electronegativity. (D) Sequence identity within individual HEAT repeats of Kapβ2 and Kap104p. The motifs recognized by the B-helix of each HEAT repeat are specified above the graph.

change to isoleucine may disrupt hydrophobic motif recognition. Kapβ2 residues E653, K696, E734, R776 and E809 form an extensive charged network that appears to position the aliphatic portions of R776 and E809 for hydrophobic contacts with the F273 ring in hnRNP A1, while the Cδ of Kapβ2 I773 contacts Cβ of hnRNP A1 F273. In yeast, these contacts are lost as the charged network is disrupted and Cδ of I773 is no longer available for hydrophobic interaction in the yeast valine residue. In contrast, of the many Kapβ2 residues that contact basic sidechains in bPY-NLS, only E653 is different in yeast (Figure 3-13). In fact, several amino acid differences at the periphery of the NLS site in yeast Kap104p increase electronegativity, thus supporting bPY-NLS recognition.

Prediction of Kapβ2 homologs that recognize hPY-NLSs

Comparison of individual HEAT repeats of Kapβ2 and Kap104p showed high identity (~50%) at H8-H10 but the similarity dropped to ~20% at H17 (Figure 3-12D). B helices that line the interface are generally more conserved than the A outer helices. However, even in the former, sequence identities in H16B-H17B dipped significantly below 40% (Figure 3-12D). These observations suggest that both helical orientations and

interface functional groups are better conserved at recognition sites for the C-terminal PY motif (H8-H10) than at the N-terminal basic/hydrophobic motifs (H16-H17). Consequently, the loss of Kap104p recognition for the N-terminal hydrophobic motif is most likely due to critical interface residue changes in H16B-H17B and to changes in helical orientations in this region. We have aligned sequences of Kapβ2 homologs, tracked interface residues and potential overall helical similarities at the N-terminal hydrophobic motif interfaces in different organisms, and used this information to predict species in which Kapβ2 would recognize hPY-NLS.

Based on the sequence alignment of Kapβ2 homologs, we can track interface residues (Figure 3-13A) and potential overall helical similarities (Figure 3-13B) at the N-terminal hydrophobic motif interfaces in different organisms and use this information to predict species that recognize hPY-NLS. In higher eukaryotes of the Animalia Kingdom such as *X. laevis* and *D. melanogaster*, all seven Kapβ2 residues that contact the hydrophobic motif are conserved (Figure 3-13A) and overall sequence identities of H16-H17 are >60% (Figure 3-13B), suggesting recognition of the hydrophobic PY-NLSs. Organisms in the Plantae Kingdom such as *A. thaliana* and *O. sativa* have few amino acid substitutions (Figure 3-13A), maintain overall sequence similarities of ~50% in this region (Figure 3-13B), and are likely to recognize the hydrophobic motif as well. However, Kapβ2 sequences diverge in the Fungi Kingdom (Figure 3-13A) and overall sequence identities in H16-H17 dip below 30% (Figure 3-13B), suggesting changes in helical orientations; most fungi Kapβ2s will likely not recognize hPY-NLSs.

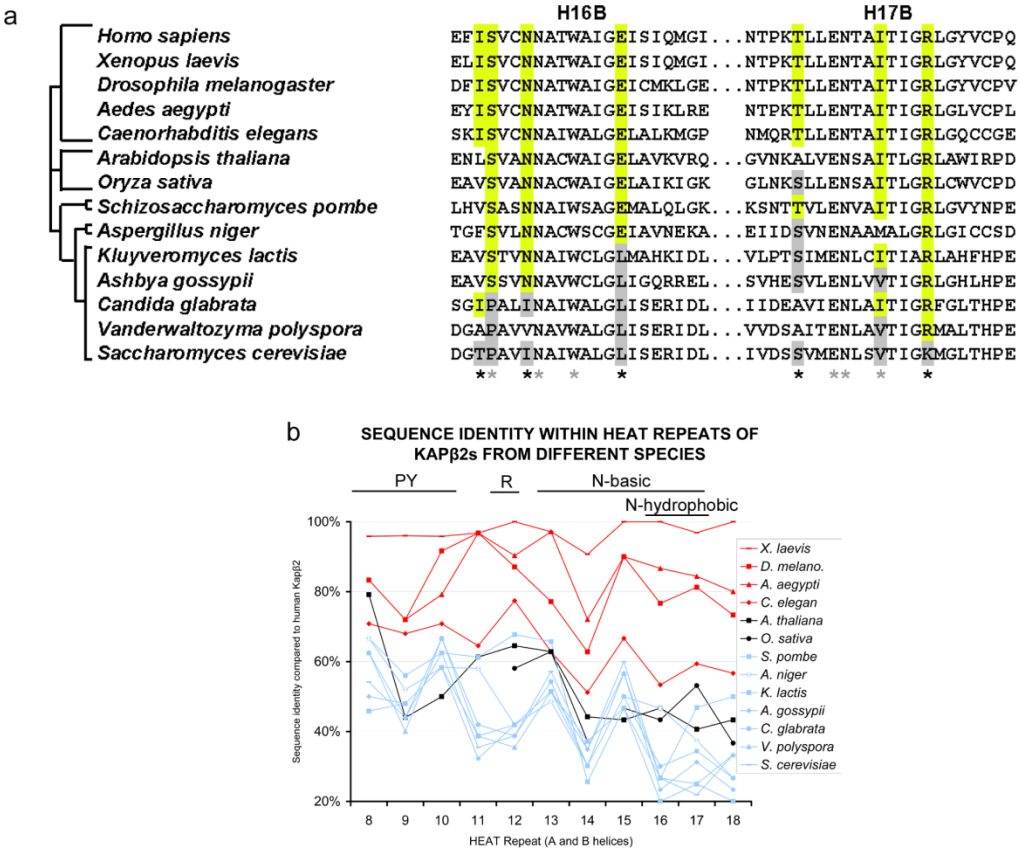


Figure 3-13. Kap104p specificity. (*a*) Alignment of Kapβ2 homologs showing residues contacting the hydrophobic motif of hPY-NLS (black asterisks) and the basic motif of bPY-NLS (grey asterisks). Contact residues that differ in *H. sapiens* (yellow) and *S. cerevisiae* (grey) are highlighted. (*b*) Sequence identity within HEAT repeats of Kapβ2s from different species. Organisms in the Animalia Kingdom are colored in red, Plantae Kingdom in black, and the Fungi Kingdom in blue. The motifs recognized by the B-helix of each HEAT repeat are specified above the graph.

Distribution of binding energy along the Hrp1p-NLS.

We have performed scanning alanine mutagenesis covering residues 506-532 of Hrp1p-NLS (Figure 3-7A, Table 3-2, Table 3-4). In the N-terminal region of Hrp1p-NLS, none of the four mutants ^{506}RSGG^{509}/AAAA, ^{512}RRNG^{515}/AAAA, ^{516}RGG^{518}/AAA and ^{519}RGGYN^{523}/AAAAA (Table 3-2) affected Kap104p binding, suggesting that this N-terminal basic-enriched region may contribute little to total binding energy. However, these mutations may be misleading as glycine to alanine mutations may decrease the entropy of the unbound NLS, thus decreasing the entropic penalty of binding and offsetting affinity loss from arginine mutations. Therefore, we also generated a quadruple mutant where all the arginines (R512, R513, R516 and R519) were mutated to alanines. This quadruple mutant decreased Kap104p binding by a marginal five-fold (Figure 3-14A and Table 3-2), suggesting that positive charges in the N-terminal basic region are somewhat important for Kapβ-NLS interaction. Quadruple mutant R512, R513, R516, R519/KKKK did not affect Kap104p binding (Table 3-2), further suggesting that stereospecific interactions with arginine guanido groups are not important for Kap104p binding.

Kap104p binding was not significantly affected when both arginine residues, ^{524}RR^{525}, in the C-terminal Rx$_{2\text{-}5}$PY motif of Hrp1p-NLS were mutated to alanines ($K_{Dmutant}/K_{Dwildtype}$ = 1.7; Figure 3-14A and Table 3-2). In contrast, the C-terminal ^{531}PY^{532}/AA mutation abolished detectable Kap104p binding (Table 3-2). The enthalpy of binding for all the PY-NLSs that we have measured by ITC are similar, and the weakest measurable K_D in this series was 10 μM (Cansizoglu, Lee et al. 2007).

Table 3-2. Summary of ITC data for Kap104p binding to Hrp1p mutants

Hrp1p	K_D[a] *(nM)*	ΔG[b] *(kcal/mol)*	ΔH *(kcal/mol)*	$T\Delta S$[c] *(kcal/mol/K)*	$\Delta\Delta G$[d] *(kcal/mol)*	$K_{Dmutant}/K_{Dwildtype}$
wildtype	32±16	-10.1±0.3	-10.0±0.5	0.09±0.7	—	—
Mutations in the N-terminal basic motif						
^{506}RSGG^{509}/AAAA	42±35	-10.0±0.6	-12.6±1.2	-2.6±0.6	0.02±0.6	1.3±1.1
^{512}RRNG^{515}/AAAA	48±26	-9.8±0.3	-8.3±0.5	1.5±0.1	0.2±0.3	1.5±0.8
^{516}RGG^{518}/AAA	24±18	-10.3±0.5	-8.3±2.0	2.0±2.4	-0.3±0.5	0.7±0.6
^{519}RGGYN^{523}/ AAAAA	17±6	-10.4±0.2	-14.2±1.9	-3.8±2.1	-0.4±0.2	0.5±0.2
R512A,R513A, R516A,R519A	194±128	-9.1±0.5	-9.4±1.3	-0.3±1.0	0.9±0.5	6.0±4.0
R512K,R513K, R516K,R519K	36±11	-10.0±0.2	-12.8±0.06	-2.8±0.1	-0.05±0.2	1.1±0.4
Mutations in the C-terminal $Rx_{2-5}PY$ motif						
R524A	37±4	-10.0±0.05	-11.7±1.5	-1.8±1.6	0.07±0.06	1.1±0.1
R525A	13±3	-10.5±0.1	-12.7±0.6	-2.2±0.5	-0.5±0.1	0.4±0.1
^{524}RR^{525}/AA	62±11	-9.6±0.1	-11.4±0.2	-1.8±0.3	0.4±0.1	1.9±0.3
Y529A	143±46	-9.2±0.2	-10.9±1.3	-1.7±1.4	0.8±0.2	4.4±1.4
Y529L	37±27	-10.0±0.4	-10.9±0.8	-0.9±0.4	-0.02±0.4	1.1±0.8
^{531}PY^{532}/AA	n.d.	n.d.	n.d.	n.d.	n.d.	>200
Mutations in two binding epitopes						
R512A, ^{524}RR^{525}/AA	44±28	-9.9±0.4	-9.9±0.4	0.06±0.1	0.1±0.4	1.4±0.9
R512A, Y529A	60±6	-9.7±0.05	-11.9±0.4	-2.2±0.5	0.4±0.05	1.9±0.2

[a] stoichiometry = 0.9-1.1; [b] $\Delta G = -RT\ln K_a$; [c] $T\Delta S = (\Delta H - \Delta G)$; [d] $\Delta\Delta G = -RT\ln(K_{d\_wt}/K_{d\_mut})$; n.d., not detectable; All binding experiments were performed 2-4 times (± SD).

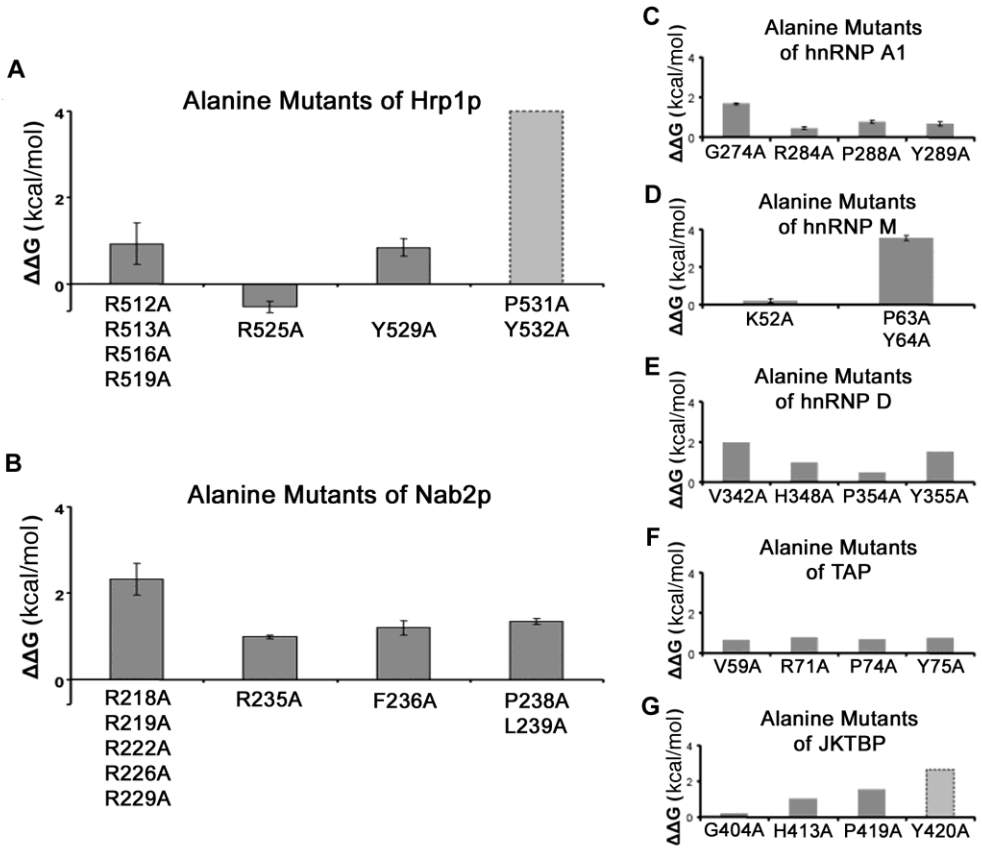


Figure 3-14. Mutagenic analyses of Hrp1p-NLS, Nab2p-NLS, and human PY-NLSs. (A, B) Loss of Kap104p binding energy in alanine mutants of Hrp1p (A) and Nab2p (B) ($\Delta\Delta G = -RT\ln(K_{D\_wt}/K_{D\_mut})$. (C-G) Loss of Kapβ2 binding energy in alanine mutants of PY-NLSs from hnRNP A1 (C), hnRNP M (D), hnRNP D (E), TAP (F), and JKTBP (G) ($\Delta\Delta G = -RT\ln(K_{D\_wt}/K_{D\_mut})$. K_{D_wt} and K_{D_mut} values for hnRNP A1 were obtained from (Lee, Cansizoglu et al. 2006), hnRNP M from (Cansizoglu, Lee et al. 2007), hnRNP D, TAP and JKTBP from (Imasaki, Shimizu et al. 2007). K_D values for Hrp1p, Nab2p, hnRNPs A1 and M were obtained by ITC whereas those for hnRNP D, TAP and JKTBP were obtained by surface plasmon resonance.

Therefore, we assume that the affinity of the Hrp1p ^{531}PY^{532}/AA mutant is likely weaker than 10 μM, and its $K_{Dmutant}/K_{Dwildtype}$ > 200 (Figure 3-14A). Thus, Hrp1p-NLS contains one strong binding hotspot at its PY motif, similar to the single significant hotspot at the C-terminal PY motif of human substrate hnRNP M ($K_{D,mutantPY/AA}/K_{D,wildtype}$ = 500 for hnRNP M-NLS) (Cansizoglu, Lee et al. 2007). Interestingly, we also located a modest binding hotspot at residue Y529 ($K_{Dmutant}/K_{Dwildtype}$ = 4 for Y529A; Figure 3-14A and Table 3-2) in the linker between the arginine and the PY of the Rx_{2-5}PY C-terminal motif. However, the Y529L mutation did not affect Kap104p binding (Table 3-2), suggesting that a hydrophobic, but not necessarily aromatic, moiety at this position might be important.

Distribution of binding energy along the Nab2p-NLS.

We have performed scanning alanine mutagenesis covering residues 210-239 of Nab2p-NLS (Figure 3-7A, Table 3-3 and Table 3-4). Binding energy along the Nab2p-NLS appears quite distributed compared to Hrp1p-NLS, with no single binding hotspot that stands out above others (Figure 3-14B and Table 3-3). In its basic N-terminal region, ^{216}KNRR^{219}, ^{222}RGG^{224} and ^{226}RGGRN^{230} were each mutated to alanines but only ^{216}KNRR^{219}/AAAA showed a small three-fold decrease in Kap104p affinity (Table 3-3). None of the single mutants K216A, R218A, R219A, R222A, R226A or R229A decreased Kap104p binding (Table 3-4), and simultaneous mutation of all the arginines to lysines also did not decrease Kap104p binding. In contrast, mutation of all five arginines to alanines decreased affinity by 60-fold (K_D = 2.25 μM; Figure 3-14B and Table 3-3), suggesting that the collective basic character of this region contributes significantly to

Table 3-3. Summary of ITC data for Kap104p binding to Nab2p mutants

Nab2p	K_D[a] *(nM)*	ΔG[b] *(kcal/mol)*	ΔH *(kcal/mol)*	$T\Delta S$[c] *(kcal/mol/ K)*	$\Delta\Delta G$[d] *(kcal/mol)*	$K_{Dmutant}/K_{D\ wildtype}$
wildtype	37±20	-10.0±0.3	-11.3±0.5	-1.3±0.6	—	—
Mutations in the N-terminal basic motif						
^{210}GGG^{212}/AAA	19±6	-10.4±0.2	-9.8±1.4	0.5±1.6	-0.4±0.2	0.5±0.2
^{216}KNRR^{219}/ AAAA	109±52	-9.6±0.3	-8.2±1.0	1.2±1.3	0.6±0.3	3.0±1.4
^{222}RGG^{224}/AAA	58±11	-9.7±0.1	-9.6±0.7	0.06±0.8	0.3±0.1	1.6±0.3
^{226}RGGRN^{230}/ AAAAA	59±21	-9.7±0.2	-9.9±0.6	-0.2±0.4	0.3±0.2	1.6±0.6
R218A,R219A, R222A,R226A, R229A	2250±1344	-7.6±0.4	-4.9±0.3	2.7±0.06	2.3±0.4	60.8 ±36.3
R218K,R219K, R222K,R226K, R229K	21±5	-10.3±0.1	-9.9±1.3	0.4±1.1	-0.4±0.1	0.6±0.1
Mutations in the C-terminal $Rx_{2-5}PY$ motif						
R235A	203±14	-8.9±0.04	-5.2±0.9	3.7±0.9	1.0±0.04	5.5±0.4
R235K	37±6	-9.9±0.1	-5.3±0.4	4.6±0.3	-0.01±0.1	1.0±0.2
R235H	63±42	-9.7±0.4	-8.4±0.4	1.3±0.6	0.2±0.4	1.7±1.1
F236A	296±81	-8.7±0.2	-7.4±0.3	1.4±0.09	1.2±0.2	8.0±2.2
F236L	66±32	-9.7±0.4	-8.4±0.2	1.3±0.5	-0.3±0.3	1.8±0.9
^{238}PL^{239}/AA	376±42	-8.6±0.07	-5.8±1.3	2.8±1.3	-1.3±0.1	10.2±1.2
Mutations in two binding epitopes						
R222A, ^{238}PL^{239}/AA	411±40	-8.5±0.06	-6.5±0.3	2.1±0.4	1.4±0.06	11.1±1.1
R235A, ^{238}PL^{239}/AA	544±4	-8.4±0.01	-5.6±0.1	2.8±0.09	1.6±0.004	14.7±0.1
R218A,R219A, R222A,R226A, R229A,P238A, L239A	n.d.	n.d.	n.d.	n.d.	n.d.	>200

[a] stoichiometry = 0.9-1.1; [b] $\Delta G = -RT\ln K_a$; [c] $T\Delta S = (\Delta H - \Delta G)$; [d] $\Delta\Delta G = -RT\ln(K_{d\_wt}/K_{d\_mut})$; All binding experiments were performed 2-4 times (± SD)

Table 3-4. Summary of ITC data for Kap104p binding to additional Hrp1p and Nab2p mutants

Hrp1p	K_D[a] *(nM)*	ΔG[b] *(kcal/mol)*	ΔH *(kcal/mol)*	$T\Delta S$[c] *(kcal/mol/K)*	$\Delta\Delta G$[d] *(kcal/mol)*	$K_{Dmutant}/K_{D\ wildtype}$
N510A, H511A[e]	4±2	-11.3±0.3	-13.8±0.2	-2.6±0.3	-1.2±0.4	0.1±0.8
R512A	19±4	-10.3±0.1	-10.5±1.0	0.2±1.0	-0.3±0.1	0.6±0.1
R513A[e]	11±5	-10.6±0.2	-10.0±0.2	0.6±0.3	-0.6±0.4	0.3±0.7
R516A[e]	7±1	-10.9±0.1	-10.7±0.08	0.2±0.1	-0.9±0.3	0.2±0.4
N526A, N527A[e]	17±9	-10.4±0.3	-11.2±0.3	-0.8±0.4	-0.3±0.4	0.5±0.8
H530A	55±12	-9.7±0.1	-9.0±1.0	0.7±1.1	0.3±0.1	1.7±0.4
Nab2p	K_D[a] *(nM)*	ΔG[b] *(kcal/mol)*	ΔH *(kcal/mol)*	$T\Delta S$[c] *(kcal/mol/K)*	$\Delta\Delta G$[d] *(kcal/mol)*	$K_{Dmutant}/K_{D\ wildtype}$
V214A	14±8	-10.5±0.3	-13.1±0.3	-2.5±0.4	-0.6±0.4	0.4±0.9
K216A	22±4	-10.2±0.1	-12.9±1.3	-2.6±1.4	-0.3±0.1	0.6±0.1
R218A	28±6	-10.1±0.1	-11.9±0.5	-1.8±0.4	-0.2±0.1	0.8±0.2
R219A	26±4	-10.2±0.07	-12.0±0.4	-1.9±0.3	-0.2±0.08	0.7±0.1
R222A	27±10	-10.1±0.2	-10.7±0.9	-0.6±0.6	-0.2±0.2	0.7±0.3
R226A	18±17	-10.5±0.7	-11.8±0.3	-1.2±1.0	-0.6±0.7	0.5±0.5
R229A	12±10	-10.7±0.6	-11.7±0.7	-1.0±0.2	-0.8±0.5	0.3±0.3
N230A, N231A, N232A	14±2	-10.5±0.1	-12.5±1.5	-2.0±1.6	-0.4±0.3	0.4±0.006
N237A[e]	30±7	-10.0±0.1	-10.2±0.1	-0.2±0.2	-0.1±0.3	0.8±0.5
P238A	17±4	-10.4±0.1	-6.6±0.03	3.7±0.09	-0.5±0.1	0.5±0.1

[a] stoichiometry = 0.9-1.1; [b] $\Delta G = -RT\ln K_a$; [c] $T\Delta S = (\Delta H - \Delta G)$; [d] $\Delta\Delta G = -RT\ln(K_{d\_wt}/K_{d\_mut})$; n.d., not detectable;

All binding experiments were performed 2-4 times (± SD) excluding those marked ([e]), which were measured once.

total binding energy of the NLS. Comparison of single arginine to alanine mutants ($K_{Dmutant}/K_{Dwildtype}$ values ~1.0) to the penta-mutant R218, R219, R222, R226, R229/AAAAA ($K_{Dmutant}/K_{Dwildtype}$ = 60.8) indicated a binding cooperativity of at least 60-fold within the N-terminal basic motif of Nab2p.

When R235 of the Nab2p C-terminal $Rx_{2-5}PL$ motif was mutated to an alanine, Kap104p affinity decreased by five-fold (Figure 3-14B and Table 3-3). Crystal structures of Kapβ2 bound to NLSs of hnRNPs A1 and M showed the equivalent arginine residues making electrostatic interactions with numerous aspartate and glutamate residues, suggesting the importance of a positively charged residue at this position (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007). We also mutated R235 to lysine and histidine, but neither mutant affected Kap104p binding significantly ($K_{Dmutant}/K_{Dwildtype}$ are 1.0 and 1.7, respectively; Table 3-3). The C-terminal $^{238}PL^{239}$/AA mutation in Nab2p-NLS decreased Kap104p binding by 10-fold (Figure 3-14B and Table 3-3). The energetic significance of this mutation suggests its equivalence to the PY motif in human Kapβ2 substrates and in Hrp1p. Furthermore, the Nab2p $^{238}PL^{239}$/PY mutant bound Kap104p with slightly higher affinity at K_D value of 13 nM. Mutagenesis of residue L239 to all other amino acids is described below.

The measurable $^{238}PL^{239}$/AA mutation in Nab2p-NLS (K_D = 376 nM) provided an opportunity to explore cooperativity across binding sites or epitopes. Mutations in the Nab2p triple mutant R222A, $^{238}PL^{239}$/AA (K_D 411 nM; $K_{Dmutant}/K_{Dwildtype}$ = 11.1; Table 3-3) show almost perfect additivity (0.9 ± 0.3) when compared to single R222A mutant (did not affect Kap104p binding; Table 3-4) and double mutant $^{238}PL^{239}$/AA ($K_{Dmutant}/K_{Dwildtype}$ = 10.2; Table 3-3). A second Nab2p triple mutant R235A, $^{238}PL^{239}$/AA

(K_D 544 nM; $K_{Dmutant}/K_{Dwildtype}$ = 14.7; Table 3-3) was also compared to single R235A mutant ($K_{Dmutant}/K_{Dwildtype}$ = 5.5; Table 3-4) and double ^{238}PL^{239}/AA mutant ($K_{Dmutant}/K_{Dwildtype}$ = 10.2; Table 3-3). Strict additivity between the R and the PL sites would give a calculated $K_{Dmutant}/K_{Dwildtype}$ of 56.1 for the triple mutant. Thus, the experimental $K_{Dmutant}/K_{Dwildtype}$ of 14.7 for the triple mutant indicated cooperativity of 3.8±0.5-fold between the two epitopes. Similarly, Hrp1p triple mutant R512A, ^{524}RR^{525}/AA and double mutant R512A, Y529A showed cooperativity of approximately 1.4±0.9 and 2.3±0.8-fold between epitopes, respectively. The couplings between binding epitopes observed here for both Nab2p and Hrp1p are still more than an order of magnitude lower than that observed within the N-terminal basic region of Nab2p (cooperativity >60-fold).

We also located a new binding hotspot at F236 in Nab2p ($K_{Dmutant}/K_{Dwildtype}$ = 8 for F236A; Figure 3-14B and Table 3-3), which is located in the linker between the R and the PL of the Rx_{2-5}PL C-terminal motif. This site is analogous to Y529 of Hrp1p discussed in the previous section and both residues are located two residues N-terminal of the PY/L motifs. As in the Hrp1p-NLS Y529L mutant, mutation of Nab2p F236 to leucine did not affect Kap104p binding (Table 3-3). Aromatic or hydrophobic residues occur at this position in many human PY-NLSs including hnRNPs M, D, F, JKTBP, TAP, HMBA-inducible protein, PABP2, PQBP-1, RB15B and WBS-16 (Truant, Kang et al. 1999; Suzuki, Iijima et al. 2005; Lee, Cansizoglu et al. 2006; Imasaki, Shimizu et al. 2007). Aromatic sidechains at this position overlap in the crystal structures of Kapβ2 bound to the NLSs of hnRNPs M, D and TAP (Cansizoglu, Lee et al. 2007; Imasaki, Shimizu et al. 2007). F61 of hnRNP M-NLS, Y352 of hnRNP D-NLS and Y72 of TAP-

NLS all make hydrophobic interactions with Kapβ2 W460A and with the backbones of the PY motifs. A hydrophobic residue here may contribute to binding energy through both favorable enthalpy and decrease of entropic penalty upon binding by preorganizing the PY motif. Thus, if present, a hydrophobic residue here may be considered as an extension of the PY motif.

Hrp1p contains a single very significant binding hotspot at its PY motif. In contrast, binding energy in Nab2p is more evenly distributed across its N-terminal basic region and the R, F and PL residues of its C-terminal consensus motif. Thus, distributions of binding energy in the two yeast NLSs are very different. From N- to C-terminus, energetic distribution across the three epitopes (N-terminal basic region, R and PY/L of the C-terminal motif) of Hrp1p and Nab2p can be roughly described as medium-weak-strong and strong-medium-medium, respectively ($\Delta\Delta G < 0.9$ kcal/mol is categorized as weak, $0.9 \leq \Delta\Delta G \leq 1.7$ kcal/mol as medium and $\Delta\Delta G > 1.7$ kcal/mol as strong; Figure 3-14 A and B). Similarly, in previously characterized PY-NLSs of hnRNPs A1, D, TAP and JKTBP (Lee, Cansizoglu et al. 2006; Imasaki, Shimizu et al. 2007), energetic distributions at the three epitopes are also quite varied, with rough patterns of strong-weak-weak, strong-medium-medium, weak-weak-weak and weak-medium-strong, respectively (Figure 3-14 C-G). In summary, all three PY-NLS epitopes are energetically highly variable; the N-terminal basic/hydrophobic and the C-terminal PY motifs appear to cover the entire energetic continuum from strong to weak and the arginine of the $Rx_{2\text{-}5}PY$ motif is medium to weakly energetically significant.

Degeneracy of tyrosine in the C-terminal PY motif.

Of the >20 sequences that bind Kapβ2 and Kap104p (Table 3-1) (Lee, Cansizoglu et al. 2006), two do not contain the PY dipeptide in their C-termini. HuR has a PG and Nab2p has a PL, thus raising the question of degeneracy at this C-terminal position. We mutated Y532 in the PY motif of Hrp1p to the other 19 amino acids (Figure 3-15A, Table 3-5). Only Y532F, Y532H and Y532M showed measurable Kap104p binding by ITC. Y532F best resembles the wildtype, with only a four-fold decrease in Kap104p affinity. Both Y532H and Y532M in Hrp1p bound significantly weaker with K_D values of 1 μM and 2 μM, respectively.

We also mutated L239 in the Nab2p PL motif to the other 19 amino acids (Figure 3-15B, Table 3-6). Binding energy along the Nab2p-NLS is very evenly distributed compared to Hr1p1p-NLS with the Nab2p ^{238}PL^{239}/AA mutation decreasing affinity only 10-fold compared to the >200-fold effect in Hrp1p. Thus, in the energetically distributed Nab2p-NLS, changes in the L239 position may be quite permissive. This is indeed the case since only L239D and L239E showed significant affinity decrease of 11- and 7-fold, respectively. L239G, L239I and L230P showed modest 3-4 fold affinity decrease. None of the other mutants (to S, T, N, Q, K, R, V, M, F, Y, W and H) decreased Kap104p binding.

Tyrosine is clearly the most preferred residue in the last position of the Hrp1p-NLS. Correspondingly, mutation of the PL motif in Nab2p to PY improves Kap104p binding. These results suggest that in general, tyrosine may be the most preferred and thus likely the most prevalent amino acid found in the last position of PY-NLSs (Table

3-1). It appears that if the PY site is energetically very significant, such as in Hrp1p, the residue type allowed at the terminal position is quite restrictive, with only 2-4 residues (Y, F, H and M) allowed. However, when the same motif is fairly silent energetically, such as in Nab2p and hnRNP A1 (Lee, Cansizoglu et al. 2006), the distribution of allowed amino acids in the terminal position is likely much wider, with only 2-5 residues disallowed.

Table 3-5 Summary of ITC data for mutations of the PY motif of Hrp1p

Hrp1p[e f]	K_D[a] *(nM)*	ΔG[b] *(kcal/mol)*	ΔH *(kcal/mol)*	$T\Delta S$[c] *(kcal/mol/K)*	$\Delta\Delta G$[d] *(kcal/mol)*	$K_{Dmutant}/K_{Dwt}$
^{531}PY^{532} (wildtype)	32±16	-10.1±0.3	-10.0±0.5	0.09±0.7	—	—
^{531}PY^{532}/PM	2237±400	-7.6±0.1	-4.4±0.2	3.2±0.2	2.5±0.3	69.9±0.4
^{531}PY^{532}/PF	116±30	-9.2±0.2	-8.0±0.2	1.2±0.2	0.8±0.3	3.6±0.5
^{531}PY^{532}/PH	1002±350	-8.0±0.2	-8.8±0.7	-0.7±0.7	2.0±0.4	27.1±0.6

[a] stoichiometry = 0.9-1.1; [b] $\Delta G = -RT\ln K_a$; [c] $T\Delta S = (\Delta H - \Delta G)$; [d] $\Delta\Delta G = -RT\ln(K_{d\_wt}/K_{d\_mut})$; [e] single experiment for each mutant (errors from curve fitting); [f] No heat was detected when Y532 was mutated to amino acids G, A, C, S, T, D, E, N, Q, K, R, V, L, I, P or W.

Table 3-6 Summary of ITC data for mutations of the PL motif of Nab2p

Nab2p	K_D[a] *(nM)*	ΔG[b] *(kcal/mol)*	ΔH *(kcal/mol)*	$T\Delta S$[c] *(cal/mol/K)*	$\Delta\Delta G$[d] *(kcal/mol)*	$K_{Dmutant}/K_{Dwt}$
^{238}PL^{239} (wt)[e]	37±20	-10.0±0.3	-11.3±0.5	-1.3±0.6	—	—
^{238}PL^{239}/PG	99±25	-9.4±0.1	-10.4±0.2	-1.1±0.3	0.6±0.3	2.7±0.5
^{238}PL^{239}/PA[e]	26±9	-10.2±0.2	-9.7±0.9	0.5±1.1	-0.2±0.2	0.7±0.3
^{238}PL^{239}/PC	32±11	-10.0±0.2	-10.6±0.2	-0.6±0.3	-0.08±0.3	0.9±0.6
^{238}PL^{239}/PS	27±16	-9.6±0.3	-13.6±0.5	-4.0±0.6	-0.2±0.4	0.7±0.9
^{238}PL^{239}/PT	15±6	-10.5±0.2	-11.0±0.2	-0.6±0.3	-0.5±0.4	0.4±0.7
^{238}PL^{239}/PD	412±77	-8.5±0.1	-7.6±0.2	1.0±0.2	1.4±0.3	11.1±0.5
^{238}PL^{239}/PE	253±54	-8.8±0.1	-5.3±0.2	3.5±0.2	1.1±0.3	6.8±0.5
^{238}PL^{239}/PN	61±32	-9.6±0.3	-11.3±0.4	-1.7±0.5	0.3±0.4	1.7±0.8
^{238}PL^{239}/PQ	20±4	-10.3±0.1	-11.6±0.1	-1.3±0.2	-0.4±0.3	0.5±0.5
^{238}PL^{239}/PK	18±5	-10.3±0.2	-9.7±0.1	0.6±0.2	-0.4±0.3	0.5±0.6
^{238}PL^{239}/PR	24±9	-10.2±0.2	-11.8±0.2	-1.7±0.3	-0.3±0.3	0.7±0.7
^{238}PL^{239}/PV	70±11	-9.6±0.1	-10.2±0.1	-0.6±0.1	0.4±0.3	1.9±0.5
^{238}PL^{239}/PI[e]	103±30	-9.3±0.2	-8.9±0.5	0.4±0.3	0.6±0.2	2.8±0.8
^{238}PL^{239}/PM	43±7	-9.8±0.1	-11.6±0.09	-1.7±0.1	0.09±0.3	1.2±0.5
^{238}PL^{239}/PF	44±7	-9.8±0.1	-8.3±0.06	1.6±0.1	0.1±0.3	1.2±0.5
^{238}PL^{239}/PY	13±4	-10.5±0.2	-15.1±0.1	-4.6±0.2	-0.6±0.3	0.4±0.6
^{238}PL^{239}/PW	21±9	-10.3±0.2	-12.0±0.3	-1.8±0.4	-0.3±0.4	0.6±0.8
^{238}PL^{239}/PH	41±7	-9.9±0.1	-9.7±0.07	0.1±0.1	0.06±0.3	1.1±0.5
^{238}PL^{239}/PP	159±26	-9.1±0.1	-6.9±0.1	2.2±0.2	0.8±0.3	4.3±0.5

[a] stoichiometry = 0.8-1.1; [b] $\Delta G = -RT\ln K_a$; [c] $T\Delta S = (\Delta H - \Delta G)$; [d] $\Delta\Delta G = -RT\ln(K_{d\_wt}/K_{d\_mut})$; [e] At least two binding experiments were performed (±SD). All other experiments were performed once (errors from curve fitting).

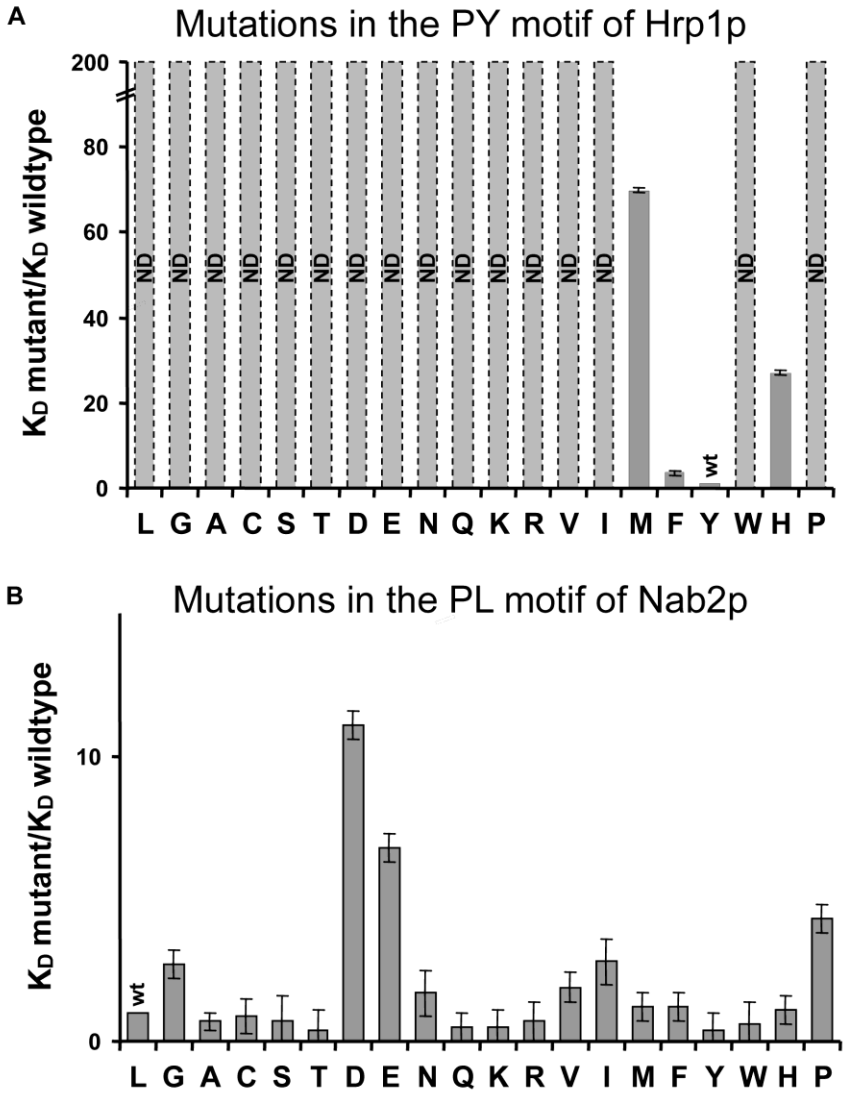


Figure 3-15. Mutagenic analysis of the PY(L) motif. (A) Mutations in the PY motif of Hrp1p and the resulting fold decrease in binding affinity for Kap104p. Only M, F, H can substitute for the Y. All other mutations had no detectable (ND) binding by ITC. (B) Mutations in the PL motif of Nab2p and the resulting fold decrease in binding affinity for Kap104p.

Hrp1p and Nab2p mutants are mislocalized in vivo.

To examine the effect of PY-NLS mutations on nucleocytoplasmic localization of Hrp1p and Nab2p *in vivo*, we expressed GFP-tagged full length Hrp1p and Nab2p wildtype and mutant proteins in yeast. Wildtype GFP-Hrp1p and Nab2p-GFP are localized in the nucleus as has been previously reported (Figure 3-16 A and B and Figure 3-17 A and B) (Aitchison, Blobel et al. 1996; Truant, Fridell et al. 1998; Lee and Aitchison 1999). Mutations in the C-terminal PY motif (^{531}PY^{532}/AA) of Hrp1p, which abolished detectable Kap104p binding, resulted in mislocalization of the GFP-fusion protein to the cytoplasm (Figure 3-16 A and B). The N-terminal basic motif of Hrp1p is also important for nuclear localization of Hrp1p: the R512,R513,R516,R519/AAAA mutant, which decreased Kap104p binding by a marginal five-fold, is also mislocalized (Figure 3-16 A and B). Xu and Henry have previously shown that substitutions of R516 and R519 with glutamines mislocalized Hrp1p, but proteins with lysine substitutions are properly localized (Marfatia, Crafton et al. 2003; Xu and Henry 2004). This further suggests that basic charges rather than stereospecific interactions are necessary for Kap104p interactions.

In the case of Nab2p, mutations in either the N-terminal motif (penta-mutant R218,R219,R222,R226,R229/AAAAA; decreases Kap104p binding by 60-fold) or the C-terminal PY motif (^{238}PL^{239}/AA; decreases Kap104p binding by 10-fold) resulted in increased cytoplasmic localization of the GFP-fusion protein (Figure 3-17 A and B). Arginine methylation of Nab2p by Hmt1p is required for its export from the nucleus, possibly explaining some nuclear accumulation of the N-terminal mutant despite its low

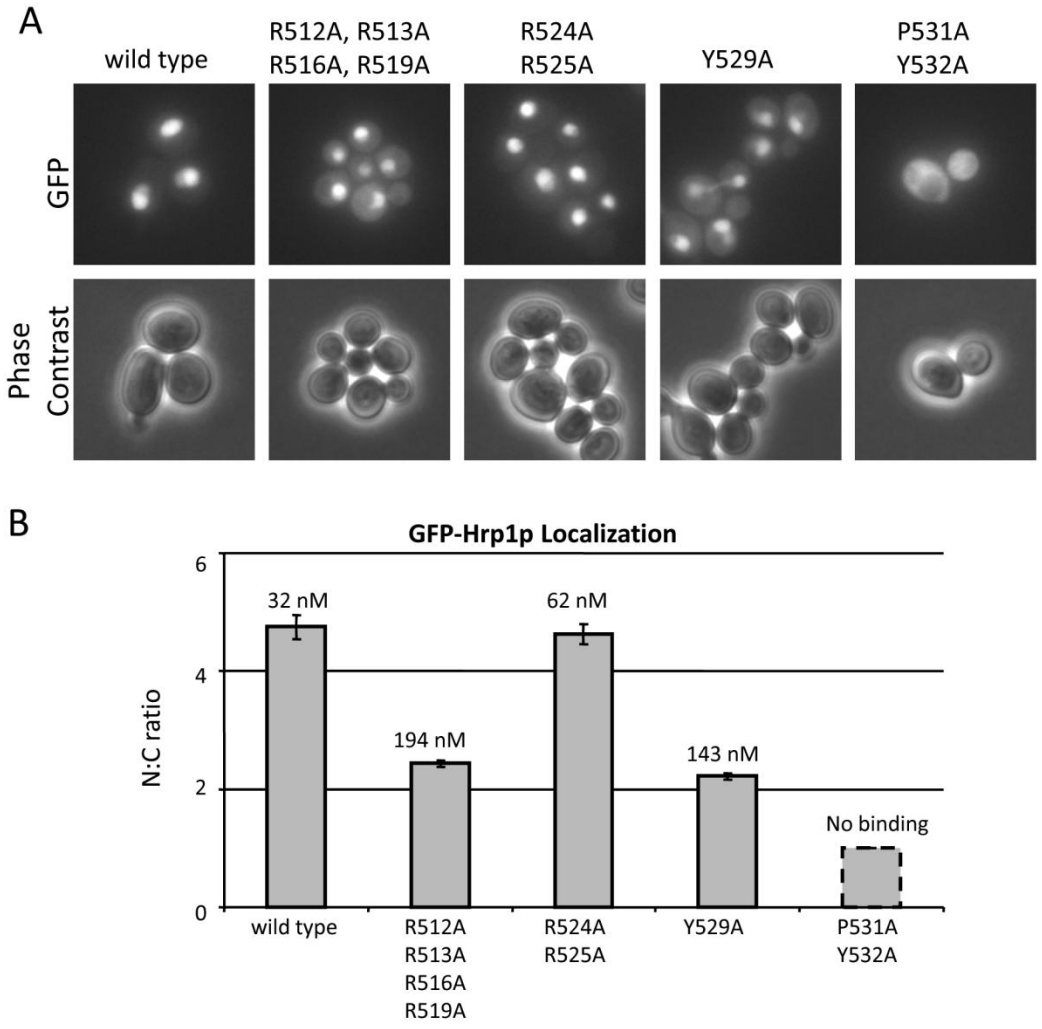


Figure 3-16. Hrp1p mutants are mislocalized *in vivo*. (A) *S. cerevisiae* cells expressing either wild type or mutant full length GFP-Hrp1p fusion proteins were analyzed by fluorescence microscopy and phase contrast. GFP is displayed in the same fluorescence scale in each panel. (B) Mean pixel values were used to determine the nuclear:cytoplasmic (N:C) ratio of fluorescence intensity for either GFP-Hrp1pfusion proteins (±SEM). Dashed lines indicate an estimated N:C ratio of 1:1 due to the diffuse nuclear and cytoplasmic localization of the fusion protein.

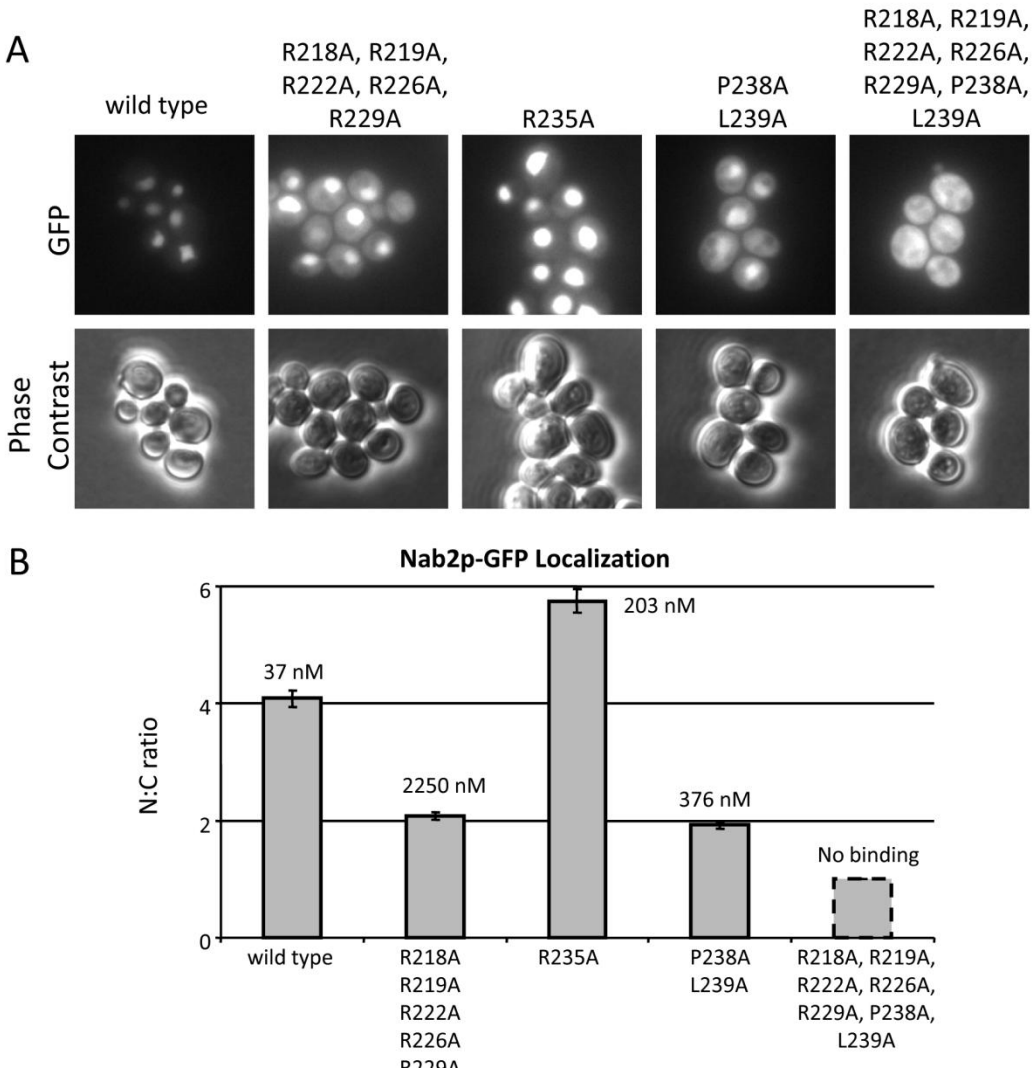


Figure 3-17. Nab2p mutants are mislocalized *in vivo*. (A) *S. cerevisiae* cells expressing either wild type or mutant full length Nab2p-GFP fusion proteins were analyzed by fluorescence microscopy and phase contrast. GFP is displayed in the same fluorescence scale in each panel. (B) Mean pixel values were used to determine the nuclear:cytoplasmic (N:C) ratio of fluorescence intensity for either Nab2p-GFP fusion proteins (±SEM). Dashed lines indicate an estimated N:C ratio of 1:1 due to the diffuse nuclear and cytoplasmic localization of the fusion protein.

affinity for Kap104p (Green, Marfatia et al. 2002; Marfatia, Crafton et al. 2003). Combined mutations of both the N- and C-terminal motifs resulted in diffuse localization of the fusion protein, consistent with further affinity reduction for Kap104p (Table 3-3 and Figure 3-17 A and B). We have shown here that mutations in the PY-NLSs of Hrp1p and Nab2p that decrease binding affinity to Kap104 also affect nuclear localization in yeast cells.

Discussion

The problem of deciphering the sequence code for substrate recognition by Kapβ2 is interesting and challenging because the transport factor exhibits obvious biologically relevant specificity for nuclear import substrates but at the same time, is able to handle a large number of different sequence-diverse substrates. Previous studies have captured the requirement for structural disorder in NLSs and the notion of a few anchoring amino acids such as the N-terminal hydrophobic/basic and $Rx_{2-5}PY$ motifs (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007). Here, we show that yeast Kap104p is a PY-NLS recognizing homolog specific for the basic subclass of this signal, and that the two different Kap104p substrates have rather different distributions of binding energy for Kap104p. The NLS in Hrp1p largely uses the PY motif and the NLS in Nab2p uses many positions distributed across three binding regions. Consistent with this, the Y position of the PY motif shows more degeneracy in Nab2p than in Hrp1p. Based on all this and on thermodynamic data from five human PY-NLSs (Lee, Cansizoglu et al. 2006;

Cansizoglu, Lee et al. 2007; Imasaki, Shimizu et al. 2007), we propose the following physical properties that govern affinity of PY-NLSs recognition by Kapβ2:

1. PY-NLSs contain at least three energetically significant binding epitopes

Structures of PY-NLSs from hnRNPs A1, M, D, TAP and JKTBP converge spatially at three distinct binding sites or epitopes separated by structurally variable linkers: 1) the N-terminal hydrophobic/basic motif, 2) the arginine residue of the C-terminal $Rx_{2-5}PY$ sequence motif and 3) the PY of the C-terminal $Rx_{2-5}PY$ motif (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007; Imasaki, Shimizu et al. 2007). We have shown here that all three structural epitopes can be energetically significant.

The N-terminal basic-enriched motifs of Hrp1p- and Nab2p-NLSs constitute epitope-1, where collective basic character and likely charge density drive Kap104p binding. Mutations of all the arginines in this region to alanines decreased binding energy by 0.9-2.3 kcal/mol for both NLSs. Similarly, the N-terminal hydrophobic motif of hnRNP A1-NLS and the equivalent region of hnRNP D-NLS that contains both hydrophobic and basic residues are also energetically significant, with mutations decreasing binding energy by ~ 2 kcal/mol (Cansizoglu, Lee et al. 2007).

Epitopes 2 and 3 are contained within the C-terminal $Rx_{2-5}PY/L$ sequence motifs. Two linkers of variable lengths, compositions and structures connect epitope-1 to epitope-2, and epitope-2 to epitope-3 (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007). Epitope-2 is located at Hrp1p $^{524}RR^{525}$ and Nab2p R235 at the first consensus position of the C-terminal $Rx_{2-5}PY/L$ sequence motifs. Of the three PY-NLS epitopes, epitope-2 tends to contribute the least to binding energy, with mutations decreasing

binding energy maximally by ~ 1 kcal/mol in Nab2p, hnRNP D and JKTBP (Figure 3-14 B, E and G).

Epitope-3 is located at Hrp1p ^{531}PY^{532} and Nab2p ^{238}PL^{239}. Mutations at these terminal positions are generally energetically significant, decreasing binding energy by 1.3 to 4 kcal/mol in Hrp1p, Nab2p, hnRNPs M, D and JKTBP. However, exceptions are seen in hnRNP A1 and TAP, where PY mutations decreased binding modestly by only ~ 0.7 kcal/mol.

Since free PY-NLSs are structurally disordered and adopt extended Kapβ2-bound conformations, epitopes 1-3 are presented as peptides that can be represented by sequence patterns or linear motifs (Puntervoll, Linding et al. 2003; Neduva, Linding et al. 2005; Neduva and Russell 2005). In epitope-1, the N-terminal basic motif may be represented by a collection of sequence patterns covering 5-19 residues and the N-terminal hydrophobic motif by sequence patterns of approximately four residues. Epitopes 2 and 3 are both relatively smaller and simpler, and together can be described by a single sequence pattern.

2. Each linear epitope can accommodate large sequence diversity

Comparison of validated and potential PY-NLSs in Table 3-1 (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007) show that sequences within each of the three linear epitopes can be quite variable. The N-terminal basic/hydrophobic motif is the largest and most variable epitope. Mutagenesis of yeast PY-NLSs has provided more information on the diversity and also suggested some limits to the diversity of individual epitopes. In particular, positive charges within the N-terminal basic motifs are important but arginine and lysine residues are interchangeable and exact positions of basic groups

may not be important (Table 3-2, Table 3-3 and Table 3-4). Additional biochemical and structural studies will be needed to understand requirements of charge density, segment size and negatively selected amino acids in this epitope. The consensus for this basic region remains elusive. The 55% accuracy for bioinformatics-derived potential yeast bPY-NLSs binding to Kap104p may reflect high sequence variability and undiscovered physical characteristics of this region.

Epitope-2 is usually composed of a single residue. Examination of validated PY-NLSs (Table 3-1) shows that arginine is most prevalent in this position, although histidines are found in this position in hnRNP D, JKTBP, HuR and lysines in potential yeast NLSs of Naf1p, Sbp1p, Arp8p and Ste20p (Figure 3-11A). Mutagenesis has shown that arginine, lysine and histidines are interchangeable in this position. Thus, the appropriate sequence pattern here is R/K/H.

Human Kapβ2 substrate HuR (Table 3-1) has a PG dipeptide, yeast Nab2p and eight bioinformatics-derived potential yeast NLSs contain PL dipeptides at the C-terminal positions of their NLSs (epitope-3). In some cases, epitope-3 matters energetically more than in others. It is unclear why the dipeptide motif is energetically significant in some peptides and relatively silent in others. We speculate that a hydrophobic amino acid two residues N-terminal of the PY motif may be necessary (though probably not sufficient) and should be included in the sequence pattern for an energetically strong epitope-3. A hydrophobic residue at this position may preorganize the short peptide segment for binding, lowering both strain and entropic penalties. We also note that if epitope-3 is energetically very significant, the terminal site tends to be phenylalanine, histidine and

methionine. If the dipeptide motif is fairly silent energetically, many other amino acids are allowed in the terminal position.

3. Energetic cooperativity observed within linear epitopes, but not between them

Mutations within a linear epitope such as within the N-terminal basic region of Nab2p show large cooperativity of >60-fold (Table 3-2 and Table 3-4). Mutations within the N-terminal basic region of hnRNP M-NLS also show cooperativity in a similar regime, of ~40-fold (Cansizoglu, Lee et al. 2007). In contrast, seven examples of simultaneous mutations between different linear epitopes in Hrp1p, Nab2p (Table 3-2 and Table 3-3), hnRNPs A1 and M (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007) show only modest cooperativities of 1.0-3.8-fold. Cooperativity between linear epitopes in PY-NLSs is also very small compared to that typically observed between spatially distinct sites in conformational epitopes. For example, in the interaction of human growth hormone with human growth hormone receptor, mutations at distant sites in the interface showed large cooperativity of ~60-fold (Walsh, Sylvester et al. 2004). Thus, by comparison the linear epitopes in PY-NLSs are energetically quasi-independent. In an analogous system, a bipartite interaction in a linear sorting signal in a SNARE and COPII coat also exhibited energetic quasi-independence, showing only a 1.5-2-fold cooperative effect between the two distant sites (Mossessova, Bickford et al. 2003). In both PY-NLSs and vesicular sorting signals, minimal coupling between linear epitopes, and thus energetic modularity of those epitopes, may be attributed to flexible or structurally variable linkers that connect the epitopes.

4. Energetically variable linear epitopes can be mixed in a combinatorial fashion

Finally, the fourth biophysical property that governs PY-NLS affinity stems from the observation that binding energy is distributed very differently amongst the three linear epitopes in all seven thermodynamically characterized PY-NLSs (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007; Imasaki, Shimizu et al. 2007). In different PY-NLSs, a given linear epitope can vary significantly in its contribution to total binding energy. For example, the N-terminal basic motif in Hrp1p contributes much less to Kap104p binding than the equivalent epitope in Nab2p (compare Figure 3-14 A and B). Similarly, PY in hnRNP A1 contributes only weakly to Kapβ2 binding while PY motifs in hnRNP M and Hrp1p are sole binding hotspots in the NLSs (Figure 3-14 A, C and D). We had previously taken advantage of the energetic variability of PY-NLS epitopes by harnessing the avidity effect of NLS hotspot at epitope-1 in hnRNP A1 fused to NLS hoptspot at epitope-3 of hnRNP M, which resulted in a chimeric peptide inhibitor that bound Kapβ2 200-fold tighter than both substrates and RanGTP (Cansizoglu, Lee et al. 2007). Despite the wide energetic variability of individual linear epitopes, the total binding energies are very similar for various PY-NLS-containing substrates. Therefore, evolution has not combined epitopes randomly, but rather tuned them to a range for appreciable Kapβ2 binding and efficient Ran dissociation. The extremely tight-binding chimeric peptide inhibitor of Kapβ2 (Cansizoglu, Lee et al. 2007) is evidence of such evolutionary pressure. Although very high affinity can be achieved easily, nuclear import function is lost as RanGTP can no longer dissociate substrates.

Binding energy in the PY-NLS is distributed over a large sequence, with three different elements contributing differently in various substrates. It is this feature that makes the PY-NLS fundamentally different from the well-known monopartite classical-

NLS. A relatively small motif is recognized in a monopartite NLS and binding energy is concentrated in a stereotypical fashion across small sequences.

Modular and combinatorial design of PY-NLS may be highly evolvable

In PY-NLSs, the three distinct linear sequence elements are presented on peptides that exhibit intrinsic structural disorder and bind Kapβ2 with extended structurally diverse conformations. This modular and flexible display of multiple sequence motifs is relatively free of spatial constraints that usually relate multiple binding sites within a folded ligand. Furthermore, when binding energy is variably distributed among multiple epitopes in PY-NLSs, single mutations or mutations within single NLS epitopes are likely to have decreased chances of abolishing Karyopherin binding. Thus, the modular, flexible and energetically combinatorial architecture of PY-NLSs may allow significant evolvability to form new interactions while maintaining Kapβ2 recognition. Similar "multi-faceted" interactions, where different ligands make energetically significant interactions with different subsets of interface residues, were recently studied in a theoretical context (Humphris and Kortemme 2007) and also suggested to be more tolerant to mutations and are therefore quite evolvable.

Multiple functions have in fact been identified in several PY-NLSs. In Nab2p, the RGG region that overlaps with NLS epitope-1 is a putative RNA binding region (Anderson, Wilson et al. 1993). PY-NLSs in Nab2p, Hrp1p, EWS and FUS interact with and are methylated by arginine methyltransferases (Henry and Silver 1996; Shen, Henry et al. 1998; Belyanskaya, Gehrig et al. 2001; Green, Marfatia et al. 2002; Rappsilber, Friesen et al. 2003). Phosphorylation sites have also evolved within PY-NLSs to regulate nucleocytoplasmic localization. Serine phosphorylation in hnRNP A2-NLS and tyrosine

phosphorylation in SAM68-NLS (Lukong, Larocque et al. 2005) both alter subcellular localization of the proteins. A PY-NLS may also evolve additional NLSs within its sequence. This could generate redundancy in nuclear import pathways, and also provide a path to switch substrates from one Karyopherin to another, and ultimately from one cellular process to another. We have identified potential classical-NLS (Horton, Park et al. 2007) in the N-terminal basic motifs of eight human bPY-NLSs in Table 3-1. It is not clear what overlapping NLSs mean in the cellular context, but this question will need to be explored in the future.

Path to comprehensive PY-NLS identification in genomes

Identifying correct sequences that will account for most of the very diverse PY-NLS is an extremely challenging task. The core problem is that binding energy is distributed across three epitopes or motifs in many different ways. Thus, simply relaxing sequence constraints in a global search will also increase "noise" and result in many wrong answers.

We predict that if a PY motif (epitope-3) is energetically very significant, the sequence tolerance for this motif is small and sequence content of the other two epitopes will likely not matter. Thus, this subset of PY-NLSs should be easily identified upon identification of PY motifs that can provide large binding energies. Given the relatively small size of this motif, the task of finding strong PY motifs should be experimentally accessible. A similar situation should apply for an energetically strong N-terminal basic/hydrophobic motif (epitope-1). However, as the need for affinity from the PY motif decreases and as more binding energy is provided by the two other motifs, sequence tolerance relaxes. The problem of multiple motifs with varying sequence

tolerances seems very complex. But, the relatively small size of each motif, and energetic independence of the motifs allows the problem to be divided into manageable pieces. Our current inability to identify sequences of individual epitopes that are energetically strong may contribute to the 55% accuracy for bioinformatics-derived potential yeast bPY/L-NLSs binding to Kap104p. For example, individual epitopes in bioinformatics-derived sequences that did not bind Kap104p may be energetically weak and thus did not provide sufficient binding energy when combined.

First, the range of energies for PY-NLSs that are import-competent *in vivo* (and to what degree) will need to be determined. The range of suitable binding energies will likely vary depending on cellular concentrations of substrates, but should not be unbounded (Hodel, Harreman et al. 2006). For example, a designed peptide with a K_D of 100 pM binds Kapβ2 too tightly for *in vivo* nuclear import (Cansizoglu, Lee et al. 2007), thus providing a high affinity boundary for Kapβ2 import. Second, binding energies of putative PY-NLSs will need to be predicted. Unfortunately, the accuracy of calculating binding affinity for protein-small molecule interaction is still questionable and prediction of binding energies for protein-protein interactions are even further behind (Gilson and Zhou 2007). Our studies here suggest that we can get around this problem by handling each epitope independently and then combining them to assess for functional NLSs. We may use computational alanine-scanning mutagenesis (Kortemme and Baker 2002) to predict binding energy differences for each of the three PY-NLS linear epitopes and then empirically determine combinations that are functional. Such predictions could be tested against a future experimental thermodynamic database obtained from the initial predicted PY-NLSs (Lee, Cansizoglu et al. 2006), and the method iteratively refined. Binding

energy calculation remains problematic. We expect prevalent sequence- and physical characteristics-based bioinformatics methods are limited to successful prediction of potential NLSs with at least one energetically strong linear epitope, but will miss those composed of multiple weak or intermediate epitopes. A computational method that combines bioinformatics, structural modeling and prediction of binding energies may be a solution. Many more Kapβ2-NLS structures will be necessary to expand a structural database to facilitate modeling interactions of new sequences by homology modeling and/or physical energy function-based predictions of protein-protein interactions (Kortemme and Baker 2004; Baker 2006; Nayeem, Sitkoff et al. 2006).

Conclusion

PY-NLSs are very diverse in sequence and structure, and thus cannot be sufficiently described by their weak consensus motifs. Instead, PY-NLSs are described by a collection of weak physical rules that also include requirements for intrinsic structural disorder and overall positive charge (Lee, Cansizoglu et al. 2006). Here, we examined the energetic organization of PY-NLS through mutagenic and thermodynamic analyses of these signals in yeast. These studies have revealed physical properties that govern the binding affinity of this variable signal. The PY-NLS is a modular signal composed of three spatially distinct but structurally conserved linear epitopes that can be represented by a series of sequence patterns. Although each linear epitope can accommodate substantial sequence diversity, we have begun to define limits for each. More importantly, in addition to structural modularity, the three linear epitopes also

exhibit energetic modularity. Modular organization of the PY-NLS suggests that the daunting search for these very diverse sequences can be performed in parts. Finally, each linear epitope can contribute very differently to total binding energy in different PY-NLSs, explaining how signal diversity can be achieved through combinatorial mixing of energetically weak and strong motifs while maintaining affinity appropriate for nuclear import function. This collection of physical rules and properties describe how functional determinants of PY-NLSs are organized and lays a path to decode this diverse and evolvable signal for future genome-wide identification of Kapβ2 import substrates. More generally, many biological recognition processes involve linear recognition motifs with weak and obscure sequence motifs. Physical understanding of how diverse PY-NLS sequences can achieve common biological function may serve as a model for decoding many other weakly conserved and complex signals throughout biology.

CHAPTER 4

IDENTIFICATION AND VALIDATION OF PUTATIVE KAP104p NLSs REVEALS REDUNDANCY AMONG YEAST KAPβs

Abstract

Proteins containing a nuclear localization signal (NLS) are recognized and transported into the nucleus by Karyopherin β proteins. The yeast import protein Kap104p recognizes substrates containing a PY-NLS, which is a multipartite and diverse signal comprised of three binding epitopes. Refinement of the weak consensus sequence for the PY-NLS is hindered due to the lack of validated Kap104p substrates. Here we identify potential NLS sequences and validate them in *in vitro* binding assays. Additionally, we verify the sequences as functional targeting signals in cells. Most of the functional NLSs identified bind other karyopherins in addition to Kap104p revealing significant redundancy among the yeast karyopherins.

Introduction

Import Karyopherin β transport proteins recognize their substrates through nuclear localization signals (NLS) and transport these proteins into the nucleus where they are dissociated upon binding RanGTP (Chook and Blobel 2001). There are ten yeast import Kapβs that each recognize a distinct NLS, but there are currently only two well characterized NLSs: 1) the classical NLS (cNLS) recognized by the dimeric complex Kapβ1/Kapα and 2) the PY-NLS recognized by Kapβ2 (Kap104p in yeast) (Conti and Izaurralde 2001; Lee, Cansizoglu et al. 2006). There are hundreds of verified proteins containing cNLSs (Lange, Mills et al. 2007), but there are only two validated PY-NLS containing substrates of Kap104p, Hrp1p and Nab2p (Suel, Gu et al. 2008). The small number of substrates hinders analysis of the PY-NLS.

The PY-NLS is an extended signal comprised of three epitopes. There are three rules for NLS recognition by Kap104p (Lee, Cansizoglu et al. 2006). The PY-NLS must be in a disordered region with an overall basic charge. Additionally, the sequence must match the weak consensus of an N-terminal basic or hydrophobic (human only) region followed by a linker and a C-terminal $RX_{(2-5)}PY$ motif. Subsequent thermodynamic analyses revealed physical properties governing PY-NLS binding affinity for Kap104p (Suel, Gu et al. 2008). The three linear epitopes that make up the PY-NLS (epitope 1: hydrophobic/basic motif, epitope 2: arginine in $RX_{(2-5)}PY$ motif, epitope 3: PY motif) can be varying strengths in different NLSs. These epitopes can be combined in a manner that mixes strong and weak epitopes to result in a functional NLS. There can be significant

sequence diversity within the different epitopes. The basic enriched region can accommodate different mixtures of arginines and lysines as long as it has an overall positive charge. Epitope 2 can accommodate a lysine or histidine in addition to an arginine. The arginine however, appears to be the preferred residue at this position. The amino acids that can substitute for the tyrosine in epitope 3 vary depending on the energetic strength of the epitope. In Hrp1p epitope 3 is necessary for Kap104p binding and only a phenylalanine, histidine and methionine can substitute for the tyrosine. In Nab2p, which has a PL instead of a PY motif, the epitope can accommodate every amino acid without a significant decrease in binding energy.

In order to further evaluate the sequence requirements of the PY-NLS we previously performed a bioinformatics search to identify potential NLSs containing a PY or PL motif in epitope 3 (Suel, Gu et al. 2008). Eleven of the twenty putative NLSs tested bound Kap104p and were dissociated by RanGTP. Here we expand upon the previous search and include PM, PH, PI, PF motifs in our search. We first evaluated the ability of the potential PY-NLSs to bind Kap104p and be dissociated by RanGTP. To test if the potential NLS was a functional targeting sequence in cells, we tagged the NLS with GFP and visualized its localization in yeast cells. Six proteins were able to target GFP to the nucleus, but only one was dependent on Kap104p. The other five NLSs bound at least one other karyopherin.

Redundancy among yeast Kapβ has been seen with other proteins and is not surprising given that only two of the ten yeast import Kapβs are essential (Mosammaparast and Pemberton 2004). In the absence of Kap123p, Kap121p is capable of importing ribosomal proteins (Rout, Blobel et al. 1997) and most histones are

recognized and imported by multiple Kapβs (Mosammaparast, Jackson et al. 2001; Mosammaparast, Guo et al. 2002). In fact 45% of known yeast substrates (not including cNLS-bearing cargos) have a secondary import pathway (See Redundancy section on page 10). However, it is somewhat unusual to see functional overlap in only a span of 50 residues. This study identifies sequences recognized by multiple karyopherins. Additionally, we obtained the sequences of NLSs that bind Kap104p and those that are not recognized by Kap104p for use in future computational studies of PY-NLSs.

Materials and Methods

Plasmids

Yeast substrate genes were obtained by PCR from a *S. cerevisiae* genomic DNA library (Novagen) and subcloned into BamHI and NotI sites in the pGexTev and/or pMalTev vector. NLSs were confirmed by nucleotide sequencing. A complete list of plasmids can be found in Table 7-6 in the appendix.

The GFP-GST-NLS construct was cloned into a modified pRS415 (*CEN6, ARS, LEU2*, and AP^R) shuttle vector containing a 5' ADH1 promoter (Sikorski RS and Hieter 1989). GFP and a ggsgg linker were inserted into SpeI and PstI sites (Figure 4-1). GST in combination with engineered BamHI and NotI sites was cloned into PstI and XhoI sites. NLSs were subcloned from pGexTev vectors and inserted into BamHI and NotI sites. A complete list of plasmids can be found in Table 7-4.

Figure 4-1: Cloning of GFP-GST-NLS.

Cell culture and microscopy

Please refer to page 59.

Bioinformatics search for new Kap104p substrates

Please refer to page 59.

Protein expression and purification

Please refer to page 60.

Binding assays

Please refer to page 67.

Results

Identification and in vitro validation of putative Kap104p NLSs

The two substrates of Kap104p, Hrp1p and Nab2p, have a PY and PL motif respectively at epitope 3. Additionally, mutagenesis experiments showed that phenylalanine, methionine, and histidine residues could substitute for the tyrosine in Hrp1p (Figure 3-15). We had previously identified putative bPY and bPL NLSs using a bioinformatics search (see page 71). In order to search for additional substrates with more variability in epitope 3 we used the sequence pattern K/R-X_{0-6}-K/R-X_{0-6}-K/R-X_{0-6}-K/R-X_{2-5}-R/K/H-X_{1-5}-PY/L/F/M/H/I to search for potential yeast bPY-NLSs within *S. cerevisiae* proteins in the UniProtKB/Swiss-Prot protein database (Bairoch, Boeckmann et al. 2004). The resulting lists were filtered for structural disorder (Linding, Jensen et al. 2003) and overall basic character. The search resulted in 24 putative bPY-NLSs, 71 putative bPL-NLSs, 18 putative bPF-NLSs, 11 putative bPM NLSs, nine putative bPH NLSs and 40 putative bPI NLSs.

Putative NLSs from each list were cloned from a yeast genomic library into a pGexTev vector and used in binding assays. Kap104p was added to immobilized putative NLS in the presence and absence of RanGTP. The putative NLSs were divided into three groups: strong binding, weak binding, or no binding and are listed in Table 4-1.

Three out of the five bPY-NLSs bound Kap104p strongly (Figure 4-2). Tfg2p bound Kap104p with an affinity of greater than 10 nM as measured by ITC. Rml2p bound with an affinity of 236±47 nM and Snp1p with an affinity of 1.4±0.9 μM (Table 4-1). Five out of 12 putative bPL-NLSs bound Kap104p strongly, three bound weakly,

and three showed no discernable binding (Figure 3-11Figure 4-3) Prb1p had a K_D=1.2±0.4 μM in ITC experiments with Kap104p while Enp1p and Naf1p showed no detectable heat.

Of the eight putative bPF-NLSs tested two bound Kap104p strongly, one bound weakly and the five did not bind Kap104p (Figure 4-4). Aft1p bound Kap104p with an affinity of 1.1±0.2 μM as measured by ITC. Only one out of the five putative bPM-NLSs bound Kap104p (Figure 4-5). Ten putative bPI-NLSs were examined. Three bound Kap104p strongly, two weakly, and five did not bind Kap104p (Figure 4-6). There was some heat detected during an ITC experiment using Gbp2p and Kap104p, but not enough to quantify. Out of the 39 putative NLSs tested 14 bound Kap104p strongly, seven bound weakly and 18 did not bind. Only four of the putative NLSs bound Kap104p with an affinity greater than 1.4 μM.

Table 4-1: Putative Kap104p NLSs.

Protein	Local-ization[a]		Putative NLS		ITC
Strong Binding					
HRP1[b]	N	498	DRDRDYNHRSGGNHRRNGRGGRGGYNRRNNGYHPYNR	534	32±16 nM
NAB2[b]	N	205	RFTQRGGGAVGKNRRGGRGGNRGGRNNNSTRFNPLAKA	242	37±20 nM
TFG2	no data	170	ELKKKQQQQKRRNNRKKFNHRVMTDRDGRDRYIPYVKT	207	>10 nM
AFT1	N/C	183	SNATVTNGPQTSPDQTSSIKPKKKRCVSRFNNCPFRVR	220	1.1±0.2 μM
CDC25	C	779	SSAASGSVFTPFNRPSHNRTFSRARVSKRKKKYPLTVD	816	
ENP1	N/C	1	MARASSTKARKQRHDPLLKD	20	no heat
GBP2	N	70	DDRNWPPRRGGRGRGGSRSFRGGRGGGRGRTLGPIVER	107	>10 μM
HAA1	C	94	GTDEVCKYHAQKRHLRKSPSSSQKKGRSISRSQPMFER	131	>10 μM
HRK1	C	664	SDNKSSQQHRGPHHKKIIHGPYRLLRLLPHASRPIMSR	701	
MMR1	no data	54	SLLYPTSLSKLSELSRGGRSKQRRGSDTMRSVSPIRFQ	88	no heat
NAF1	N	274	KKRDNRKLANDSDNVKVKRARQPKANSLPKLVPPLGMS	311	no heat
POL12	N	89	TKKPVIKKSLNSSPLFGLSIPKTPTLKKRKLHGPFSLS	126	
PRB1	C	130	NKVEEKKMKGKKVKGKKHHEKTLEKGRHHNRLAPLVST	167	1.2±0.4 μM
RML2	N/C	79	LIKRRRKLSKEVTQMKRLKPVSPGLRWYRSPIYPYLYK	116	236±47

Protein	Local-ization[a]		Putative NLS		ITC
					nM
SBP1	C	143	FRGGYRGGFRGRGNFRGRGGARGGFNGQKREKIPLDQM	180	
SNP1	N/C	1	MNYNLSKYPDDVSRLFKPRPPLSYKRPTDYPYAKR	35	1.4±0.9 µM
Weak Binding					
APC2	N/C	410	NQDTNITKRDKNKKSPFLWNLKVKGKRELNKDLPIRHA	447	
ARP8	N	119	SRDKRAPPAVQTSKRYKKYPKLDPAKAPPGKKVPLHLL	156	
DIG1	N	71	ADHEDSETATAKKRKAQPLKNPKKSLKRGRVPAPLNLS	108	
ICL1	no data	23	AAEIEKWWSDSRWSKTKRNYSARDIAVRRGTFPPIEYP	60	
RPS1A	N/C	1	AVGKNKRLSKGKKGQKKRVVDPFTRK	26	
SGV1	N	429	GPKKDDASFLPPSKNVLAKPPPSKIRELHQNPRPYHVN	466	
STE20	C	231	NPKHKQHKPKVKPSKPEAKSKPVSVKKSFPSKNPLKNS	268	
No Binding					
BMS1	cell peri.	24	QGHNAKAFAVAAPGKMARTMQRSSDVNERKLHVPMVDR	61	
CDC5	N	1	MSLGPLKAINDKQLNTRSKLVHTPIKGN	28	
CTF8	no data	67	LIRFGSLQIDGERATLFVGKKQRLLGKVTKLDVPMGIM	104	
CWC25	N	136	LLKDDPMSKFKVTKQQRRTPDSTKKRAMSQRGKPLSKP	172	
GCN5	N	272	AALRRKIRTISKSHIVRPGLEQFKDLNNIKPIDPMTIP	309	
HAA1	C	1	MVLINGIKYACERCIRGHRVTTCNHTDQPLMMI	33	
HAA1	C	567	SIHSVPQSINSPRMPKTGSRQDKNIHTKKEERNPLNNI	604	
LGE1	N	219	KGYPVHVPENRSNSDGSSSSVVKKKRILDMKDSPFIYL	256	
LRS4	N	114	KHNKERTPSTGRDEQQRNSKAAHTSKPTIHLLSPIVNR	151	
MRPS8	no data	69	STRRLWVGLKYRDNKPVLSSCKLISKPNSRIHLPMEDM	106	
NDT80	no data	35	LTQLNEDGTTSNYFDKRKLKIAPRSTLQFKVGPPFELV	72	
RGD1	C	328	MNHGQKNKSPPKFAVDPSRNSIPKRMISTHNESPFLSS	365	
RPO41	punc.	465	VKEEINHCRSLLSEKLSDKKSLNKVDTNRLGYGPYLTL	502	
SGD1	N	66	EINSSRLKSAPTSEKRSANAGVKNVGKQANGKNPISSD	103	
SGV1	N	550	PNDSRYHNPRYVNKPETNFNRQPQKYSRQESNAPINKN	587	
SPO13	no data	201	YPNVNVYDSPLFKKTRLPHQTKSLDKEKNYQYLPIYPV	238	
TRM8	N	28	LKHVKINESSLVQEGQKIDLPKKRYYRQRAHSNPFSDH	65	
YHI9	no data	45	ANWTNLSETTFLFKPSDKKYDYKLRIFTPRSELPFAGH	82	

[a]UCSF GFP localization database (Huh, Falvo et al. 2003)
[b]Validated substrate of Kap104p.

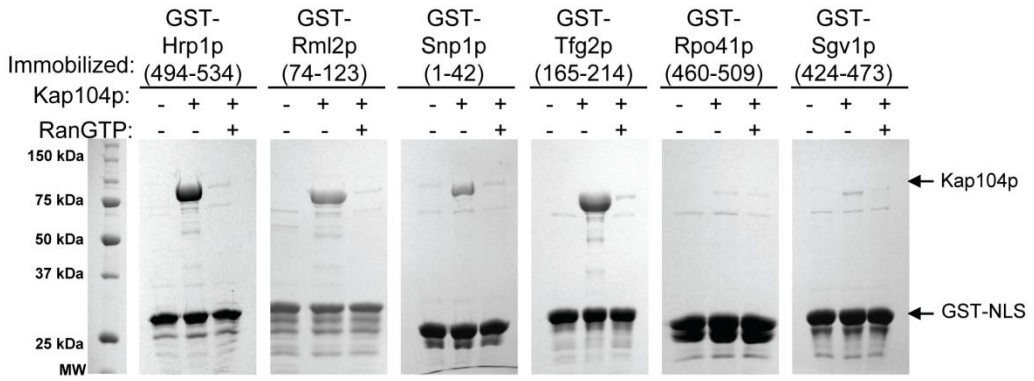


Figure 4-2: Putative bPY NLSs. Kap104p was added to immobilized GST-NLS in the presence and absence of RanGTP.

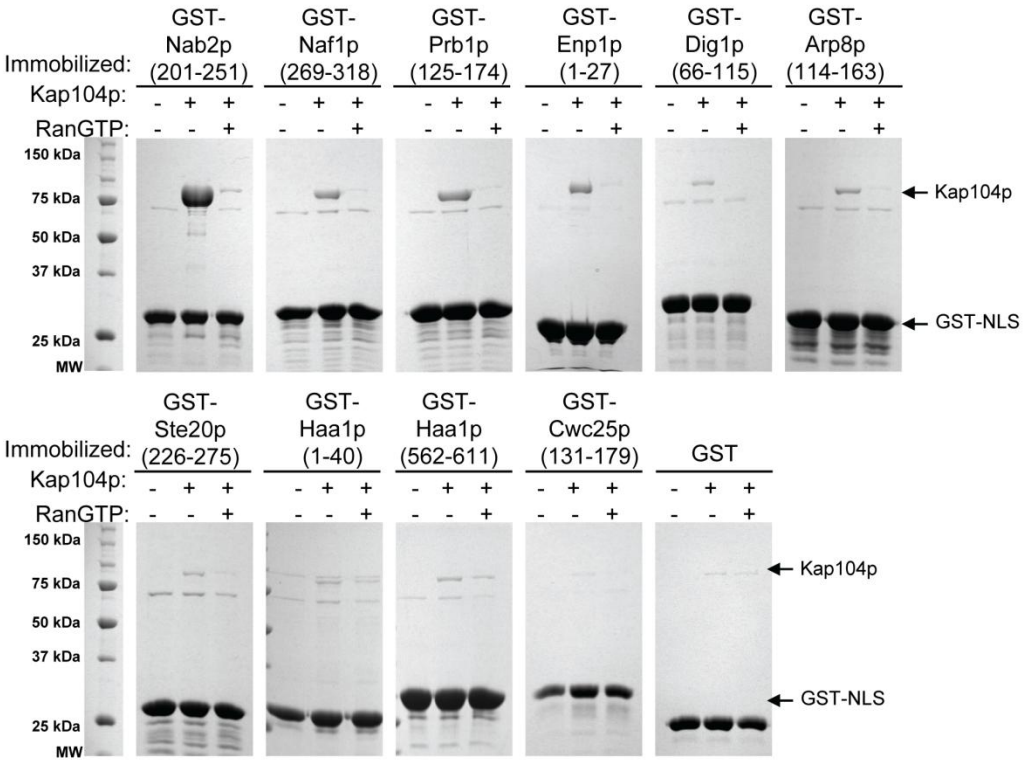


Figure 4-3: Putative bPL NLSs. Kap104p was added to immobilized GST-NLS in the presence and absence of RanGTP.

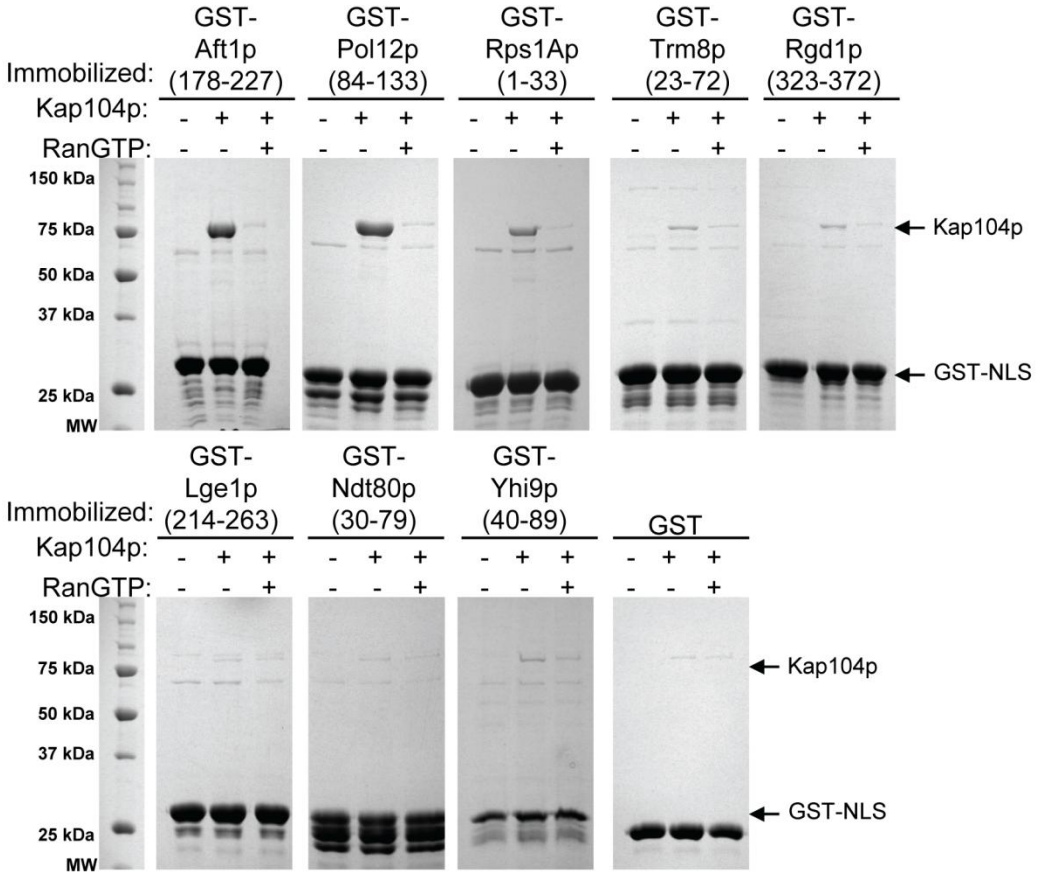


Figure 4-4: Putative bPF NLSs. Kap104p was added to immobilized GST-NLS in the presence and absence of RanGTP.

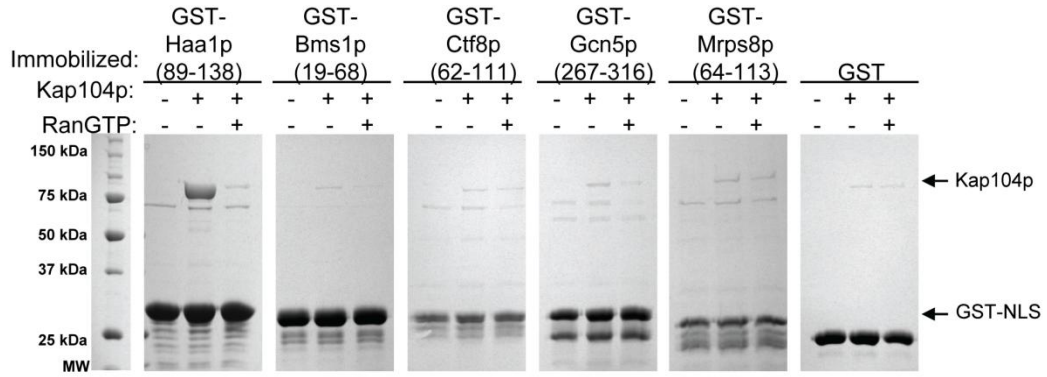


Figure 4-5: Putative bPM NLSs. Kap104p was added to immobilized GST-NLS in the presence and absence of RanGTP.

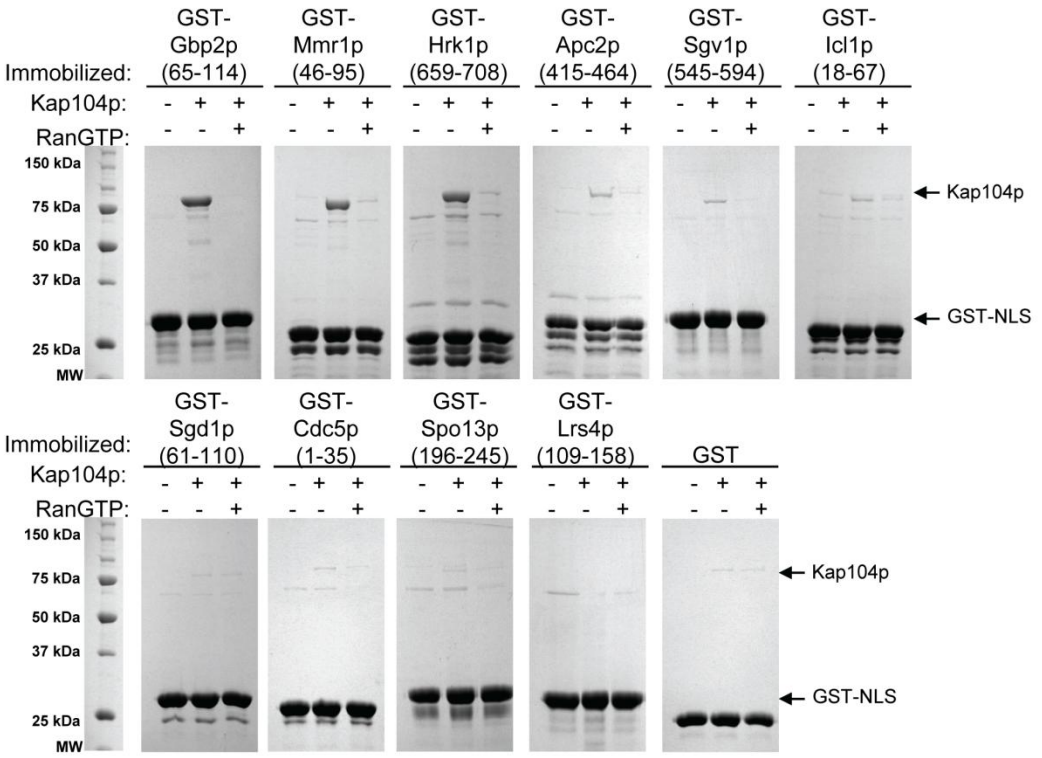


Figure 4-6: Putative bPI NLSs. Kap104p was added to immobilized GST-NLS in the presence and absence of RanGTP.

Validation of putative NLSs as functional targeting signals in cells

In order to test if the putative NLSs that bound Kap104p are functional targeting signals in cells we fused GFP-GST proteins to the N-terminuses of the NLSs and visualized their localization in wild type cells. Eight of the NLSs tested were cytoplasmic at steady state (Figure 4-7) while six were nuclear (Figure 4-8). Thus, the NLSs of Tfg2p, Aft1p, Cdc25p, Naf1p, Arp8p, and Dig1p are functional nuclear targeting sequences.

Dependence on Kap104p for nuclear import

We have shown that the six putative NLSs were functional targeting sequences, however we wanted to test if the nuclear import of these NLSs was dependent on Kap104p. The six GFP-GST-NLS fusion proteins and full length Tfg2p-GFP were transformed into *kap104-16* temperature sensitive cells. Protein localization was visualized at permissive (room temperature) and non-permissive (37°C) temperatures. The NLSs of Aft1p, Cdc25p, Naf1p, Arp8p and Dig1p remained nuclear at the non-permissive temperature (data not shown). Both GFP-GST-Tfg2p and the control Nab2p-GFP mislocalized to the cytoplasm at the non-permissive temperature (Figure 4-9). Thus, Tfg2p NLS is the only newly identified NLS whose primary import Kapβ is Kap104p and the other five NLSs are imported by another pathway.

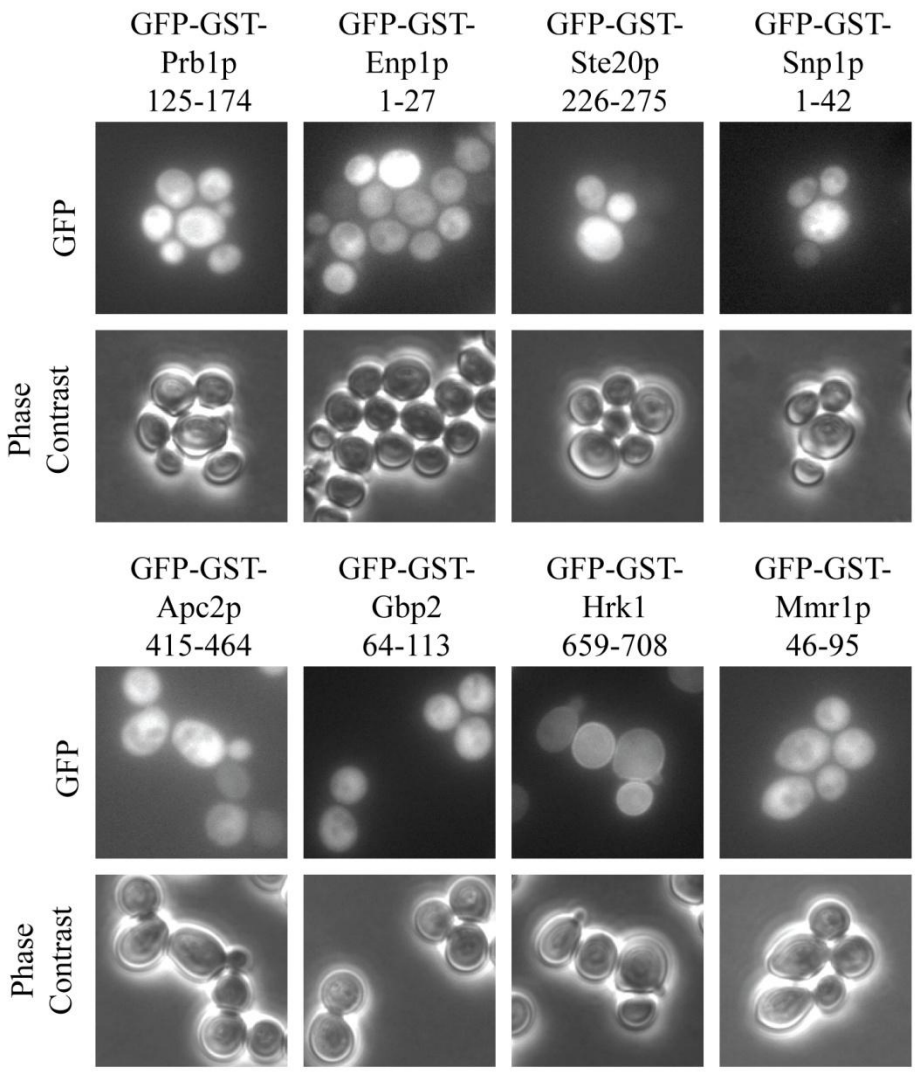


Figure 4-7: Cytoplasmic localization of putative GFP-GST-NLSs. The localization of GFP-GST-NLS fusion proteins were examined in wild type cells by fluorescence microscopy and phase contrast microscopy.

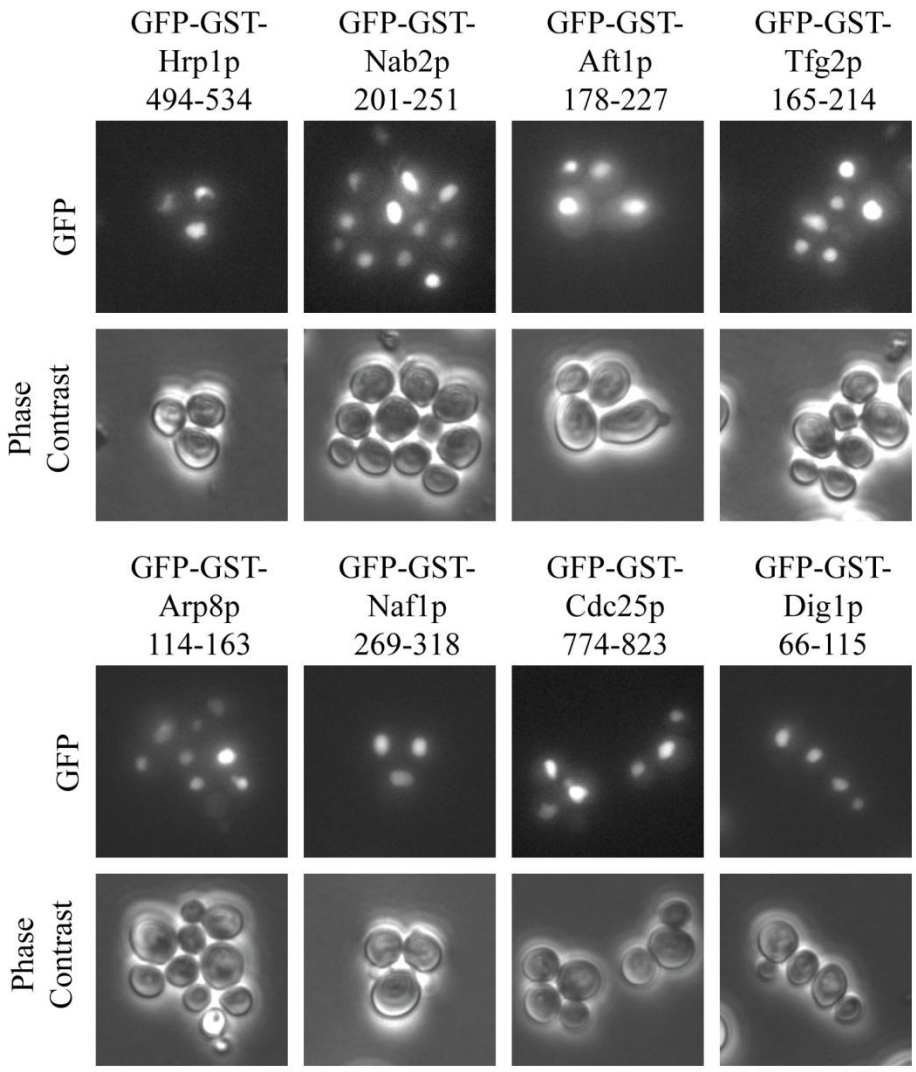


Figure 4-8: Nuclear localization of putative GFP-GST-NLSs. The localization of GFP-GST-NLS fusion proteins were examined in wild type cells by fluorescence microscopy and phase contrast microscopy.

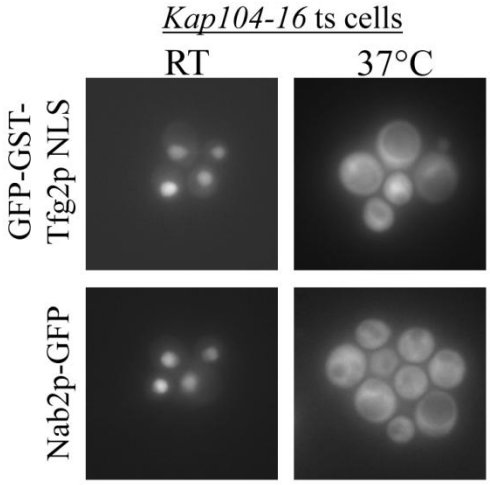


Figure 4-9: GFP-GST-Tfg2p NLS was mislocalized in *kap104-16* temperature sensitive cells at the non-permissive temperature, but not at the permissive temperature. Nab2p-GFP is shown as a positive control.

Recognition of NLS by other karyopherins

The NLSs in Aft1p, Cdc25p, Naf1p, Arp8p and Dig1p are being imported into the nucleus by another mechanism, possibly by recognition by another import karyopherin. The best characterized and most abundant NLS is the cNLS recognized by Kap60p (Kapα in humans). A putative monopartite cNLS is located in all five NLSs (Table 4-2). In order to test if the cNLS is recognized by Kap60p, protein was added to immobilized NLSs. Kap121p was also tested for binding because it is an essential Kapβ and has been shown to recognize rg-NLSs (Leslie, Zhang et al. 2004). All five of the NLSs bound Kap60p stochiometrically (Figure 4-10). The NLSs of Arp8p, Naf1p, and Dig1p bound Kap60p stronger than they bound Kap104p. Cdc25p and Aft1p bound all three Kapβs strongly. Tfg2p bound sub-stochiometrically to Kap121p and did not bind Kap60p.

Surprisingly, Nab2p bound Kap121p almost as well as it bound Kap104p. There is significant redundancy among the yeast Karyopherin βs even within a small 50-residue NLS.

Table 4-2: Putative cNLSs in PY-NLSs.

Gene	Putative NLS
HRP1	DRDRNDRDRDYNHRSGGNHRRNGRGGRGGYNRRNNGYHPYNR
NAB2	VDNSQRFTQRGGGAVGKNRRGGRGGNRGGRNNNSTRFNPLAKALGMAGES
AFT1	EYASPSNATVTNGPQTSPDQTSSIKP**KKKR**CVSRFNNCPFRVRATYSLKR
TFG2	QEREEELKKKQQQQKRRNNRKKFNHRVMTDRDGRDRYIPYVKTIPKKTAI
ARP8	NELGSSRDKRAPPAVQTS**KRYK**KYPKLDPAKAPPGKKVPLHLLEKRRLGR
NAF1	QQRQRKKRDNRKLANDSDNVKV**KRAR**QPKANSLPKLVPPLGMSSNAPMQH
CDC25	IDLKASSAASGSVFTPFNRPSHNRTFSRARVS**KRKK**KYPLTVDTLNTMKK
DIG1	SGDKEADHEDSETATA**KKRK**AQPLKNPKKSL**KRGR**VPAPLNLSDSNTNTH

cNLS are underlined and in bold. Epitope 1 of the PY-NLS is in blue and epitopes 2 and 3 are in red.

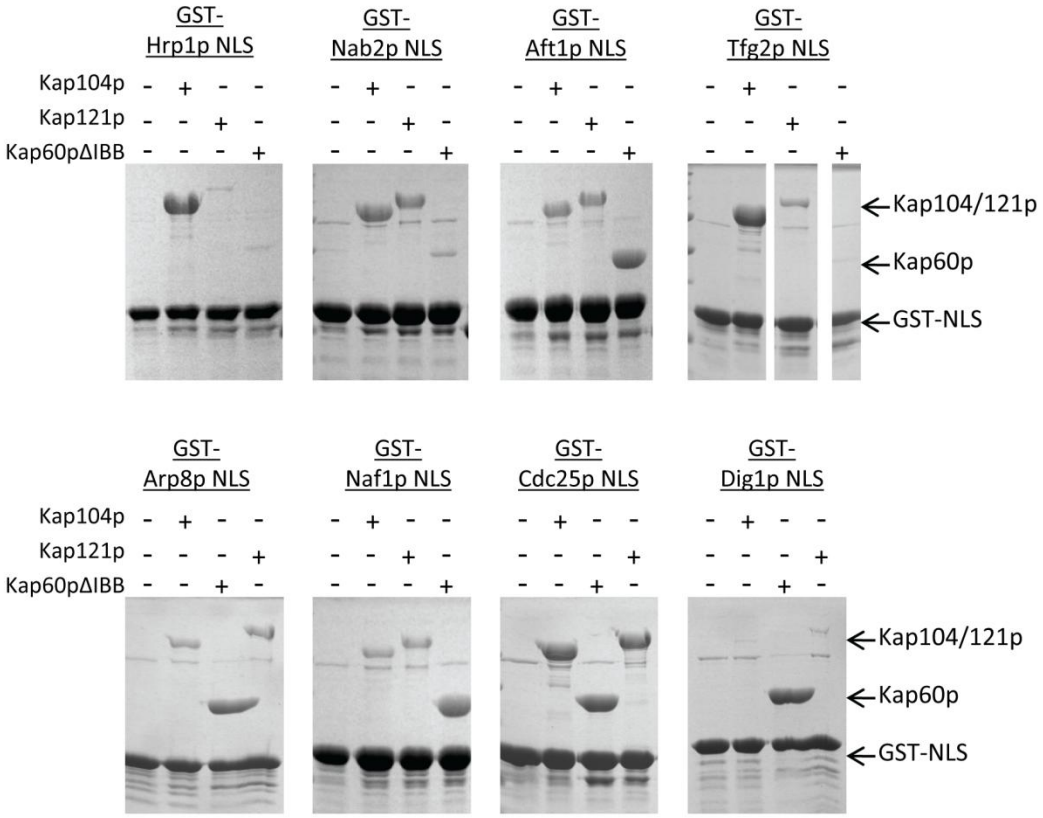


Figure 4-10: Binding assay. Kap104p, Kap121p, or Kap60pΔIBB was added to immobilized putative GST-NLSs.

Discussion

The PY-NLS is an extended signal with a weak consensus sequence. A hindrance to refinement of the consensus sequence is the lack of Kap104p sequences. In order to identify additional Kap104p substrates we used the weak consensus sequence plus additional information from previous mutagenesis studies in a bioinformatics search (Suel, Gu et al. 2008). We identified 173 putative NLSs and tested 39 for Kap104p binding. Of those 39 putative sequences 14 bound Kap104p strongly, seven bound weakly to Kap104p and 18 were not recognized by Kap104p. Six of the predicted NLSs were able to target GFP-GST to the nucleus and one, Tfg2p, was primarily imported by Kap104p.

All of the putative NLSs that bind Kap104p contribute to our knowledge of sequence requirements for Kap104p recognition. The NLSs do not need to be functional targeting sequences in the cell for this purpose. If we examine just the ten sequences that bind Kap104p and have a detectable ITC signal we witness the diverse sequence requirements for Kap104p recognition (Table 4-3). We have previously examined the sequence requirements of the three epitopes through mutagenic experiments of Hrp1p and Nab2p (Suel, Gu et al. 2008). Arginines are the only basic residues in epitope 1 of Hrp1p and Nab2p, however the arginines can be substituted with lysines without a decrease in binding affinity. Therefore we concluded that the only requirement for epitope 1 is a basic charge. Epitope 1 of the identified sequences have a mixture of arginines and lysines, supporting this conclusion. Additionally, an arginine is the residue in epitope 2 in Hrp1p and Nab2p, however thermodynamic studies found that a lysine or

Table 4-3: Sequences recognized by Kap104p

Protein	Local-ization[a]		Putative NLS		ITC
Strong binding on beads/ heat detected by ITC					
HRP1[b]	N	498	DRDRDYNHRSGGNHRRNGRGGRGGYNRRNNGYHPYNR	534	32±16 nM
NAB2[b]	N	205	RFTQRGGGAVGKNRRGGRGGNRGGRNNNSTRFNPLAKA	242	37±20 nM
TFG2	no data	170	ELKKKQQQQKRRNNRKKFNHRVMTDRDGRDRYIPYVKT	207	>10 nM
hnRNP M[c]		27	VPSQNGAPGEGERPAQNEKRKEKNIKRGGNRFEPYANP	67	12±5 nM
RML2	N/C	79	LIKRRRKLSKEVTQMKRLKPVSPGLRWYRSPIYPYLYK	116	236±47 nM
AFT1	N/C	183	SNATVTNGPQTSPDQTSSIKPKKKRCVSRFNNCPFRVR	220	1.1±0.2 µM
PRB1	C	130	NKVEEKKMKGKKVKGKKHHEKTLEKGRHHNRLAPLVST	167	1.2±0.4 µM
SNP1	N/C	1	MNYNLSKYPDDVSRLFKPRPPLSYKRPTDYPYAKR	35	1.4±0.9 µM
GBP2	N	70	DDRNWPPRRGGRGRGGSRSFRGGRGGGRGRTLGPIVER	107	>10 µM
HAA1	C	94	GTDEVCKYHAQKRHLRKSPSSSQKKGRSISRSQPMFER	131	>10 µM
Strong binding on beads/ NO heat detected by ITC (or not tested)					
ENP1	N/C	1	MARASSTKARKQRHDPLLKD	20	no heat
MMR1	no data	54	SLLYPTSLSKLSELSRGGRSKQRRGSDTMRSVSPIRFQ	88	no heat
NAF1	N	274	KKRDNRKLANDSDNVKVKRARQPKANSLPKLVPPLGMS	311	no heat
POL12	N	89	TKKPVIKKSLNSSPLFGLSIPKTPTLKKRKLHGPFSLS	126	not tested
CDC25	C	779	SSAASGSVFTPFNRPSHNRTFSRARVSKRKKKYPLTVD	816	not tested
HRK1	C	664	SDNKSSQQHRGPHHKKIIHGPYRLLRLLPHASRPIMSR	701	not tested
SBP1	C	143	FRGGYRGGFRGRGNFRGRGGARGGFNGQKREKIPLDQM	180	not tested
FUS[c]			RGGFRGGRGGDRGGFGPGKMDSRGEHRQDRRERPY	526	not tested

[a]UCSF GFP protein localization database (Huh, Falvo et al. 2003)
[b]Validated substrate of Kap104p
[c]human proteins

histidine could substitute in this position without a significant decrease in binding affinity (Suel, Gu et al. 2008). All ten sequences that bind Kap104p with some discernable affinity, as measured by ITC, have an arginine in this position (Table 4-3). The other eight sequences that bind Kap104p in binding assays have a mixture of R,K and H at this position. However, arginine appears to be the most prevalent and preferred residue in epitope 2.

Epitope 3 is the PY motif. If epitope 3 is energetically strong it can only accommodate a F, M, or H at this position, however if this epitope is energetically weak in an NLS it can accommodate all residues therefore we increased our search to include PF, PM, PH, and PI motifs in epitope 3. Only nine putative substrates containing a PH motif were identified and none tested bound Kap104p. There is a representative of every other motif in the list of sequences that bound to Kap104p. The majority of sequences containing PL and PY motifs tested bound Kap104p, while the majority of sequences containing PF, PM, and PI did not bind Kap104p. While Kap104p is capable of binding proteins containing a diverse range of sequences at epitope 3, PL and PY may be the most preferred and common motif in Kap104p substrates.

While the identification of new NLSs and examination of their sequences confirms previously identified sequence requirements, they provide little additional insight into restraints on the consensus sequence. However, the sequences will be advantageous in future computational studies of the PY-NLS. The PY-NLS is a modular signal with binding energy distributed across three epitopes. This feature complicates the identification of new NLSs. The negative results obtained could be the result of combining weak epitopes. This data may be helpful in training a program to identify strong versus weak binding epitopes and assist in the prediction of new substrates.

The function of an NLS is to target a protein to the nucleus. Eight of the putative NLSs identified are unable to target an unrelated protein, GFP, to the nucleus. The full length protein of two of the eight cytoplasmic putative NLSs are localized to the cytoplasm, three have both nuclear and cytoplasmic localization and only one is nuclear according to the UCSF GFP fusion localization database (Huh, Falvo et al. 2003). In

contrast, of the six nuclear putative NLSs, one has cytoplasmic localization of its full length protein, one is localized to both the nucleus and cytoplasm and four are nuclear (Huh, Falvo et al. 2003). The localization of the full length protein is sometimes used as an indicator of whether or not a putative NLS will be a functional targeting signal in the cell. However, this can be misleading and nuclear localization should not be used as a restriction on a search for NLSs. There are other factors affecting the localization of a protein besides its NLS. The NLS of Cdc25p is able to target GFP to the nucleus but the full length protein is cytoplasmic. Cdc25p is a plasma membrane bound Ras-GEF (Garreau, Geymonat et al. 1996). The sequestration of Cdc25p in the cytoplasm counteracts its nuclear localization and results in cytoplasmic steady state localization. Additionally, the fact that the full length protein is nuclear does not guarantee that the putative NLS identified is responsible for targeting the protein to the nucleus. For example, full length Gbp2p is nuclear, but the putative PY-NLS is not a functional targeting signal. It has been shown that there is a functional NLS in the N-terminus of Gbp2p recognized by Kap111p/Mtr10p (Windgassen and Krebber 2003). Other cellular factors, including the presence of additional NLSs and cytoplasmic retention, determine the localization of a protein.

Of the six NLSs that are functional targeting sequences only one is affected by the loss of Kap104p mediated transport. The other five NLSs bind at least one other karyopherin. Aft1p and Cdc25p bind Kap104p, Kap60p and Kap121p almost equally well. Ueta et al. have shown that nuclear import of Aft1p is primarily mediated by Kap121p/Pse1p (Ueta, Fukunaka et al. 2003). They identified two NLSs in Aft1p: 332-365 and 198-225 which includes the putative PY-NLS recognized by Kap104p and the

putative cNLS recognized by Kap60p. Other NLSs have been identified that are recognized by multiple karyopherins. Nop1p is primarily imported by Kap121p, but Kap104p can rescue import in the absence of Kap121p (Leslie, Zhang et al. 2004). Fries et al. identified an NLS in Asr1p that is recognized strongly by Kap114p and weakly by Kap95p, Kap121p and Kap123p (Fries, Betz et al. 2007). It is only mislocalized in the absence of two or more Kapβs. What is the cause of this redundancy? In the case of the yeast PY-NLS, epitope 1 is a basic motif which in Aft1p, Cdc25p, Naf1p, Arp8p and Dig1p is a putative cNLS. As mentioned in Chapter 3, the PY-NLS is an evolvable signal. A specific NLS may have either evolved to recognize other karyopherins or evolved from other signals.

Cellular conditions determine if an NLS will be effectively recognized and transported. We show that the NLS of Nab2p also binds Kap121p, but it is well known that it is mislocalized to the cytoplasm in *kap104-16* temperature sensitive cells (Aitchison, Blobel et al. 1996). Nab2p is an essential gene while Kap104p is not. In the absence of Kap104p, Kap121p may be capable of transporting the mRNA protein Nab2p into the nucleus, albeit at a slower rate. The ability of NLSs to bind multiple karyopherins may allow the cell to withstand stochastic fluctuations in the concentrations of Kapβ proteins.

CHAPTER 5

KAP104p IMPORTS THE PY-NLS-CONTAINING TRANSCRIPTION FACTOR TFG2p INTO THE NUCLEUS

Abstract

A previous bioinformatics study identified a putative PY-NLS in the yeast transcription factor Tfg2p (Suel, Gu et al. 2008). In this study, we validate Tfg2p as a Kap104p substrate and examine the energetic organization of its PY-NLS. The Tfg2p PY-NLS can target a heterologous protein into the cell nucleus through interactions with Kap104p. Surprisingly, full length Tfg2p is still localized to the nucleus of Kap104p temperature sensitive cells and similarly, Tfg2p with a mutated PY-NLS is nuclear in wild type cells. Other Kapβs such as Kap108p and Kap120p also bind Tfg2p and may import it into the nucleus. More importantly, we demonstrate that Tfg2p is retained in the nucleus through DNA binding. Mutations of DNA binding residues relieve nuclear retention and unmask the role of Kap104p in Tfg2p nuclear import. More generally, steady state localization of a nuclear protein is dictated by its nuclear import and export activities as well as its interactions in the nucleus and the cytoplasm.

Introduction

The majority of nucleocytoplasmic transport is mediated by Karyopherinβ proteins (Kapβ, Importins/Exportins). There are 19 Kapβs in humans and 14 in yeast. Ten of the yeast Kapβs import substrates from the cytoplasm to the nucleus (Fried and Kutay 2003; Mosammaparast and Pemberton 2004). Import Kapβs recognize and bind substrates via nuclear localization signals (NLSs) for transport through the nuclear pore complex. Once inside the nucleus, import Kapβs bind the small GTPase RanGTP and release their substrates (Gorlich and Kutay 1999; Chook and Blobel 2001; Conti and Izaurralde 2001; Weis 2003). Only a few substrates have been identified for each Kapβ and many yeast proteins are imported by multiple karyopherins.

Each import Kapβ recognizes a different set of substrates with distinct NLSs. The best characterized NLS is the short, basic classical NLS (cNLS), which is recognized by the Kapα/Kapβ1 heterodimer (yeast Kap60p/Kap95p) (Conti and Izaurralde 2001). Monopartite cNLSs consist of a single cluster of basic residues with a consensus sequence of K(K/R)X(K/R) whereas bipartite NLSs have two clusters of basic residues separated by 10-12 amino acids (Kalderon, Richardson et al. 1984; Conti, Uy et al. 1998; Conti and Kuriyan 2000; Fontes, Teh et al. 2000; Catimel, Teh et al. 2001; Hodel, Corbett et al. 2001; Fontes, Teh et al. 2003; Lange, Mills et al. 2007). The cNLS is a relatively small well-defined NLS with concentrated binding energy.

In contrast to the small monopartite cNLS, the PY-NLS recognized by Kapβ2 (Kap104p in yeast) is a larger linear signal that is quite diverse in sequence (Lee, Cansizoglu et al. 2006; Suel, Gu et al. 2008). Structural and biochemical studies of

Kapβ2 bound to the NLS of hnRNP A1 revealed that PY-NLSs should be (1) structural disordered in the native substrate, (2) have overall basic charge and (3) contain an N-terminal hydrophobic or basic motif and a C-terminal R/K/HX$_{2-5}$PY motif (Lee, Cansizoglu et al. 2006). These rules are general among eukaryotes as the *S. cerevisiae* homolog Kap104p recognizes PY-NLSs in its two substrates, Hrp1p and Nab2p (Aitchison, Blobel et al. 1996; Siomi, Fromont et al. 1998; Truant, Fridell et al. 1998; Lee and Aitchison 1999; Lange, Mills et al. 2008; Suel, Gu et al. 2008). However, unlike human Kapβ2, Kap104p recognizes only the basic but not the hydrophobic PY-NLS subclass (Suel, Gu et al. 2008).

Recent thermodynamic analyses of Kap104p-PY-NLS interactions revealed biophysical properties that govern binding affinity (Suel, Gu et al. 2008). The PY-NLS is a modular signal that contains at least three energetically significant binding epitopes. Epitope 1 is the N-terminal basic/hydrophobic motif, epitope 2 is the arginine, lysine or histidine residue of the R/K/HX$_{2-5}$PY motif and epitope 3 is the PY motif. Hrp1p and Nab2p have an aromatic residue two residues N-terminal of the PY motif that also contributes to binding energy and is potentially an extension of epitope 3 (Suel, Gu et al. 2008). This residue has not been included in the consensus sequence due to the small numbers of validated substrates. Each linear epitope can also accommodate large sequence diversity, can contribute differently to overall binding energy in different NLSs and are energetically quasi-independent.

A bioinformatics search identified several putative Kap104p substrates that contain PY-NLSs (Suel, Gu et al. 2008). Among these potential Kap104p substrates is Tfg2p, which is one of three subunits of the yeast general transcription factor TFIIF.

TFIIF is involved in the initiation and elongation of RNA polymerase II (RNAPII) mediated gene transcription and mutations in TFIIF can cause upstream shifts in start site utilization (Ghazy, Brodie et al. 2004; Freire-Picos, Krishnamurthy et al. 2005). Additionally, TFIIF stimulates early phosphodiester bond formation and stabilizes a short DNA-RNA hybrid in the active center of RNAPII (Khaperskyy, Ammerman et al. 2008). In cells, Tfg2p exists in a large assembly with RNA polymerase II proteins and two other TFIIF subunits, Tfg1p and Tfg3p (Chung, Craighead et al. 2003). Tfg1p and Tfg2p are essential for cell viability and are homologous to human RAP74 and RAP30 respectively whereas Tfg3p is not essential and has no known human homolog (Henry, Campbell et al. 1994). Tfg2p contains an N-terminal Tfg1p interaction domain (residues 1-230) and a C-terminal winged helix DNA binding domain (residues 280-400) (Figure 5-2) (Henry, Campbell et al. 1994; Groft, Uljon et al. 1998; Gaiser, Tan et al. 2000; Chen, Warfield et al. 2007).

The scarcity of Kap104p substrates hinders additional evaluation and refinement of the rules and physical properties that govern PY-NLS recognition. Here we validate Tfg2p as a substrate of Kap104p and examine the physical properties of its PY-NLS. We show that the PY-NLS is a functional targeting signal in cells and its import is primarily mediated by Kap104p. However, full length Tfg2p accumulates in the nucleus even in the absence of Kap104p mediated nuclear import. When Kap104p import is compromised, other Kapβs such as Kap108p and Kap120p can import Tfg2p into the nucleus where it is retained through its DNA binding domain. Mutations of DNA binding residues alleviate nuclear retention and reveal Kap104p as the primary nuclear import pathway for Tfg2p.

Materials and Methods

Plasmids

Full length *Tfg2* was cloned by PCR from a *S. cerevisiae* genomic library (Novagen) and subcloned into the SpeI and SmaI sites of a modified pRS415 (*CEN6, ARS, LEU2, and* AP^R) shuttle vector containing an ADH1 promoter and a C-terminal GFP gene (Sikorski and Hieter 1989). The PY-NLS and other *Tfg2* fragments were cloned into the pRS415 plasmid containing N-terminal GFP and GST genes using inserted BamHI and NotI restriction sites. There is a GGSGG linker between the genes. The PY-NLS was also subcloned into the BamHI and NotI sites of the pGEX-Tev vector. All point mutations were cloned using the QuikChange method (Stratagene) and confirmed by nucleotide sequencing.

The pRS314 vector used for the viability assays was a gift from A. Ponticelli (Ghazy, Brodie et al. 2004). The vector contains *Tfg2* under the control of an endogenous promoter with a Myc tag at its C-terminus. Point mutations were cloned using the QuikChange method. The C-terminus deletion mutant was cloned by inserting a stop codon after residue 343. The NLS deletion clone (Δ176-206) was made using QuikChange. A ggsgg linker was inserted in place of the NLS. All clones were confirmed by nucleotide sequencing.

All Kapβ genes (except *Kap104* which was a gift from J. Aitchison) were obtained by PCR from the *S. cerevisiae* genomic library and subcloned into pGEX-Tev vectors. Please see Table 7-1 and Table 7-5 for a complete list of plasmids.

Cell culture and microscopy

BY4741 (*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) cells (Brachmann, Davies et al. 1998) harboring pRS415 plasmids were grown at 30°C overnight in SC-leu media. Cells were then transferred to a 1.5% low melting agarose pad mad with SC-leu media in a coverslip bottom dish (Willco from Ted Pella). Cells were visualized on an Olympus IX-81 inverted microscope at room temperature with a 60X phase objective with fluorescence capability (numerical aperture=1.35). Images were acquired with a Hamamtsu ORCA-ER camera using Image-Pro Plus software (Media Cybernetics) and were also analyzed using Image Pro. All the cell images in individual figure panels were taken on the same day and GFP signals in the panels were displayed using the same fluorescence scale. To obtain a nucleus:cytoplasm ratio, the mean fluorescence intensity was measured in the nucleus and cytoplasm using a 36-pixel box for at least 50 cells of each mutant.

Kap104-16 temperature sensitive cells (a gift from J. Aitchison) were grown at room temperature in SC-leu-trp media overnight (Aitchison, Blobel et al. 1996). Cells at the permissive temperature were analyzed as above. The remaining cells were shifted to 37°C for 1 hour and then transferred to a pre-warmed agarose pad. Cells were observed in a 37°C temperature chamber and analyzed as above.

Protein expression and purification

GST-Kapβ proteins were expressed in *E. coli* BL21 (DE3) cells and lysed using an EmulsiFlex-C5 homogenizer (Avestin). After centrifugation, the supernatant was applied to glutathione sepharose (GE Healthcare) and washed extensively. The protein

was eluted with 20 mM glutathione and the GST tag was cleaved using Tev protease. The protein was then applied to an anion exchange column (GE Healthcare). The final product was put over Ni Sepharose (GE Healthcare) to remove excess Tev. Please refer to page 60 for detailed information regarding buffers.

GST-Tfg2 PY-NLSs were expressed in *E. coli* BL21 (DE3) cells and lysed by sonication. The protein was immobilized on glutathione sepharose and washed three times. The protein was left on the beads for binding assays.

Binding assays

Binding assays were preformed as on page 67.

Viability assays

Wild type and mutant pRS314-Tfg2 (TRP) plasmids were transformed into a *Tfg2* plasmid shuffle strain (a gift from A. Ponticelli) (Ghazy, Brodie et al. 2004). The FP290 strain (*MATα ura3-52 trp1Δ63 tfg2Δ1[+p316/g2(URA3)])* has a deletion of the endogenous *Tfg2* gene but harbors a plasmid containing a wild type *Tfg2* gene (URA3). Transformants were plated on SC-trp-ura medium. Single colonies were grown in liquid –ura-trp media overnight at 30ºC and then diluted so that all mutants were at the same cell density. Cells were then spotted on either YPD media as a control or on plates containing 5-fluoroorotic acid (5-FOA). 5-FOA is toxic to cells with a functional URA3 gene and will select for cells that have lost the wild type *Tfg2* containing plasmid. The presence of cell growth indicates that the mutant *Tfg2* gene is able to support cell growth in the absence of wild type *Tfg2*.

Results

Tfg2p contains a PY-NLS recognized by Kap104p

A bioinformatics search identified a putative PY-NLS in Tfg2p (Suel, Gu et al. 2008). In addition, a previous high-throughput mass spectrometric protein complex identification study identified Tfg2p along with Nab2p, Hrp1p, and Ran as a binding partner of Kap104p (Ho, Gruhler et al. 2002). Residues 165-214 of Tfg2p match the PY-NLS consensus sequence and are located in a long loop in the Tfg1p interaction domain. Sequence alignment with RAP30, the human homolog of Tfg2p, suggests that this loop is an insertion in the yeast protein that is absent in RAP30 (Figure 5-1A and B). Epitope 1 of the PY-NLS would correspond to two stretches of basic residues, ^{179}KRR^{181} and ^{184}RKK^{186}, while epitope 2 is R200 (Figure 5-2B). Epitope 3 is located at ^{203}PY^{204}, with a predicted additional binding hotspot at Y201A. Epitope 1 of the Tfg2p NLS differs from the epitope 1 in Hrp1p and Nab2p, but epitope 3 in all three proteins (Y/FxPY) are almost identical.

We had shown recently that the predicted PY-NLS and full length Tfg2p bound Kap104p and were dissociated by RanGTP *in vitro* (Figure 3-10) (Suel, Gu et al. 2008). Tfg2p bound Kap104p with higher affinity than either Nab2p or Hrp1p. Tfg2p bound Kap104p with a dissociation constant or K_D of less than 10 nM as measured by isothermal titration calorimetry, at least four times stronger than Nab2p or Hrp1p (Figure 5-2C). The Tfg2p PY-NLS is able to dissociate Kap104p from immobilized GST-Nab2p NLS to the similar degree as RanGTP (Figure 5-3). While these data show that Tfg2p

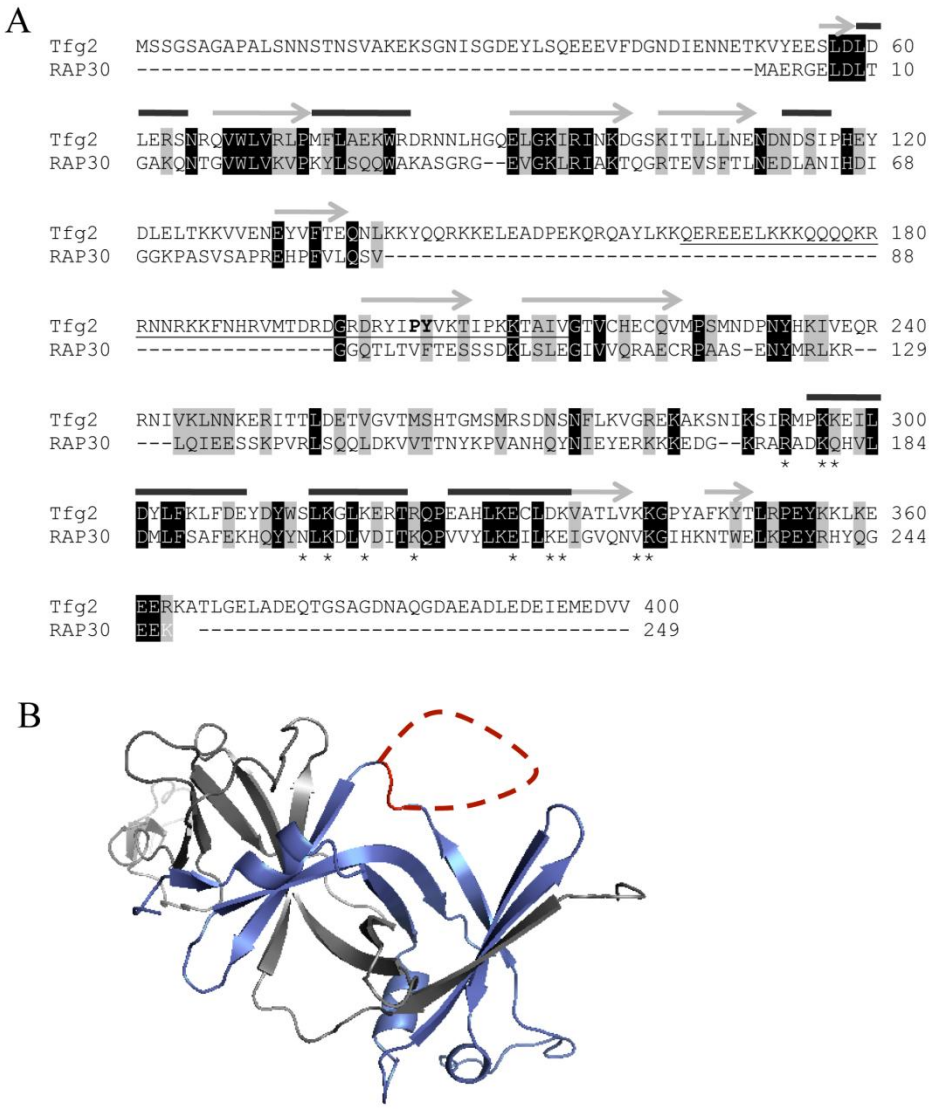


Figure 5-1: (A) Alignment of *S. cerevisiae* Tfg2p with its human homolog RAP30. The PY-NLS is underlined and the PY motif is in bold. Identical residues are highlighted in black and similar residues in gray. Beta strands are represented by gray arrows and alpha helices by black bars. Secondary structure labels are from the RAP30 crystal (1f3u) and NMR structures (1bby). Asterisks denote residues mutated in the DNA binding domain. (B) Crystal structure of RAP74 (gray)/ RAP30 (blue) homodimerization domains. The PY-NLS insertion (red dashed line) is between RAP30 residues 88 and 89 (red solid line). Figure was made from pdb file 1f3u in PyMOL (Gaiser, Tan et al. 2000; DeLano 2002).

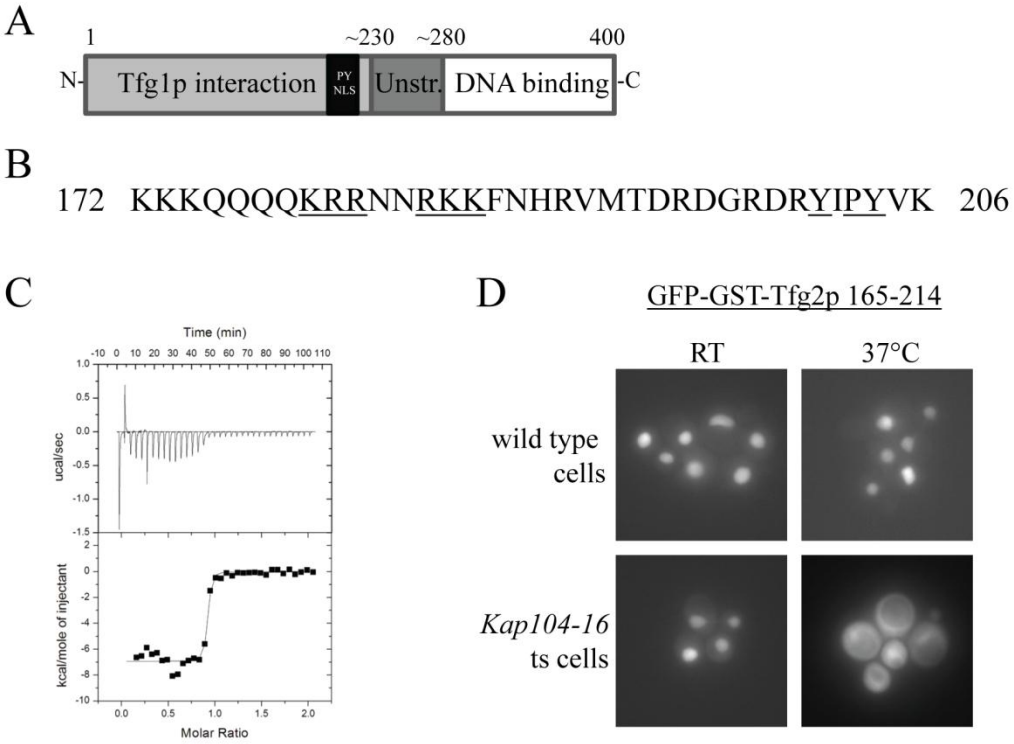


Figure 5-2: Tfg2p has a PY-NLS that is recognized by Kap104p. A) Domain organization of Tfg2p. B) Sequence of the PY-NLS. Residues mutated in C and D are underlined. C) ITC profile of MBP-Tfg2p PY-NLS injected into Kap104p. D) Wild type and *kap104-16* temperature sensitive cells expressing GFP-GST-Tfg2p PY-NLS were analyzed by fluorescence microscopy at permissive (room temperature) and non-permissive (37°C) temperatures.

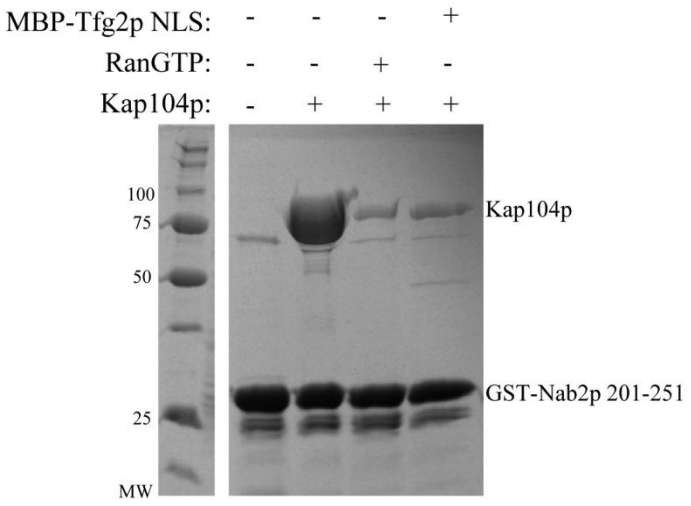


Figure 5-3: Tfg2p PY-NLS dissociates Kap104p from Nab2p PY-NLS. Kap104p was added to immobilized GST-Nab2p PY-NLS in the absence and presence of RanGTP or MBP-Tfg2p PY-NLS.

PY-NLS and Kap104p interact *in vitro*, it is not known if the PY-NLS-like sequence in Tfg2p is a functional NLS that is capable of targeting Tfg2p to the cell nucleus.

To determine if the Tfg2p PY-NLS is a functional Kap104p-dependent NLS in cells, a GFP-GST reporter was fused to the N-terminus of Tfg2p PY-NLS. GST was added to make the reporter larger to prevent free diffusion into the nucleus. The fusion protein was transformed into wild type and *kap104-16* temperature sensitive cells (Aitchison, Blobel et al. 1996). At the permissive temperature, GFP-GST-Tfg2p PY-NLS is localized to the nucleus in both cell types (Figure 5-2D). However, at the non-permissive temperature GFP-GST-Tfg2p PY-NLS mislocalized to the cytoplasm in the *kap104-16* temperature sensitive cells. Thus, the Tfg2p PY-NLS is a functional targeting signal that is transported by Kap104p into the nucleus.

Tfg2p PY-NLS has a strong epitope 3

The three linear epitopes of PY-NLSs can contribute differently to total binding energy in different substrates (Suel, Gu et al. 2008). We performed site-directed mutagenesis of the Tfg2p PY-NLS to examine energetic contributions of the three linear epitopes to Kap104p binding. Recombinant Kap104p bound immobilized wild type GST-Tfg2 PY-NLS stoichiometrically (Figure 5-4A). Alanine mutations within either stretches of basic residues, ^{179}KRR^{181} or ^{184}RKK^{186}, had no visible effect on Kap104p binding. Mutagenesis of Y201 to alanine resulted in a modest decrease in binding, while the ^{203}PY^{204}/AA mutant showed a large decrease in Kap104p binding (Figure 5-4A).

GFP-GST-Tfg2 PY-NLS fusion proteins with the above point mutations were visualized in wild type cells to examine the effects of the mutations on cellular localization. Wild type Tfg2p PY-NLS is localized to the nucleus with a nucleus to cytoplasm ratio (N:C) of 4:1 (Figure 5-4B). K179A, R180A, R181A and R184A, K185A, K186A mutants are predominantly nuclear with N:C ratios of 3.7:1 and 3.2:1 respectively (SEM < 0.14 for each mutant). These data are in agreement with the *in vitro* data that showed no decrease in Kap104p binding. These mutants also confirm that these basic residues do not form a bipartite cNLS within the PY-NLS. Mutations in epitope 3 that decreased Kap104p binding also caused protein mislocalization in cells. GFP-GST-Tfg2p Y201A showed significant cytoplasmic localization, with an N:C ratio of 1.9:1 and the fusion protein harboring ^{203}PY^{204}/AA mutations is predominantly cytoplasmic (Figure 5-4B). Therefore, the Tfg2p NLS has an energetically strong epitope 3 with a PY motif that is necessary for steady state nuclear localization.

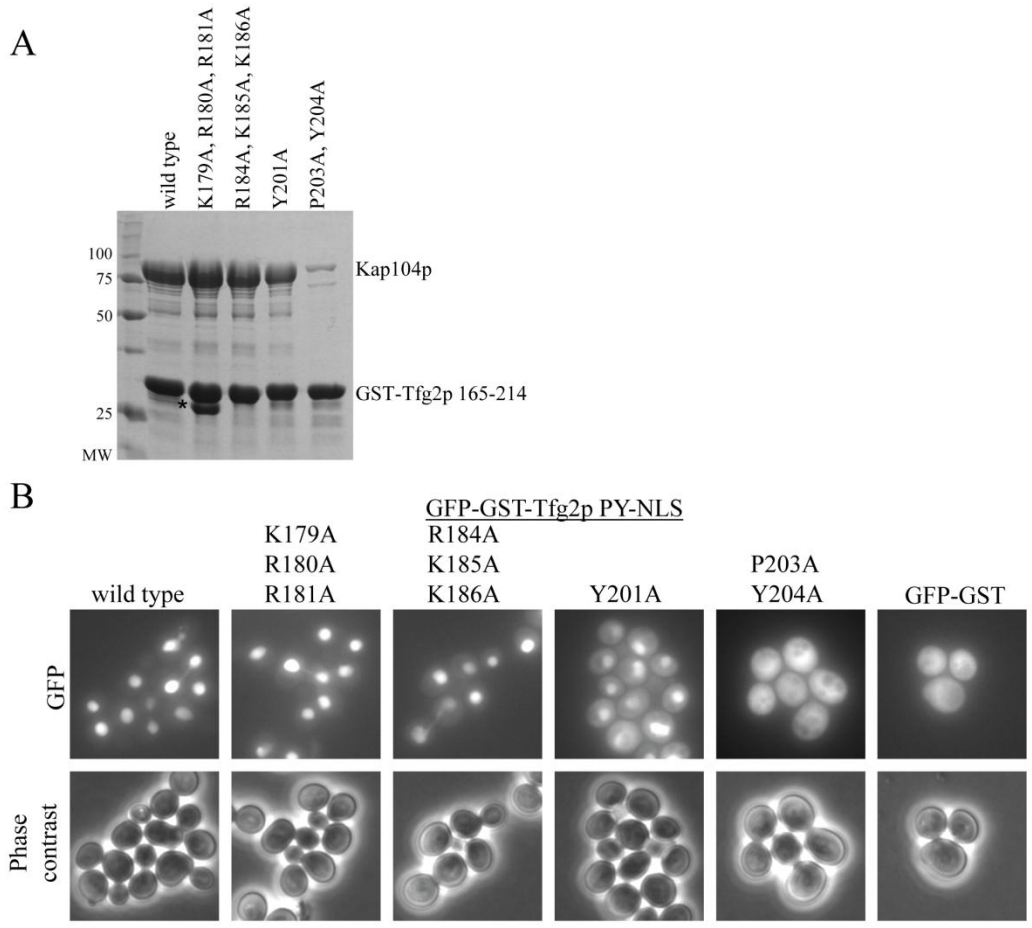


Figure 5-4: Tfg2p has a strong epitope 3. A) Binding assays of recombinant Kap104p with immobilized wild type or mutant Tfg2p PY-NLS. Proteins were visualized by Coomassie staining. B) Cellular localization of GFP-GST-Tfg2 PY-NLS mutants in wild type yeast cells analyzed by fluorescence microscopy and phase contrast. GFP-GST is included as a control.

Only Kap104p binding is affected by mutations in the PY motif

The ^{203}PY^{204}/AA mutations mislocalized Tfg2 PY-NLS to the cytoplasm. We wanted to see if this effect was solely due to a decrease in Kap104p binding or if the PY motif may be utilized by another import-Kapβ. We cloned and purified all 10 import-Kapβs and the bidirectional transporter Msn5 for *in vitro* binding assays (Yoshida and Blobel 2001). Kap104p is the only Kapβ that bound stoichiometrically to wild type PY-NLS (Figure 5-5A). Kap108p, Kap121p, Kap114p, Kap119p and Kap95p all bound sub-stoichiometrically suggesting very weak affinity for the NLS. Mutagenesis of the PY motif only caused a decrease in Kap104p binding and did not affect other Kapβ binding (Figure 5-5B). Other karyopherins weakly recognize the PY-NLS, but the PY motif is only necessary for recognition by Kap104p. Therefore, mislocalization of Tfg2p PY-NLS caused by the ^{203}PY^{204}/AA mutations can be attributed solely to a decrease in Kap104p binding and nuclear import.

Nuclear import of full length Tfg2 appears unaffected by loss of Kap104p recognition

While the PY-NLS of Tfg2p is localized to the nucleus, the localization of full length Tfg2p is unknown. Full length Tfg2p was tagged at the C-terminus with GFP and its localization was examined in wild type yeast cells. Wild type Tfg2p is nuclear at steady state (Figure 5-6A). Surprisingly, the ^{203}PY^{204}/AA mutations did not affect the localization of Tfg2p-GFP (Figure 5-6A), suggesting that nuclear import of Tfg2p may not be entirely dependent on Kap104p.

To examine if nuclear accumulation of full length Tfg2p is dependent on Kap104p, Tfg2p-GFP was transformed into *kap104-16* temperature sensitive cells. At

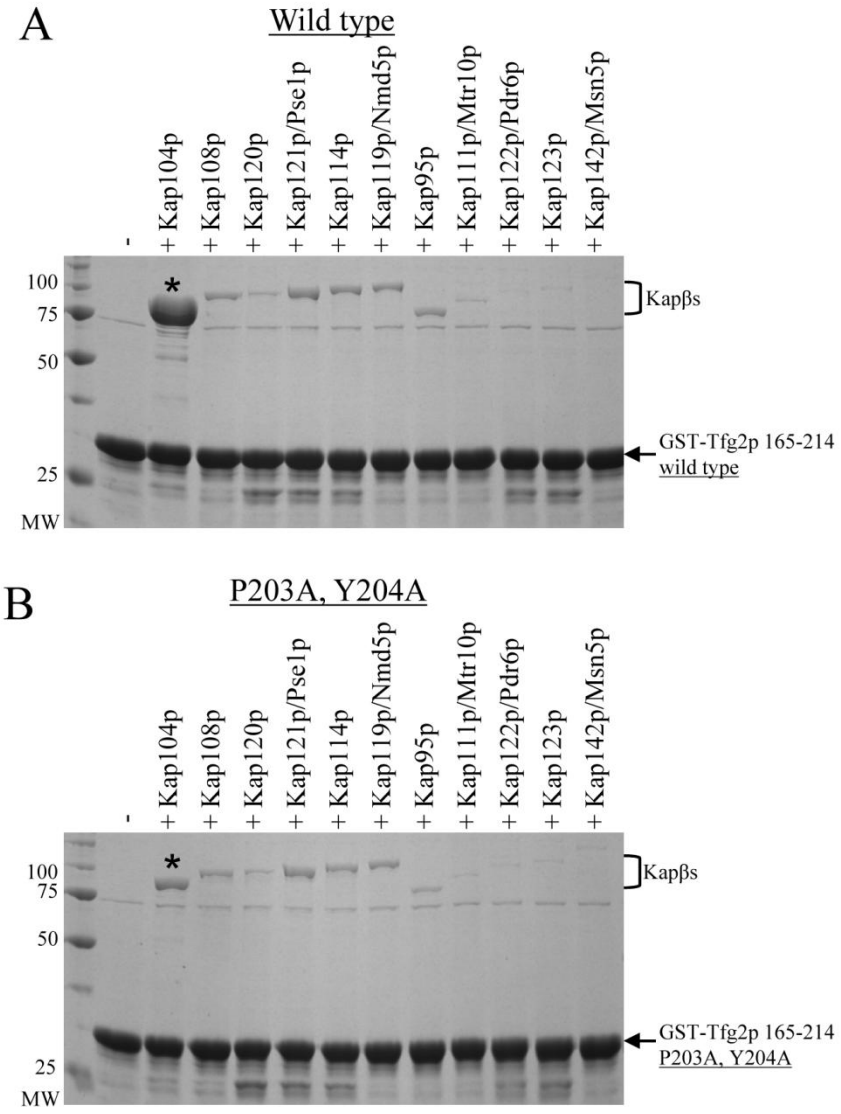


Figure 5-5: Only Kap104p binding is decreased by mutations in the PY motif. A) Binding assay of recombinant Kapβs with immobilized wild type Tfg2p PY-NLS. B) Same assay as in A, except with ^{203}PY^{204}/AA mutation in the PY-NLS. Proteins were visualized by Coomassie staining.

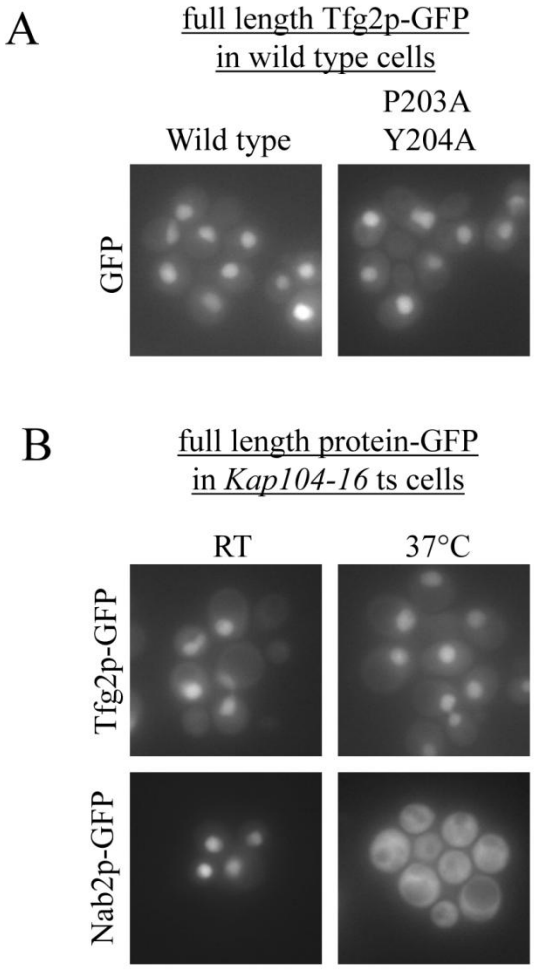


Figure 5-6: Nuclear import of full length Tfg2p is not dependent on Kap104p. A) Wild type yeast cells expressing either wild type full length Tfg2p-GFP fusion protein or ^{203}PY^{204}/AA mutant fusion protein were analyzed by fluorescence microscopy. B) *Kap104-16* temperature sensitive cells expressing either Tfg2p-GFP or Nab2p-GFP were analyzed by fluorescence microscopy at permissive and non-permissive temperatures.

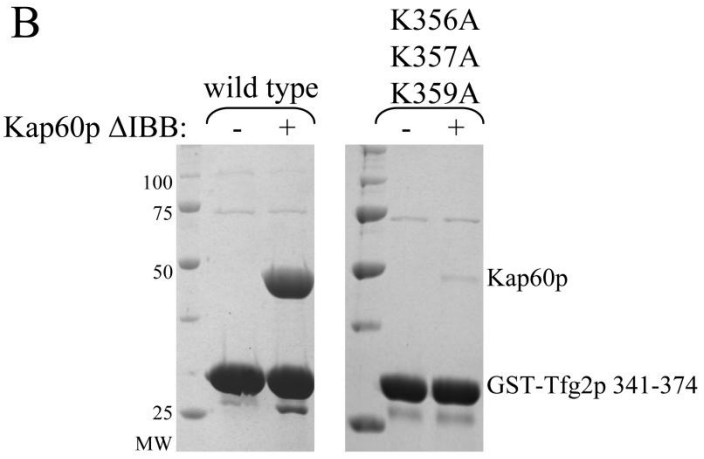
both permissive and non-permissive temperatures, Tfg2p-GFP is nuclear (Figure 5-6B). Nab2p-GFP was used as a positive control to ensure that the *kap104-16* temperature sensitive cells had not reverted. In the same experiment, Nab2p-GFP is mislocalized to the cytoplasm at non-permissive temperature. Nuclear import and accumulation of full length Tfg2p is therefore not solely dependent on Kap104p.

Other karyopherins bind Tfg2p

We examined if other Kapβs can also bind and import Tfg2p into the nucleus. The best characterized import pathway is the Kap95p/Kap60p (Kapβ1/Kapα) pathway utilizing cNLS recognition by Kap60p. Examination of the Tfg2p protein sequence revealed a putative monopartite cNLS (^{356}KKLK^{359}) at the C-terminus of Tfg2p (Figure 5-7A). Immobilized GST-Tfg2p (341-374) bound Kap60p stoichiometrically and alanine mutagenesis of the three lysine residues abolished Kap60p binding (Figure 5-7B). We made these mutations in the background of the ^{203}PY^{204}/AA mutations in full length Tfg2p-GFP and examined its localization in wild type cells. The mutant Tfg2p (^{356}KK^{357}/AA, K359A, ^{203}PY^{204}/AA)-GFP fusion protein accumulated in the nucleus (Figure 5-7C). Additionally, GFP-GST-Tfg2p 341-374 was cytoplasmic (data not shown) suggesting that these residues (^{356}KKLK^{359}) are not a functional cNLS in cells.

All ten import Karyopherin βs and the bidirectional transporter Msn5p were used in binding assays with immobilized full length Tfg2p. Kap108p bound Tfg2p stoichiometrically, Kap120p bound almost stoichiometrically and Kap121p and Kap114p bound Tfg2p sub-stoichiometrically (Figure 5-8A). None of the karyopherins bound

A 341 KGPYAFKYTLRPEYKKLKEEERKATLGELADEQ 374



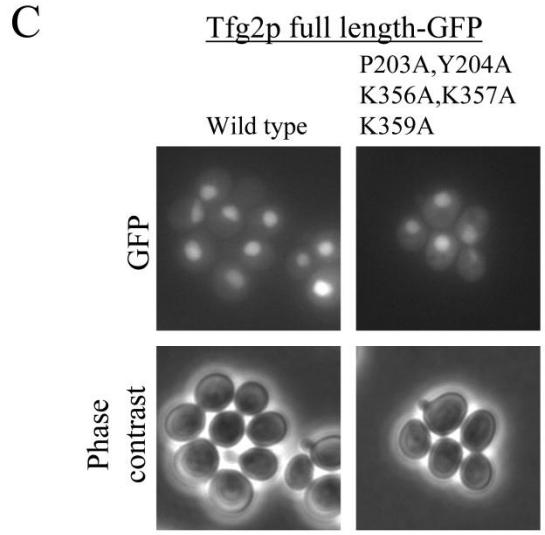


Figure 5-7: Tfg2p has a non-functional cNLS. A) Sequence of a predicted cNLS in Tfg2p. The cNLS is underlined. B) Kap60p binds the putative NLS and mutagenesis of the cNLS abolished binding. Kap60p was added to immobilized wild type or mutant GST-Tfg2p cNLS. C) The cellular localization of Tfg2p-GFP is not affected by mutations in the PY motif and in the cNLS. Localization in wild type yeast cells were analyzed by fluorescence microscopy and phase contrast.

GST (data not shown). Therefore, Kap108p and Kap120p may be alternative nuclear import factors for Tfg2p.

Kap108p and Kap120p did not bind the Tfg2p PY-NLS (Figure 5-5A) and therefore must recognize different NLSs in Tfg2p. We cloned GFP-GST fusions of several Tfg2p fragments and visualized their localization in wild type yeast cells (Figure 5-8B). The fragments were designed to encompass entire secondary structure elements of Tfg2p (based on alignment with RAP30 (Figure 5-1A)). All of the fragments were cytoplasmic at steady state, suggesting that although Kap108p and Kap120p bind Tfg2p, they do not import it into the nucleus at discernable rates.

The DNA binding domain of Tfg2p is required for its nuclear accumulation

Tfg2p binds Tfg1p and DNA through its Tfg1p interaction domain (residues 1-230) and DNA binding domain (residues 280 to 400), respectively (Figure 5-2A). Tfg2p also interacts with at least seven polymerase subunits (Chung, Craighead et al. 2003) but details of these interactions are still unclear. It is possible that nuclear accumulation of Tfg2p is mediated through its interaction with Tfg1p, DNA or polymerase subunits.

Deletion of RAP30 residues 15-30 abolished its interaction with the Tfg1p human homolog (Tan, Conaway et al. 1995). We deleted the first 86 residues of Tfg2p (corresponding to the first 36 residues of RAP30) and visualized its localization in cells. Both the wild type and PY mutant of Tfg2p (87-400)-GFP were nuclear (Figure 5-9A). A larger N-terminal truncation of Tfg2p also had no effect on localization (Figure 5-9A) implying that interaction with Tfg1p is not required for Tfg2p import.

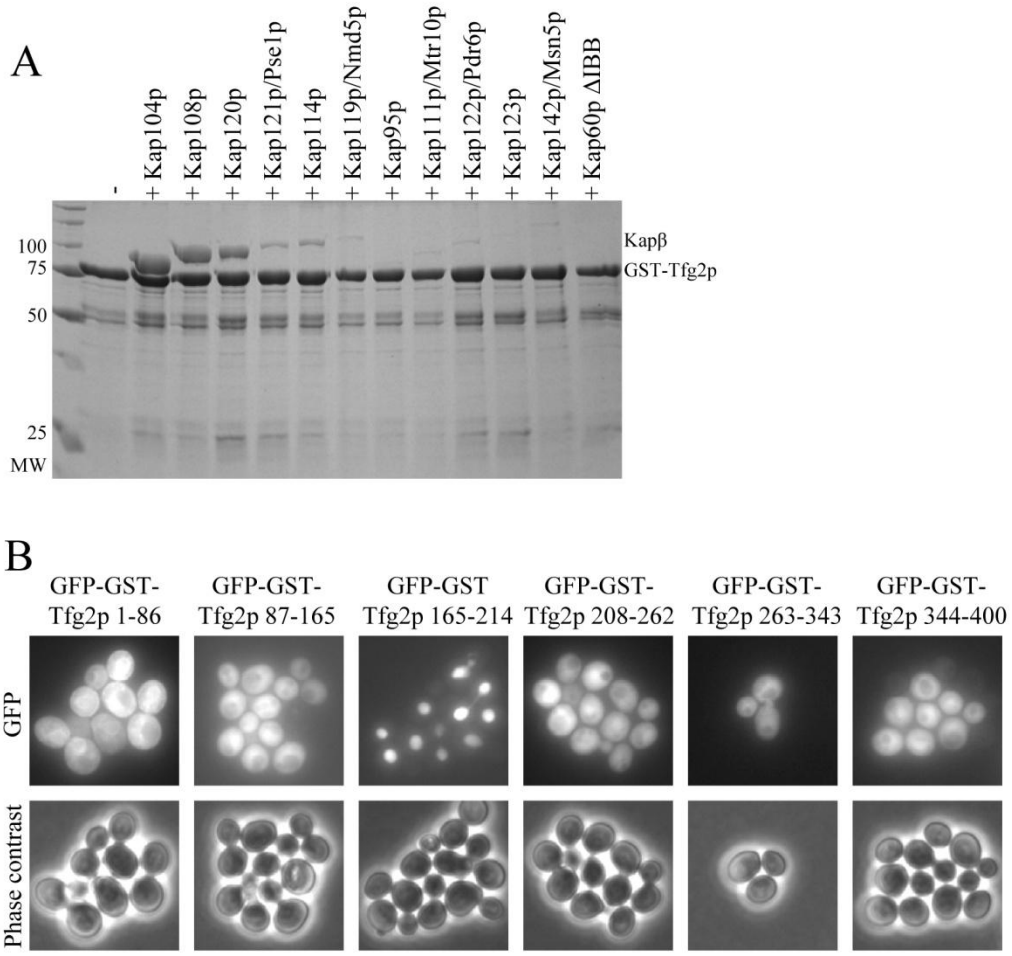


Figure 5-8: Other karyopherins bind Tfg2p, but they do not import it into the nucleus. A) Binding assay of recombinant karyopherins with immobilized full length Tfg2p. Proteins were visualized by Coomassie stain. B) GFP-GST-Tfg2p fragments were expressed in wild type yeast cells and analyzed by fluorescence microscopy and phase contrast. Only the PY-NLS (165-214) is nuclear.

To examine the role of DNA binding in Tfg2p nuclear accumulation, we made C-terminal truncations of Tfg2p (1-214, 1-262, and 1-343) to disrupt its DNA binding domain and test localization of the mutants in the absence and presence of the ^{203}PY^{204}/AA mutations. Although DNA binding domain mutants and the PY mutant of full length Tfg2p were nuclear, combined DNA binding domain and PY mutations mislocalized the proteins to the cytoplasm (Figure 5-9A). Even a truncation of only 56 residues in Tfg2p(1-343) had a dramatic effect on localization. These data show that the DNA binding domain of Tfg2p is necessary for nuclear accumulation.

Tfg2 is sequestered in the nucleus via DNA binding

The DNA binding domains of Tfg2p and RAP30 share 80% sequence similarity and the NMR structure of the RAP30 DNA binding domain places it in the winged helix turn helix family (Groft, Uljon et al. 1998). The domain has three N-terminal α-helices (helix 1-3) followed by a 2-strand β-sheet (Figure 5-10). A 5-residue loop between two strands is usually refered to as the "wing" (Groft, Uljon et al. 1998). The DNA-binding surface of RAP30 was previously mapped by DNA titration and ^{15}N-^{1}H correlation NMR spectroscopy (Groft, Uljon et al. 1998). We used the RAP30 data to mutate DNA binding residues in the highly homologous Tfg2p (Figure 5-1).

We made four site-specific DNA binding domain mutants of Tfg2p: 1) the helix 1-Tfg2p mutant with residues R293, K296, K297 on one side of helix 1 mutated to alanines, 2) the helix 2-Tfg2p mutant with residues S314, K316, K319, R323 of helix 2 mutated, 3) the helix 3-Tfg2p mutant with residues E331, D334, K335 in helix 3 mutated, and 4) the helix 3-wing-Tfg2p mutant with residues K335 of helix 3, K341 and K342 of

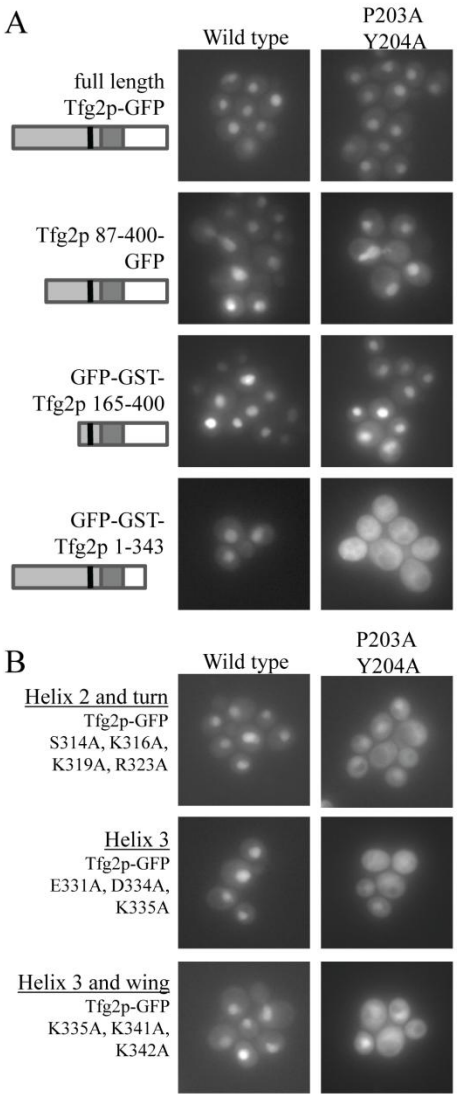


Figure 5-9: The DNA binding domain of Tfg2p is required for nuclear accumulation. A) Wild type cells expressing N-terminal and C-terminal truncations of Tfg2 tagged with either GFP or GFP-GST as indicated in the figure. Localization of proteins with ^{203}PY^{204}/AA mutations are in the right panels and localization of wild type proteins are in the left panels. Cells were analyzed by fluorescence microscopy. A small domain structure is under each label in corresponding shades of gray as to A. B) Wild type cells expressing GFP-tagged full length Tfg2p harboring point mutations in the DNA binding domain in either a wild type background or ^{203}PY^{204}/AA background. Cells were analyzed by fluorescence microscopy.

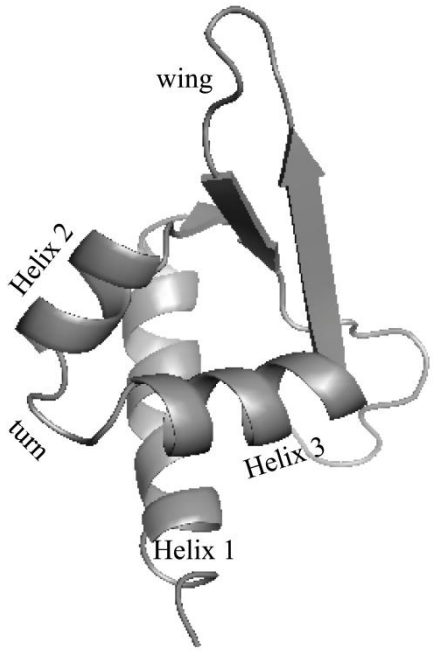


Figure 5-10: NMR structure of RAP30 DNA binding domain. Figure was made from pdb file 1bby using PyMOL (Groft, Uljon et al. 1998; DeLano 2002).

the wing mutated. We analyzed cellular localization of these mutants in a wild type NLS or ^{203}PY^{204}/AA background (Figure 5-9B). Mutations within helices 2, 3 and the wing did not affect nuclear localization in the wild type NLS background. However, when these DNA-binding mutations were coupled with the ^{203}PY^{204}/AA mutations, the Tfg2p mutants were mislocalized to the cytoplasm (Figure 5-9B). Mutations within helix α1 had no effect on localization (data not shown). These data suggest that sequestration of Tfg2p in the nucleus through its DNA binding domain accounts for its observed nuclear accumulation even when its primary nuclear import pathway Kap104p is compromised.

The PY-NLS is required for cell viability

In order to assess whether the PY-NLS and more specifically, the PY motif are necessary for cell viability we performed a plasmid shuttle assay. A *Tfg2* deletion strain containing a wild type *Tfg2* URA3 maintenance plasmid was transformed with either wild type, P203A,Y204A or ΔPY-NLS (Δ176-206) containing plasmids or vector alone. Cells were serially diluted on control plates or 5-fluoroorotic acid (5-FOA) containing plates to select against the *Tfg2* wild type plasmid (URA3). Since *Tfg2* is an essential gene, the loss of its expression would result in a loss of cell viability as is seen in the vector row of Figure 5-11. The expression of wild type Tfg2p rescues cell viability. The expression of the P203A,Y204A mutant is also able to support cell growth. However, cells expressing Tfg2p with a PY-NLS deletion (Δ176-206) are not viable. Thus, the PY-NLS is necessary for cell viability.

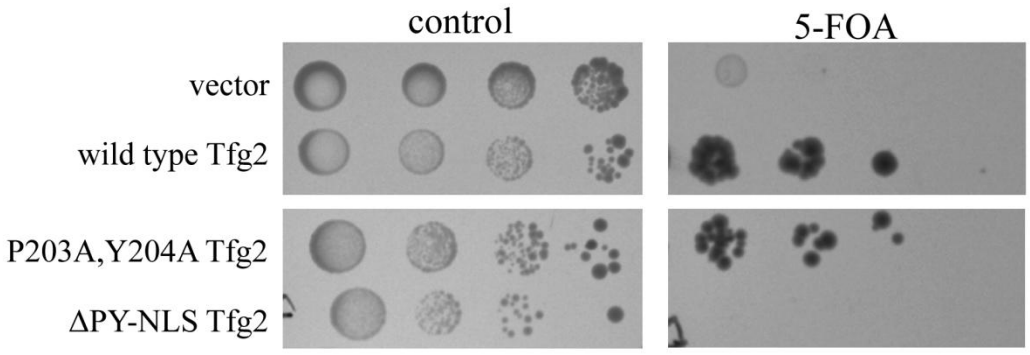


Figure 5-11: The PY-NLS of Tfg2p is necessary for cell viability. ΔTfg2 cells maintained by a wild type *Tfg2* plasmid (URA3) expressing either wild type or mutant Tfg2p. Cells were diluted on either a control plate or a plate containing 5-FOA to select against the wild type *Tfg2* maintenance plasmid (URA3).

Discussion

We show that Kap104p imports Tfg2p into the nucleus. Tfg2p is the first PY-NLS-bearing substrate identified through a bioinformatics search to be validated *in vivo*. Many human PY-NLSs were validated as NLSs *in vivo* prior to their identification as PY-NLSs (Suel, Gu et al. 2008). We show that the PY-NLS of Tfg2p is sufficient to target GFP-GST to the nucleus of wild type but not Kap104p temperature sensitive cells, and mutation of the PY motif mislocalized the NLS to the cytoplasm. Binding assays confirmed that mislocalization of mutant PY-NLS was solely due to loss of recognition by Kap104p and not other Kapβs. However, disruption of Kap104p-mediated import, either in Kap104p temperature sensitive cells or by mutagenesis of the PY motif, did not mislocalize the full length protein. Instead, Tfg2p was mislocalized to the cytoplasm only when both DNA binding residues and the PY motif were mutated. It appears that Tfg2p is still imported into the nucleus probably via a lower affinity and slower backup method when Kap104p-mediated transport is disrupted. Tfg2p can bind other Kapβs. Its PY-NLS interacts weakly with Kap108p, Kap121p, Kap114p, Kap119p, and Kap95p, and it also binds Kap108p and Kap120p through unidentified NLSs. Tfg2p may be imported into the nucleus by any of these karyopherins, albeit inefficiently. However, once full length Tfg2p gets in the nucleus, it binds DNA and other members of the PIC and is retained there. Importance of the Kap104p import pathway in targeting Tfg2p to the nucleus is revealed only when DNA binding and thus nuclear retention is disrupted.

Identification of Tfg2p as the third Kap104p substrate reinforces the physical properties we have previously found to govern PY-NLS recognition by Kap104p. The

PY-NLS is a modular signal with three epitopes, each of which contributes differently to total binding energy in different NLSs. Like another Kap104p substrate, Hrp1p, Tfg2p has a strong epitope 3 and a weak epitope 1. Similarly, its PY dipeptide is necessary for Kap104p binding and nuclear import by Kap104p (Lange, Mills et al. 2008; Suel, Gu et al. 2008). In contrast, another Kap104p substrate, Nab2p, has a strong epitope 1 and a medium strength epitope 3. Mutagenesis of Nab2p's epitope 3 only slightly mislocalized the protein to the cytoplasm (Suel, Gu et al. 2008). Despite different distribution of epitope strengths, Tfg2p, Hrp1p and Nab2p all bind Kap104p with similar affinity (10-37 nM), consistent with the concept that combinatorial mixing of different epitope strengths can result in NLSs that are diverse in sequence but still functional.

We had previously predicted an additional binding hotspot at the aromatic residue two amino acids N-terminal of the PY motif. Alanine mutation of this residue in both Hrp1p and Nab2p decreased Kap104p binding affinity (Suel, Gu et al. 2008). Similarly, the Y201A mutant of Tfg2p PY-NLS showed decreased Kap104p binding and nuclear accumulation, suggesting that this residue is a binding hotspot. Since the positions two residues N-terminal of the PY motif in all three Kap104p substrates and several Kapβ2 substrates are occupied by aromatic residues, this sequence pattern could be used as an additional restraint for future substrate searches (Lee, Cansizoglu et al. 2006; Suel, Gu et al. 2008).

Damelin et al. suggested four criteria to verify a putative NLS as a functional targeting signal recognized by a specific Kapβ pathway (Damelin, Silver et al. 2002). First, the NLS and full length substrate must bind the Kapβ directly and be dissociated by RanGTP. Second, the sequence must be necessary for import such that deletion or

mutagenesis of the sequence should mislocalize the substrate. Third, the NLS must be sufficient to target another protein into the nucleus. Finally, mutations in the transport pathway should inhibit nuclear import of the substrate. These criteria work well under simple circumstances but validation according to the criteria can often be thwarted by complexities in the cell. In the case of Tfg2p it was straight-forward to invalidate the cNLS as a functional NLS, but it was significantly more difficult to validate the PY-NLS as a functional targeting sequence. Cellular functions of Tfg2p as a transcription factor and as a nuclear resident rather than a shuttling protein dictated additional experiments to verify it as a Kap104p substrate.

Verification of NLSs in other proteins has also been complicated by cellular conditions and exemplifies the many caveats of the above validation criteria. Import substrates may be transported into the nucleus in a piggyback manner through interactions with another substrate, be imported by multiple karyopherins, be retained in the nucleus, or imported only when modified appropriately. For example, although Kap121p imports Nop1p into the nucleus, Leslie et al. did not see mislocalization in a Kap121p temperature sensitive strain since Nop1p is retained in the nucleolus (Leslie, Zhang et al. 2004). They had to first inactivate Kap121p and then turn on Nop1p expression using a galactose inducible promoter to see mislocalization of Nop1p (Leslie, Zhang et al. 2004). Many yeast proteins are imported into the nucleus by multiple mechanisms. Leslie et al. also showed that Sof1p can be imported directly by Kap121p or through a piggyback mechanism where it binds Nop1p (Leslie, Zhang et al. 2004). Additionally, many proteins such as histones and ribosomal proteins are imported into the nucleus by multiple Kapβs (Rout, Blobel et al. 1997; Mosammaparast, Jackson et al.

2001; Mosammaparast, Guo et al. 2002). Post-translational modifications can also affect protein localization, making it more difficult to determine how they are imported into the nucleus. The criteria outlined by Damelin et al. are useful guidelines and can be modified according to cellular functions of individual substrates.

The steady state localization of a nuclear protein is not solely dictated by the presence or absence of functional nuclear import or export signals. Strengths of the respective targeting signals and strengths of interactions with other binding partners in either the nucleus or the cytoplasm (nuclear or cytoplasmic retention) all contribute to protein localization. Retention in cellular compartments may result from signal masking, anchorage to cellular structures or promotion of degradation (Ziegler and Ghosh 2005). Additionally, localization is often altered in response to changes in the extra- or intracellular environment such as resulting from stimuli, stress and variations in the cell cycle or developmental period (Cardoso and Leonhardt 1999; Makhnevych, Lusk et al. 2003). Here, we show the example of Tfg2p whose nuclear steady state localization is a result of a combination of nuclear import through its PY-NLS and nuclear retention through its DNA binding domain.

Many of the Kapβs responsible for nuclear import of general transcription factors in yeast have been identified. Kap114p imports TFIIB and the TATA-binding protein (TBP) (Morehouse, Buratowski et al. 1999; Pemberton, Rosenblum et al. 1999; Hodges, Leslie et al. 2005), Kap119p imports TFIIS and Kap122p imports TFIIA (Albertini, Pemberton et al. 1998; Titov and Blobel 1999). These Kapβs also import proteins other than transcription factors. For example, Kap114p also imports proteins involved in chromatin assembly such as histones H2A, H2B and chromatin assembly factor Nap1p

(Mosammaparast, Jackson et al. 2001; Mosammaparast, Ewart et al. 2002). The two previously identified Kap104p substrates, Hrp1p and Nab2p, are both mRNA binding proteins. *In vivo* validation of Tfg2p as a PY-NLS bearing substrate of Kap104p implies that Kap104p substrates are not limited to mRNA binding proteins. Similarly, 16 of the 17 validated human Kapβ2 substrates bind RNA, and of these three are also transcription factors (Bertolotti, Melot et al. 1998). Both Kap104p and Kapβ2 appear to transport a range of functionally related substrates such as transcription factors and RNA binding or processing proteins. However many more Kap104p and Kapβ2 substrates will need to be identified and validated before we can truly define a functional network that relates the groups. Cellular and energetic dissections of the Tfg2p PY-NLS corroborates recently reported physical properties that govern Kap104p-PY-NLS interactions and suggests additional sequence restraints which will be useful for large-scale substrate searches in the future (Suel, Gu et al. 2008).

CHAPTER 6

LEUCINE-RICH NUCLEAR EXPORT SIGNAL RECOGNITION BY CRM1

Abstract

The first structure of the export karyopherin CRM1 bound to its substrate Snurportin (SPN1) has recently been solved (Dong, Biswas et al. 2009). The nuclear export signal (NES) of SPN1 is bipartite; epitope I is a previously unidentified canonical leucine-rich NES at the N-terminus of SPN1 and epitope II is a conformational NES in the nucleotide binding domain. We utilized biochemical and cellular techniques to analyze SPN1's leucine-rich NES. We show that the 15-residue NES binds CRM1 and is sufficient to target a SV40 NLS-GFP-GFP fusion protein to the nucleus of yeast cells in a CRM1 dependent manner. Additionally, residues L4A and L8A of SPN1 are required for efficient CRM1 binding and nuclear export. These results confirm that residues 1-15 of SPN1 are a functional leucine-rich NES recognized by CRM1.

Introduction

Nuclear localization and export signals (NLSs and NESs) direct macromolecules in and out of the nucleus, respectively. In human cells, a family of 19 different Karyopherinβ (Kapβ/Importin/Exportin) proteins recognizes distinct nuclear transport signals to target their transport substrates to the nuclear pore complex (NPC) (Tran, Bolger et al. 2007). Kapβ-substrate interactions and transport directionality are regulated through interactions with the Ran GTPase (Weis 2003).

Although several classes of NLSs have been characterized (Dingwall, Sharnick et al. 1982; Kalderon, Richardson et al. 1984; Lanford and Butel 1984; Lee, Cansizoglu et al. 2006), only one class of NES termed the leucine-rich NES (LR-NES) is currently known (Wen, Meinkoth et al. 1995). LR-NESs are recognized by the export-Kapβ CRM1 (also known as Exportin1 or Xpo1) (Fornerod, Ohno et al. 1997; Fukuda, Asano et al. 1997; Ossareh-Nazari, Bachelerie et al. 1997; Richards, Carey et al. 1997; Stade, Ford et al. 1997). The 10-15 residue export signal was first identified in HIV-1 Rev and cAMP-dependent protein kinase inhibitor (Fischer, Huber et al. 1995; Wen, Meinkoth et al. 1995). NESs in both proteins contain several leucines and early mutagenic studies of the signal led to a consensus sequence L-X_{2-3}-[LIVFM]-X_{2-3}-L-X-[LI] where X is any amino acid (Bogerd, Fridell et al. 1996). However, a study of 75 experimentally identified LR-NESs showed their sequences to be diverse with no strict requirement for leucine residues and only 36% matched the consensus (la Cour, Kiemer et al. 2004). The traditional leucine-rich consensus was thus expanded to ϕ-X_{2-3}-ϕ-X_{2-3}-ϕ-X-ϕ where ϕ is L,

V, I, F or M. Although the broad consensus matches most known LR-NESs, it is so vague that the motif would be found in most proteins. Like the multipartite PY-NLS (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007) and so many other vague organelle targeting signals (Mancias and Goldberg 2005; Swanton and High 2006), the LR-NES is a complex and diverse signal that cannot be sufficiently described by consensus sequence alone. Instead, the signal will need to be described within the structural and physical context of CRM1-LR-NES interactions. Nevertheless, effective CRM1 inhibition by antifungal antibiotic Leptomycin B coupled with vague NES sequence patterns has facilitated identification of hundreds of CRM1 substrates that function in a broad range of cellular processes (Nishi, Yoshida et al. 1994; Matsuyama, Arai et al. 2006).

Although >50% of CRM1 substrates contain sequences that match LR-NES consensus sequences, others lack recognizable canonical LR-NES. One such substrate is Snurportin1 (SPN1), which imports spliceosomal U snRNPs into the nucleus (Paraskeva, Izaurralde et al. 1999). SPN1 contains an N-terminal Importinβ-binding domain (IBB) and a C-terminal nucleotide-binding domain (NBD) (Huber, Cronshagen et al. 1998). The latter binds the 5'-2,2,7-terminal trimethylguanosine (m_3G) cap of snRNAs and the IBB domain of SPN1 binds Importinβ, thus importing snRNP particles into the nucleus. Following nuclear import, SPN1 is recycled back to the cytoplasm by CRM1. The export-Kapβ interacts with a region of SPN1 that spans both IBB domain and the NBD. Therefore, unlike the small linear LR-NES, the NES of SPN1 was predicted to be a large conformational signal (Paraskeva, Izaurralde et al. 1999).

Dong *et al.* solved the 2.9 Å crystal structure of CRM1 bound to SPN1 (Dong, Biswas et al. 2009). The 20 HEAT repeats of CRM1 form a ring-shaped protein and SPN1 binds in a bipartite fashion to the central convex surface of CRM1 (Figure 6-1). Epitope I is a previously undetected LR-NES (Figure 6-1). Residues 1-16 of SPN1 form an amphipathic α-helix that occupies a hydrophobic groove between outer helices of CRM1 HEAT repeats 11 and 12. The second NES epitope is a basic conformational epitope located on the SPN1 nucleotide-binding domain, which interacts with an acidic patch on CRM1 adjacent to the LR-NES site (Figure 6-1).

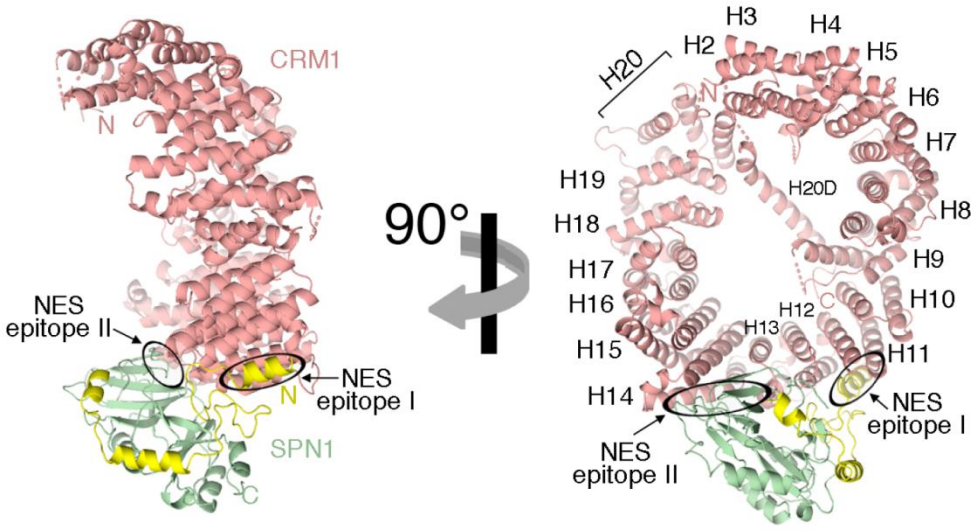


Figure 6-1: Overall structure of the CRM1-SPN1 complex. Orthogonal views of the CRM1-SPN1 complex. CRM1 is pink and the IBB and NBD domains of SPN1 are yellow and green respectively. HEAT repeats 2-20 of CRM1 are labeled H2-H20.

The N-terminus of SPN1 has the sequence MEELSQALASSFSVSQ and residues 4-14 matches the broad ϕ-X_{1-3}-ϕ-X_{2-3}-ϕ-X-ϕ LR-NES consensus. The LR-NES (SPN1 residues 1-16) is a structurally independent module that makes no contact with the nucleotide-binding domain and protrudes away from the rest of the protein. Furthermore, its location at the N-terminus and immediately preceding a long random coil region agrees with previous prediction that the LR-NES is a highly accessible α-helix that is located in surface exposed and very flexible regions of proteins (la Cour, Kiemer et al. 2004).

The LR-NES of SPN1 binds a hydrophobic groove between CRM1 helices H11A and H12A such that the three helices run parallel to each other (Figure 6-2 A and B). SPN1 sidechains M1, L4, L8, F12 and V14 line one side of the LR-NES and are buried in the groove through hydrophobic contacts with CRM1 (Figure 6-2). The solvent accessible face of the LR-NES helix is composed of polar residues E2, E3, S5, Q6 and S10. Interactions at the LR-NES interface bury 844 $Å^2$ on each protein. Biochemical and cellular studies are needed to support the discovery of the LR-NES from the structural studies.

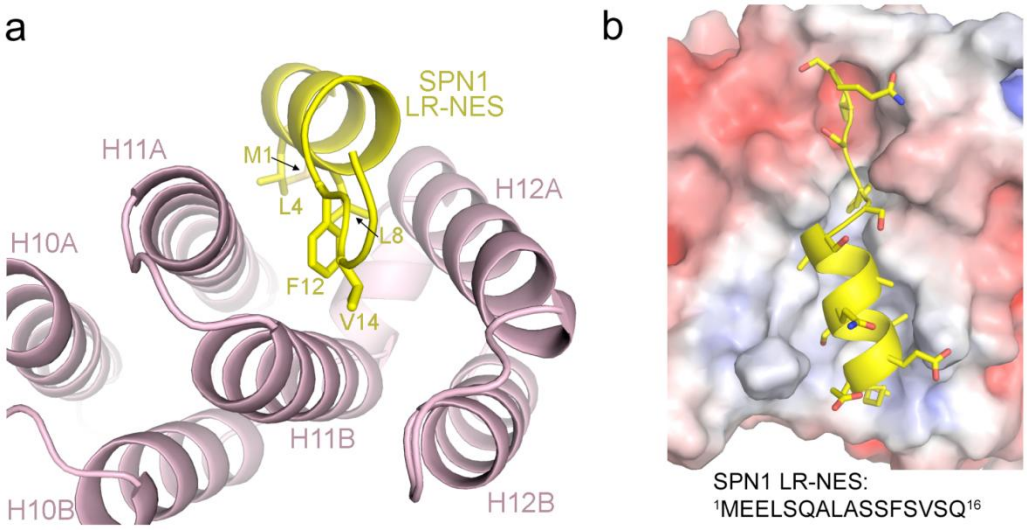


Figure 6-2: The LR-NES binding site. A) Arrangement of the CRM1 (pink) and the SPN1 LR-NES (yellow) helices. The SPN1 segment containing residues 1-16 (yellow) is a LR-NES. The signal binds between helices H11A and H12A of CRM1 (pink). B) The LR-NES interface showing the electrostatic surface potential of CRM1 and a ribbon diagram of the SPN1 LR-NES.

Materials and Methods

In vitro binding assays

All of the binding assays were performed by Anindita Biswas in the Chook lab. Site–directed mutagenesis of SPN1 and CRM1 were performed using the QuickChange method (Stratagene) and all constructs were sequenced. The SPN1(1-15) NES as well as NESs of Rev, NS2 and Nmd3 were generated by ligation of annealed oligonucleotides into the pGEX-Tev vector. ~15 µg of immobilized GST-SPN1 or GST-NESs proteins (glutathione sepharose; GE Healthcare) were incubated with ~120 µg of CRM1 proteins in the presence or absence of RanGTP (5 fold molar excess). After extensive washing, bound proteins were visualized by SDS-PAGE and Coomassie staining.

Plasmids

The NLS-GFP-GFP and NLS-GFP-GFP-SPN1 fusion genes were cloned into a modified pRS415 (*CEN6, ARS, LEU2,* and AP^R) shuttle vector containing a 5' ADH1 promoter (Sikorski RS and Hieter 1989). The first GFP gene was cloned into SpeI and PstI sites and included a 5' ggsgg linker (Figure 6-3). The second GFP was inserted into PstI and XhoI sites and included a 5' ggsgg linker and engineered BamHI and NotI sites. The SV40 NLS (PKKKRKV) and a ggsgg linker were inserted upstream of the GFP genes using an extended primer. The SPN1 gene was cloned into the BamH1 and NotI sites 3' of the GFP genes. The whole NLS-GFP-GFP-SPN1 1-15 insert was also subcloned into a pRS416 (*CEN6, ARS, URA3*, and AP^R) shuttle vector. Plasmids are listed in Table 7-7.

Figure 6-3: SV40 NLS-GFP-GFP-SPN1 in pRS415/4.

Cell culture and microscopy

For the transport competition assay BY4741 cells (*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) expressing pRS415 plasmids were grown overnight at 30°C in SC-Leu media. Cells were transferred to a 1.5% low melting agarose pad made with SC-leu in a coverslip bottom Wilco dish. Cells were observed on an Olympus IX-81 inverted microscope (60X objective) and images were acquired with a Hamamatsu ORCA-ER camera. All images were analyzed in Image-Pro Plus software (Media Cybernetics).

For the temperature sensitive cell assay pRS416 NLS-GFP-GFP-SPN1 1-15 was transformed into the *xpo1-1* strain (a gift from K. Weis, plasmid KWY121, Stade et al 1997) and BY4741 cells. Cells were grown overnight at room temperature and analyzed as above. After the fluorescence was examined at room temperature, the temperature in the incubation chamber was increased to 37°C. After one hour, the reporter fluorescence was examined in the same cells.

Results

SPN1 1-15 is a functional NES exported by CRM1 in S. cerevisiae cells

SPN1 binds CRM1 with a tighter affinity than other CRM1 substrates due to the presence of two binding epitopes. Unlike other substrates, SPN1 bound CRM1 even in the absence of RanGTP (Figure 6-4). The addition of RanGTP increased the amount of CRM1 bound by full length SPN1. We wanted to test if the LR-NES identified in the structure of SPN1 bound to CRM1 is capable of binding CRM1. CRM1 bound immobilized GST-SPN1 NES (1-15) in the presence of RanGTP (Figure 6-4).

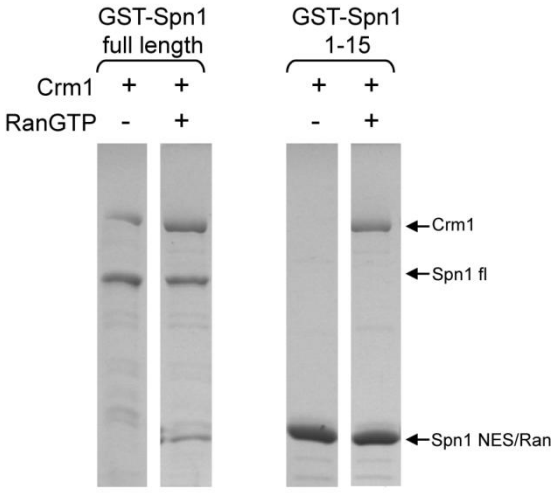


Figure 6-4: SPN1 NES binds CRM1p in the presence of RanGTP. CRM1 was added to immobilized GST-SPN1 full length or NES (1-15) in the presence and absence of RanGTP.

To determine whether the SPN1 NES is a functional export signal we utilized a nuclear transport competition assay in *S. cerevisiae* (Stade et al 1997). The SV40 NLS was fused to two GFP molecules and expressed in wild type yeast cells. The NLS-GFP-GFP fusion protein was nuclear at steady state (Figure 6-8). The addition of the fifteen residue SPN1 NES resulted in enrichment of the fusion protein in the cytoplasm.

To examine if SPN1 is being exported by the CRM1 yeast homolog, Xpo1, the NLS-GFP-GFP-SPN1 NES fusion protein was expressed in *xpo1-1* temperature sensitive cells (Stade et al 1997). Thirteen of the fourteen residues in CRM1 that make significant contacts with epitope I of SPN1 are identical in Xpo1 (Figure 6-6). At the permissive temperature, the reporter was predominantly localized in the cytoplasm (Figure 6-5). After shifting the cells to a non-permissive temperature, the reporter protein accumulated in the nucleus (Figure 6-5). The change in temperature had no effect on localization of the reporter in wild type cells. Thus SPN1 NES is a functional NES exported by CRM1.

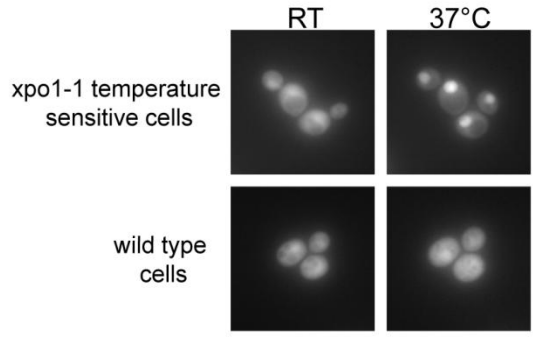


Figure 6-5: *Xpo1-1* temperature sensitive cells (top) and wild type cells (bottom) expressing NLS-GFP-GFP-SPN1 1-15 fusion protein. The location of the fusion protein was analyzed by fluorescence microscopy in the same cells at the permissive temperature (room temperature) and at the non-permissive temperature (37°). While the reporter remains cytoplasmic in the wild type cells, it accumulates in the nucleus of the *xpo1-1* temperature sensitive cells at the non-permissive temperature.

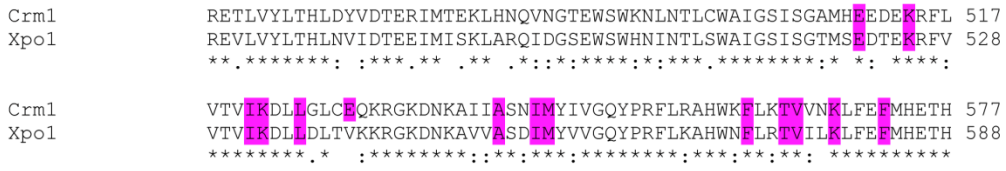


Figure 6-6: Alignment of human CRM1 with yeast Xpo1. Residues that contact the NES of SPN1 are highlighted in magenta.

L4 and L8 are necessary for efficient export of SPN1

The two leucine residues in SPN1 1-15 are buried in the groove through hydrophobic contacts with CRM1. In order to test if these residues are necessary for recognition by CRM1 we performed *in vitro* binding assays. CRM1 did not bind immobilized full length L4A, L8A SPN1 even in the presence of RanGTP (Figure 6-7). Thus, residues L4 and L8 are required for CRM1 binding.

To examine if these residues are necessary for NES recognition in cells, we mutated these residues in the NLS-GFP-GFP-SPN1 NES reporter. When L4 and L8 of SPN1 were both mutated to alanines, the reporter protein accumulated in the nucleus (Figure 6-8). The competition assay was repeated with SPN1 1-342 with similar results. The increase in nuclear accumulation of wild type SPN1 1-342 compared to the NES is presumably due to a NLS between residues 25-65 (Huber J et al 1998). These data show that L4 and L8 are necessary for efficient export of SPN1 by CRM1.

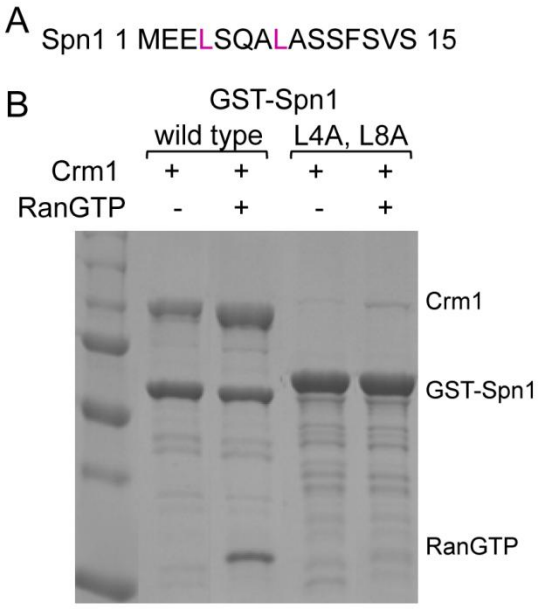


Figure 6-7: L4 and L8 are necessary for recognition of SPN1 by CRM1. A) Sequence of SPN1 NES. B) CRM1 was added to immobilized wild type or mutant GST-SPN1 in the presence or absence of RanGTP.

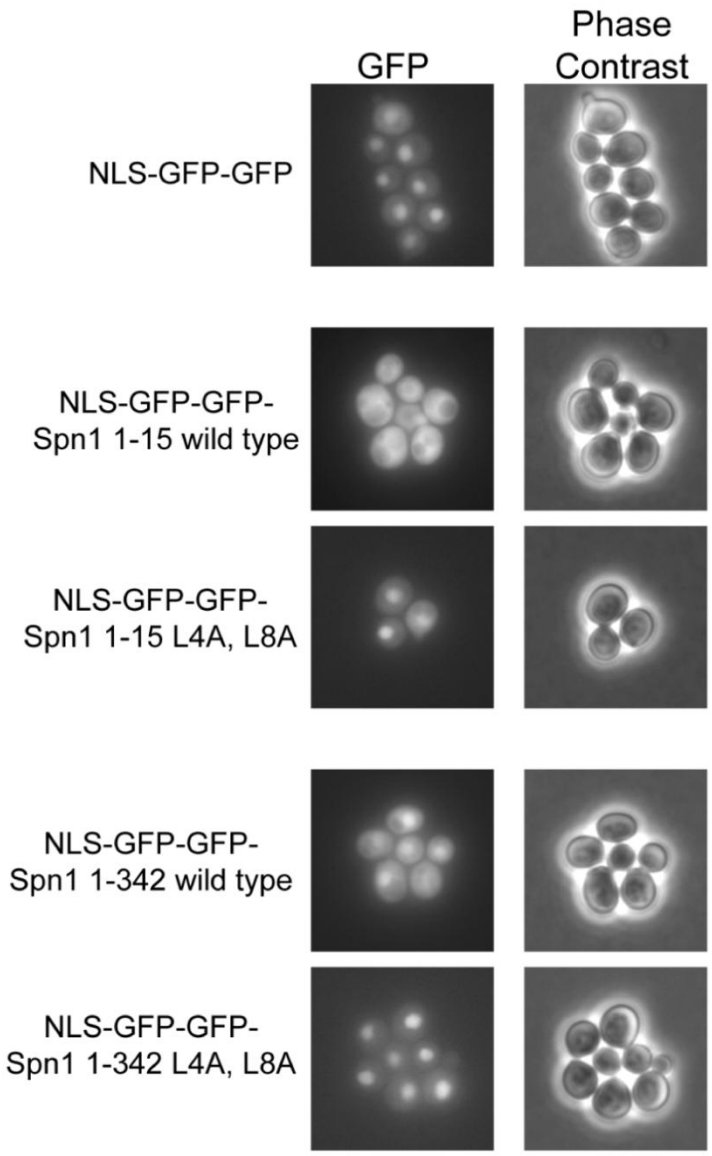


Figure 6-8: Wild type *S. cerevisiae* cells expressing either NLS-GFP-GFP or NLS-GFP-GFP-SPN1 fusion proteins were analyzed by fluorescence microscopy and phase contrast. The reporter is localized to the cytoplasm in the cells expressing NLS-GFP-GFP-SPN1 wild type protein, but accumulates in the nucleus of the L4A, L8A mutants. GFP is displayed in the same fluorescence scale in each panel.

Discussion

The biochemical and cellular evidence presented here supports the conclusion from the x-ray crystallography structure that the N-terminus of SPN1 is a leucine rich NES recognized by CRM1. Not only does the 15-residue NES bind CRM1 in the presence of RanGTP, but its export from the nucleus of yeast cells is dependent on Xpo1. SPN1 is not a naturally occurring protein in yeast, yet it is recognized and exported by the yeast homolog of CRM1 demonstrating the generality of this export mechanism.

The two leucines in the SPN1 NES, L4A and L8A, are required for efficient recognition and export of the protein. When the two leucines are mutated in full length SPN1, it does not bind CRM1. Additionally, when these two residues are mutated in either SPN1 NES or full length, the fusion protein is mislocalized to the nucleus. SPN1 1-342 only binds CRM1 with an affinity of 1.5 μM and the NES is only able to bind CRM1 in the presence of RanGTP, but the NES is sufficient to target GFP to the cytoplasm even in the presence of a SV40 NLS. Almost full length SPN1 (1-342) is also able to target SV40 NLS-GFP-GFP to the cytoplasm even though it contains a functional NLS in addition to the SV40 NLS (Huber, Cronshagen et al. 1998). Timney et al. have suggested that the fast import rate of the import factor Kap123p is due to its high cellular concentration. It is possible that a discrepancy in export versus import Kap concentrations results in a steady state cytoplasmic localization even in the presence of a functional NLS. An accurate measurement of Xpo1, Kap95p and Kap60p physiological concentrations will be necessary to further examine this idea.

The NES of SPN1 has two binding epitopes, epitope I is a leucine rich NES and epitope II is a conformational epitope in the nucleotide binding domain. This is similar to the multipartite PY-NLS recognized by Kap104p/Kapβ2 (Lee, Cansizoglu et al. 2006; Suel, Gu et al. 2008). The PY-NLS has three binding epitopes that vary in strength in different NLSs. Epitope 3 of Hrp1 is strong and is necessary for Kap104p binding, but mutagenesis of epitope 3 in Nab2p only results in a ten-fold decrease in binding affinity (Suel, Gu et al. 2008). Strong and weak mixing of different epitopes increases the diversity of this signal. It appears that in the case of SPN1 epitope I is the strong epitope since mutagenesis of only two leucine residues significantly decreases CRM1 binding and CRM1-mediated export. It remains to be seen if this will be a general property of CRM1 substrates due to the lack of structures and thermodynamic analyses of different CRM1 substrates.

CHAPTER 7

CONCLUSIONS

Karyopherin βs recognize their substrates through binding nuclear localization or nuclear export signals. Previously, the only characterized NLS was the classical NLS (cNLS) recognized by Kapα/Kapβ1. However, through structural, biochemical, biophysical and cellular studies we have identified and characterized a second class of NLS, the PY-NLS recognized by Karyopherin β2 (Kapβ2). While the cNLS is characterized by a well-defined, compact consensus sequence, the NLS recognized by Kapβ2 is larger and more diverse in sequence. The 20-30 residue PY-NLS is characterized by three rules as revealed from structural and biochemical studies of Kapβ2 bound to the NLS of hnRNP A1 (Lee, Cansizoglu et al. 2006). These rules include overall basic character, structural disorder, and a weak consensus sequence composed of an N-terminal hydrophobic (Φ_1-G/A/S-Φ_3-Φ_4, where $\Phi_1$is a hydrophobic residue and Φ_3 and Φ_4 are hydrophobic residues or R or K) or basic (K/R-X_{0-6}-K/R-X_{0-6}-K/R-X_{0-6}-K/R) motif and a C-terminal R/K/H-X_{2-5}-PY motif. The characterization of the PY-NLS has led to the identification of approximately one hundred substrates of Kapβ2 by bioinformatics (Lee, Cansizoglu et al. 2006). Thirteen proteins have previously been identified and their NLSs validated in vivo and in vitro while ten potential NLSs have been validated in vitro.

Analysis of the two previously reported substrates of the yeast homolog of Kapβ2, Kap104p, which have previously been termed rg-NLSs, revealed that they are PY-NLSs, demonstrating that the PY-NLS is recognized by two evolutionarily distant homologs (Lange, Mills et al. 2008; Suel, Gu et al. 2008). While Kapβ2 recognizes both the hydrophobic and basic subclasses of PY-NLS, Kap104p only recognizes the basic subclass of PY-NLSs (Suel, Gu et al. 2008). This difference can be explained through analysis of previously reported Kapβ2-NLS structures (Lee, Cansizoglu et al. 2006; Cansizoglu, Lee et al. 2007; Imasaki, Shimizu et al. 2007; Suel, Gu et al. 2008).

Mutagenesis and thermodynamic analysis of Kap104p-NLS interactions revealed physical properties that govern PY-NLS binding affinity (Suel, Gu et al. 2008). PY-NLSs contain at least three energetically significant binding epitopes which can each accommodate large sequence diversity. While energetic cooperativity is observed within linear epitopes, each epitope is energetically quasi-independent. Finally, each linear epitope can vary significantly in its contribution to total binding energy. The modular organization and energetic independence of each binding epitope in the PY-NLS enables the affinity of the NLS to be tuned to optimize binding to the karyopherin and release by RanGTP.

The lack of Kap104p substrates is a hindrance to the refinement of the consensus sequence and further analysis of the energetic properties of the PY-NLS. A bioinformatics search utilizing new insights from the thermodynamic study identified 173 potential PY-NLSs. 21 of the 39 potential NLSs tested bound Kap104p and were dissociated by RanGTP. Six of these NLSs were functional targeting signals in cells,

capable of targeting GFP to the nucleus. Only one of these potential PY-NLSs, Tfg2p, was solely dependent on Kap104p for import into the nucleus. The other five NLSs bound at least one other karyopherin demonstrating the redundancy among yeast Kapβ.

We identified and validated Tfg2p as a PY-NLS bearing substrate of Kap104p. While Kap104p is the primary nuclear import pathway of Tfg2p it also employs a secondary Kapβ in the absence of Kap104p. Once inside the nucleus it is retained via specific residues in its DNA binding domain. Thus, the cellular localization of Tfg2p is determined by at least two factors: the presence of an NLS and retention in the nucleus. This demonstrates the general concept that the localization of a protein is not dictated solely by the presence or absence of a nuclear localization or export signal.

Proteins bearing a nuclear export signal (NES) are targeted to the cytoplasm through recognition by an export Kapβ. Dong *et al*. have recently solved the x-ray crystallography structure of the export Kapβ CRM1 bound to one of its substrates, SPN1 (Dong, Biswas et al. 2009). The structure revealed that the NES of SPN1 has two binding epitopes: epitope I is a canonical leucine-rich NES while epitope II is a larger conformational epitope. Through biochemical and cellular experiments we confirmed that the leucine rich NES is sufficient for CRM1 recognition and export to the cytoplasm. We also showed that the two leucine residues are necessary for efficient export by CRM1.

The NES of SPN1 recognized by CRM1 and the PY-NLS recognized by Kapβ2/Kap104p are both multipartite targeting signals. The properties of the PY-NLS that make it difficult to identify also make it robust against perturbations. Additionally, the combinatorial mixing of energetically strong and weak epitopes increases the

diversity of the signal. In this work we analyzed the energetic contributions of specific residues in strong, medium and weak epitopes. The identification of sequences that bind Kap104p, in addition to those that are not recognized by Kap104p will be useful in future computational studies of the PY-NLS. The NLSs that do not bind Kap104p may contain three weak epitopes. Further dissection of the sequence requirements of the three epitopes is needed. Additionally, it is possible that the multipartite nature of the NLS may not solely be a property of the PY-NLS. The lack of consensus sequences and the significant diversity among known NLSs recognized by other import Kapβs suggests that these NLSs may have multiple binding epitopes as well. Biochemical and thermodynamic analysis of known NLSs/substrates for other Kapβs will be needed to investigate this idea. Currently, there are only consensus sequences for the classical NLS and the PY-NLS.

While it might be useful to limit the search for new PY-NLSs in the beginning to mRNA binding proteins, the search must be expanded to include all proteins. Nab2p and Hrp1p are both mRNA binding proteins, but Tfg2p is a transcription factor. Three of the validated substrates of Kapβ2 are also transcription factors. While it is intriguing to think that each Kapβ has its own sub-set of substrates, this may not be the case for every karyopherin. Kap111p is the only yeast Kapβ where all of its substrates (three) have the same function: mRNA processing. Additionally, it is known that many functional groups of proteins are transported by multiple Kapβs. For example, general transcription factors are imported by at least four different Kapβ proteins. The separate allocation of transport responsibility may be a viability mechanism of the cell.

Another mechanism may be the significant redundancy among yeast Kapβs and the substrates they recognize and transport. We have identified five 50-residue long NLSs recognized by at least two karyopherins. While the specific NLS consensus sequences for the majority of Kapβs are unknown, several NLSs are known to be basic enriched, including those recognized by Kap121p and the well studied cNLS. The basic enriched region of the PY-NLS may be the source of the redundancy for these five sequences. Further biochemical, mutagenesis and cellular studies will be needed to parse the specific contributions of each residue to each Kapβ binding in addition to evaluating the cellular implications of the redundancy.

We have identified a new class of NLS, rules for its recognition and physical properties that govern its interactions with Kapβ2/Kap104p. The PY-NLS varies from the classical NLS because it is an extended signal made up of three binding epitopes of different energetic strengths. The modular organization of the PY-NLS may emerge as a common feature among other targeting signals, as we have already shown that the NES of SPN1 recognized by the export protein CRM1 is also a multipartite signal. Given the sequence diversity among NLSs recognized by other Kapβs, it is possible that these signals contain multiple binding epitopes. Identification of the PY-NLS and its subsequent analysis establishes a paradigm for the characterization of other targeting signals.

APPENDIX

Plasmids

Table 7-1: Karyopherins

Gene	Plasmid	Sites
Kap104p	pGexTev	BamHI/NotI
Kap95p	pGexTev	BamHI/NotI
Mtr10p/Kap111p	pGexTev	BamHI/NotI
Kap123p	pGexTev	BamHI/NotI
Kap121p/Pse1p	pGexTev	BamHI/NotI
Kap119p/Nmd5p	pGexTev	SmaI/NotI
Kap108p/Sxm1p	pGexTev	EcoRI/NotI
Kap114p	pGexTev	SalI/NotI
Kap120p	pGexTev	SalI/NotI
Kap122p/Pdr6p	pGexTev	BamHI/NotI
Msn5p/Kap142p	pGexTev	BamHI/NotI
Kap60p	pGexTev	BamHI/NotI
Kap60p ΔIBB (88-543)	pGexTev	BamHI/NotI

Table 7-2. Hrp1 plasmids.

Gene	Vector	Fragment	Mutations	KS#
Hrp1	pGexTev	full length	wild type	6
Hrp1	pMalTev	full length	wild type	6
Hrp1	pGexTev	494-534	wild type	7
Hrp1	pMalTev	494-534	wild type	7
Hrp1	pGexTev	474-534	wild type	8
Hrp1	pMalTev	474-534	wild type	8
Hrp1	pGexTev	494-534	R506A, S507A, G508A, G509A	421g
Hrp1	pMalTev	494-534	R506A, S507A, G508A, G509A	421m
Hrp1	pGexTev	494-534	R512A, R513A, N514A, G515A	422g
Hrp1	pMalTev	494-534	R512A, R513A, N514A, G515A	422m
Hrp1	pGexTev	494-534	R516A, G517A, G518A	423g
Hrp1	pMalTev	494-534	R516A, G517A, G518A	423m
Hrp1	pGexTev	494-534	R519A, G520A, G521A, Y522A, N523A	424g
Hrp1	pMalTev	494-534	R519A, G520A, G521A, Y522A, N523A	424m
Hrp1	pGexTev	494-534	R524A	425g
Hrp1	pMalTev	494-534	R524A	425m
Hrp1	pGexTev	494-534	R525A	426g
Hrp1	pMalTev	494-534	R525A	426m
Hrp1	pGexTev	494-534	Y529A	427g
Hrp1	pMalTev	494-534	Y529A	427m
Hrp1	pGexTev	494-534	P531A	428g
Hrp1	pMalTev	494-534	P531A	428m
Hrp1	pGexTev	494-534	Y532A	429g
Hrp1	pMalTev	494-534	Y532A	429m
Hrp1	pGexTev	494-534	P531A, Y532A	430g
Hrp1	pMalTev	494-534	P531A, Y532A	430m
Hrp1	pGexTev	494-534	R524A, P531A, Y532A	431g
Hrp1	pMalTev	494-534	R524A, P531A, Y532A	431m
Hrp1	pGexTev	494-534	R525A, P531A, Y532A	432g
Hrp1	pMalTev	494-534	R525A, P531A, Y532A	432m
Hrp1	pGexTev	494-534	R516A	437g
Hrp1	pMalTev	494-534	R516A	437m
Hrp1	pGexTev	494-534	R519A	438g
Hrp1	pMalTev	494-534	R519A	438m
Hrp1	pGexTev	494-534	Y522A	439g

Gene	Vector	Fragment	Mutations	KS#
Hrp1	pMalTev	494-534	Y522A	439m
Hrp1	pGexTev	494-534	Y532L	441g
Hrp1	pMalTev	494-534	Y532L	441m
Hrp1	pMalTev	494-534	H530A	447m
Hrp1	pMalTev	494-534	N526A, N527A	448m
Hrp1	pMalTev	494-534	R512A, R513A, N514A, R516A, R519A, G520A, G521A, Y522A, N523A	467m
Hrp1	pMalTev	494-534	R512A, R513A, R516A, R519A	539m
Hrp1	pMalTev	494-534	R512K, R513K, R516K, R519K	540m
Hrp1	pGexTev	494-534	Y532I	541g
Hrp1	pMalTev	494-534	Y532I	541m
Hrp1	pGexTev	494-534	Y532F	542g
Hrp1	pMalTev	494-534	Y532F	542m
Hrp1	pGexTev	494-534	Y532M	543g
Hrp1	pMalTev	494-534	Y532M	543m
Hrp1	pMalTev	494-534	Y529L	563m
Hrp1	pMalTev	494-534	R506A	573m
Hrp1	pMalTev	494-534	R512A	574m
Hrp1	pMalTev	494-534	R513A	575m
Hrp1	pMalTev	494-534	R524A, R525A	576m
Hrp1	pMalTev	494-534	Y532S	577m
Hrp1	pMalTev	494-534	Y532T	578m
Hrp1	pMalTev	494-534	Y532V	579m
Hrp1	pMalTev	494-534	Y532H	580m
Hrp1	pMalTev	494-534	Y532C	581m
Hrp1	pMalTev	494-534	Y532D	582m
Hrp1	pMalTev	494-534	Y532E	583m
Hrp1	pMalTev	494-534	Y532N	584m
Hrp1	pMalTev	494-534	Y532Q	585m
Hrp1	pMalTev	494-534	Y532K	586m
Hrp1	pMalTev	494-534	Y532R	587m
Hrp1	pMalTev	494-534	Y532W	588m
Hrp1	pMalTev	494-534	Y532P	589m
Hrp1	pMalTev	494-534	P531A, R516A	428/437
Hrp1	pMalTev	494-534	P531A, R524A, R525A	428/576
Hrp1	pMalTev	494-534	N510A, H511A	602m
Hrp1	pMalTev	494-534	R512A, P531A	428/574

Gene	Vector	Fragment	Mutations	KS#
Hrp1	pMalTev	494-534	Y532G	590m
Hrp1	pMalTev	494-534	R512A, R524A, R525A	576/574
Hrp1	pMalTev	494-534	R512A, Y529A	574/427

Table 7-3. Nab2 plasmids.

Gene	Vector	Fragment	Mutations	KS#
Nab2	pGexTev	full length	wild type	1
Nab2	pMalTev	full length	wild type	1
Nab2	pGexTev	181-241	wild type	2
Nab2	pMalTev	181-241	wild type	2
Nab2	pGexTev	181-251	wild type	3
Nab2	pMalTev	181-251	wild type	3
Nab2	pGexTev	201-241	wild type	4
Nab2	pMalTev	201-241	wild type	4
Nab2	pGexTev	201-251	wild type	5
Nab2	pMalTev	201-251	wild type	5
Nab2	pGexTev	181-201	wild type	9
Nab2	pMalTev	181-201	wild type	9
Nab2	pGexTev	201-251	G210A, G211A, G212A	410g
Nab2	pMalTev	201-251	G210A, G211A, G212A	410m
Nab2	pGexTev	201-251	K216A, N217A, R218A, R219A	411g
Nab2	pMalTev	201-251	K216A, N217A, R218A, R219A	411m
Nab2	pGexTev	201-251	R222A	412g
Nab2	pMalTev	201-251	R222A	412m
Nab2	pGexTev	201-251	R222A, G223A, G224A	413g
Nab2	pMalTev	201-251	R222A, G223A, G224A	413m
Nab2	pGexTev	201-251	R226A, G227A, G228A, R229A, N230A	414g
Nab2	pMalTev	201-251	R226A, G227A, G228A, R229A, N230A	414m
Nab2	pGexTev	201-251	R235A	415g
Nab2	pMalTev	201-251	R235A	415m
Nab2	pGexTev	201-251	F236A	416g
Nab2	pMalTev	201-251	F236A	416m
Nab2	pGexTev	201-251	P238A	417g
Nab2	pMalTev	201-251	P238A	417m
Nab2	pGexTev	201-251	L239A	418g
Nab2	pMalTev	201-251	L239A	418m
Nab2	pGexTev	201-251	P238A, L239A	419g

Gene	Vector	Fragment	Mutations	KS#
Nab2	pMalTev	201-251	P238A, L239A	419m
Nab2	pGexTev	201-251	R235A, P238A, L239A	420g
Nab2	pMalTev	201-251	R235A, P238A, L239A	420m
Nab2	pGexTev	201-251	R226A	435g
Nab2	pMalTev	201-251	R226A	435m
Nab2	pGexTev	201-251	R229A	436g
Nab2	pMalTev	201-251	R229A	436m
Nab2	pGexTev	201-251	L239Y	440g
Nab2	pMalTev	201-251	L239Y	440m
Nab2	pGexTev	201-251	F236A, P238A, L239A	442g
Nab2	pMalTev	201-251	F236A, P238A, L239A	442m
Nab2	pGexTev	201-251	K216A, N217A, R218A, R219A, R222A, R226A, G227A, G228A, R229A, N230A	443g
Nab2	pMalTev	201-251	K216A, N217A, R218A, R219A, R222A, R226A, G227A, G228A, R229A, N230A	443m
Nab2	pMalTev	201-251	N327A	444m
Nab2	pMalTev	201-251	N230A, N231A, N232A	445m
Nab2	pMalTev	201-251	R209A, G210A, G211A, G212A	446m
Nab2	pMalTev	201-251	L239H	514m
Nab2	pMalTev	201-251	L239I	515m
Nab2	pMalTev	201-251	L239M	516m
Nab2	pMalTev	201-251	L239F	517m
Nab2	pMalTev	201-251	R235K	518m
Nab2	pMalTev	201-251	R235H	519m
Nab2	pMalTev	201-251	R218A, R219A, R222A, R226A, R229A	537m
Nab2	pMalTev	201-251	R218K, R219K, R222K, R226K, R229K	538m
Nab2	pMalTev	201-251	L239V	551m
Nab2	pMalTev	201-251	F236L	562m
Nab2	pMalTev	201-251	K216A	570m
Nab2	pMalTev	201-251	R218A	571m
Nab2	pMalTev	201-251	R219A	572m
Nab2	pMalTev	201-251	P238A, L239A, R218A, R219A, R222A, R226A, R229A	419/537m
Nab2	pMalTev	201-251	P238A, L239A, R222A	419/412
Nab2	pMalTev	201-251	S233A, T234A	600m
Nab2	pMalTev	201-251	V214A	601m
Nab2	pMalTev	201-251	R222A, R235A	412/415

Table 7-4: Shuttle plasmids for yeast transformations (except Tfg2p).

Gene	Vector	Frag-ment	Mutations	Notes	KS#	-
Vectors						
vector only	pRS415-ADH1 promoter-gene-GFP			MCS: XbaI, SpeI, BamH1, SmaI, PstI		leu
vector only	pRS415-gal promoter-gene-GFP			MCS: XbaI, SpeI, BamH1, SmaI, PstI		leu
vector only	pRS415-ADH1-gfp-gfp-gene			MCS: BamH1, NotI GGSGG bt GFPs		leu
vector only	pRS415-ADH1-gfp-gst-gene			MCS: BamH1, NotI GGSGG bt GFP and GST		leu
vector only	pRS415-ADH1 promoter-TAP tag-gene			MCS: BamHI, SmaI, PstI, HindIII, SalI		leu
vector only	pRS415-ADH1 promoter-GFP-gene			MCS: SpeI, BamHI, SmaI, PstI, HindIII, SalI, XhoI		
Full length wild type genes with GFP C-terminal fusion (ADH1 constitutive promoter)						
Nab2	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	501a	leu
Hrp1	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	502a	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	449a	leu
Rml2	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	462a	leu
Gbp2	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	499a	leu
Enp1	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	466a-2	leu
Snp1	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	472a-1	leu
Sgv1	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	514a-1	leu
Aft1	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	554a-1	leu
Pol12	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	555a-1	leu
Naf1	pRS415-ADH1 promoter-gene-GFP	full length		SpeI/SmaI	556a-1	leu
Full length wild type genes with GFP C-terminal fusion (Gal 1 inducible promoter)						
Nab2	pRS415-gal promoter-gene-GFP	full length		SpeI/SmaI	y501g-a	leu
Hrp1	pRS415-gal promoter-gene-GFP	full length		SpeI/SmaI	y502g-a	leu
Tfg2	pRS415-gal promoter-gene-GFP	full length		SpeI/SmaI	y449g-f	leu
Rml2	pRS415-gal promoter-gene-GFP	full length		SpeI/SmaI	y462g-b	leu
Gbp2	pRS415-gal promoter-gene-GFP	full length		SpeI/SmaI	y499g-b	leu
Enp1	pRS415-gal promoter-gene-GFP	full length		SpeI/SmaI	y466g-a	leu
Snp1	pRS415-gal promoter-	full		SpeI/SmaI	y472g-b	leu

Gene	Vector	Frag-ment	Mutations	Notes	KS#	-
	gene-GFP	length				
Aft1	pRS415-gal promoter-gene-GFP	full length		SpeI/SmaI	y553g-d	leu
Naf1	pRS415-gal promoter-gene-GFP	full length		SpeI/SmaI	y556g-b	leu
Full length point mutants of Hrp1 and Nab2 with GFP C-terminal fusion (ADH1 constitutive promoter)						
Hrp1	pRS415-ADH1 promoter-gene-GFP	full length	R512A, R513A, R516A, R519A	SpeI/SmaI	539i	leu
Hrp1	pRS415-ADH1 promoter-gene-GFP	full length	R524A, R525A	SpeI/SmaI	576c	leu
Hrp1	pRS415-ADH1 promoter-gene-GFP	full length	Y529A	SpeI/SmaI	427e	leu
Hrp1	pRS415-ADH1 promoter-gene-GFP	full length	P531A, Y532A	SpeI/SmaI	430e	leu
Hrp1	pRS415-ADH1 promoter-gene-GFP	full length	P531A, Y532A, R512A, R513A, R516A, R519A	SpeI/SmaI	430/539b	leu
Nab2	pRS415-ADH1 promoter-gene-GFP	full length	R218A, R219A, R222A, R226A, R229A	SpeI/SmaI	537	leu
Nab2	pRS415-ADH1 promoter-gene-GFP	full length	R235A	SpeI/SmaI	415e	leu
Nab2	pRS415-ADH1 promoter-gene-GFP	full length	F236A	SpeI/SmaI	416e	leu
Nab2	pRS415-ADH1 promoter-gene-GFP	full length	P238A, L239A	SpeI/SmaI	419e	leu
Nab2	pRS415-ADH1 promoter-gene-GFP	full length	R235A, P238A, L239A	SpeI/SmaI	420	leu
Nab2	pRS415-ADH1 promoter-gene-GFP	full length	P238A, L239AR218A, R219A, R222A, R226A, R229A	SpeI/SmaI	419/537b	leu
NLSs of Hrp1 and Nab2, wild type and with point mutations with GFP-GST N-terminal (ADH1 constitutive promoter)						
Hrp1	pRS415-ADH1-gfp-gst-gene	494-534	wild type	BamH1/Not1		leu
Hrp1	pRS415-ADH1-gfp-gst-gene	494-534	R512A, R513A, R516A, R519A	BamH1/Not1	539g	leu
Hrp1	pRS415-ADH1-gfp-gst-gene	494-534	R524A, R525A	BamH1/Not1	576g	leu
Hrp1	pRS415-ADH1-gfp-gst-gene	494-534	Y529A	BamH1/Not1	427g	leu
Hrp1	pRS415-ADH1-gfp-gst-gene	494-534	P531A, Y532A	BamH1/Not1	430g	leu
Hrp1	pRS415-ADH1-gfp-gst-gene	494-534	P531A, Y532A, R512A, R513A, R516A, R519A	BamH1/Not1	430/539g	leu
Nab2	pRS415-ADH1-gfp-gst-gene	201-251	wild type	BamH1/Not1		leu
Nab2	pRS415-ADH1-gfp-gst-gene	201-251	R218A, R219A, R222A, R226A, R229A	BamH1/Not1	537g	leu

Gene	Vector	Frag-ment	Mutations	Notes	KS#	-
Nab2	pRS415-ADH1-gfp-gst-gene	201-251	R235A	BamH1/Not1	415g	leu
Nab2	pRS415-ADH1-gfp-gst-gene	201-251	F236A	BamH1/Not1	416g	leu
Nab2	pRS415-ADH1-gfp-gst-gene	201-251	P238A, L239A	BamH1/Not1	419g	leu
Nab2	pRS415-ADH1-gfp-gst-gene	201-251	R235A, P238A, L239A	BamH1/Not1	420g	leu
Nab2	pRS415-ADH1-gfp-gst-gene	201-251	P238A, L239A, R218A, R219A, R222A, R226A, R229A	BamH1/Not1	419/537g	leu
NLSs of putative Kap104 substates with GFP-GST N-terminal (ADH1 constitutive reporter)						
Rml2	pRS415-ADH1-gfp-gst-gene	74-123		BamH1/Not1	486	leu
Snp1	pRS415-ADH1-gfp-gst-gene	1-42		BamH1/Not1	472nls	leu
Tfg2	pRS415-ADH1-gfp-gst-gene	165-214		BamH1/Not1	449nls	leu
Naf1	pRS415-ADH1-gfp-gst-gene	269-318		BamH1/Not1	523nls	leu
Prb1	pRS415-ADH1-gfp-gst-gene	125-174		BamH1/Not1	524nls	leu
Enp1	pRS415-ADH1-gfp-gst-gene	1-27		BamH1/Not1	466nls	leu
Cdc25p	pRS415-ADH1-gfp-gst-gene	774-823		BamH1/Not1	544nls	leu
Dig1p	pRS415-ADH1-gfp-gst-gene	66-115		BamH1/Not1	452nls	leu
Arp8p	pRS415-ADH1-gfp-gst-gene	114-163		BamH1/Not1	450nls	leu
Ste20	pRS415-ADH1-gfp-gst-gene	226-275		BamH1/Not1	525nls	leu
Aft1	pRS415-ADH1-gfp-gst-gene	178-227		BamH1/Not1	536nls	leu
Rps1A	pRS415-ADH1-gfp-gst-gene	1-33		BamH1/Not1	530nls	leu
Haa1	pRS415-ADH1-gfp-gst-gene	89-138		BamH1/Not1	483nls	leu
Gbp2	pRS415-ADH1-gfp-gst-gene	64-114		BamH1/Not1	499nls	leu
Mmr1	pRS415-ADH1-gfp-gst-gene	46-95		BamH1/Not1	528nls	leu
Hrk1	pRS415-ADH1-gfp-gst-gene	659-708		BamH1/Not1	497nls	leu
Apc2	pRS415-ADH1-gfp-gst-gene	415-464		BamH1/Not1	496nls	leu
N-terminal TAP-tag						
Kap104	pRS415-ADH1 promoter-TAP tag-gene	full length		BamHI/PstI		leu

Gene	Vector	Frag-ment	Mutations	Notes	KS#	-
Full length point mutants of Hrp1 with GFP N-terminal fusion (ADH1 constitutive promoter)						
Hrp1	pRS415-ADH1 promoter-GFP-gene	full length	wild type	SpeI/XhoI		leu
Hrp1	pRS415-ADH1 promoter-GFP-gene	full length	R524A, R525A	SpeI/XhoI	576	leu
Hrp1	pRS415-ADH1 promoter-GFP-gene	full length	Y529A	SpeI/XhoI	457	leu
Hrp1	pRS415-ADH1 promoter-GFP-gene	full length	P531A, Y532A	SpeI/XhoI	430	leu
Hrp1	pRS415-ADH1 promoter-GFP-gene	full length	P531A, Y532A, R512A, R513A, R516A, R519A	SpeI/XhoI	430/539	leu

Table 7-5: Tfg2p plasmids

Gene	Vector	Frag-ment	Mutations	Notes	KS#	-
Full length and fragments of Tfg2, wild type						
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	none	SpeI/SmaI	449a	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	165-214	none	BamH1/Not1	449nls	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	341-374	none	BamH1/Not1	628	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	1-86	none	BamH1/Not1	640	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	87-165	none	BamH1/Not1	641	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	208-262	none	BamH1/Not1	642	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	263-343	none	BamH1/Not1	643	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	344-400	none	BamH1/Not1	644	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	87-400	none	Spe1/Pst1	648	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	1-214	none	BamH1/Not1		leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	165-400	none	BamH1/Not1		leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	Δ66-85	none	BamH1/Not1	664	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	1-165	none	BamH1/Not1		leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	208-400	none	BamH1/Not1		leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	263-400	none	BamH1/Not1		leu

Gene	Vector	Frag-ment	Mutations	Notes	KS#	-
Tfg2	pGexTev	full length	none	BamH1/Not1	449	
Tfg2	pGexTev	165-214	none	BamH1/Not1	449nls	
Tfg2	pGexTev	87-400	none	BamH1/Not1	641	
Tfg2	pGexTev	1-214	none	BamH1/Not1		
Tfg2	pGexTev	165-400	none	BamH1/Not1		
Tfg2	pGexTev	Δ66-85	none	BamH1/Not1	664	
Tfg2	pRS314-endo. promoter-gene-myc	1-343	none		681 wt	trp
Fragment point mutants of GFP-GST-Tfg2: NLS						
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	165-214	K179A, R180A, R181A	BamH1/Not1	624	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	165-214	R184A, K185A, K186A	BamH1/Not1	625	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	165-214	Y201A	BamH1/Not1	626	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	165-214	P203A, Y204A	BamH1/Not1	627	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	87-400	P203A, Y204A	Spe1/Pst1	-	leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	1-214	P203A, Y204A	BamH1/Not1		leu
Tfg2	pRS415-ADH1 promoter-GFP-ggsgg-GST-gene	165-400	P203A, Y204A	BamH1/Not1		leu
Tfg2	pGexTev	165-214	K179A, R180A, R181A	BamH1/Not1	624	
Tfg2	pGexTev	165-214	R184A, K185A, K186A	BamH1/Not1	625	
Tfg2	pGexTev	165-214	Y201A	BamH1/Not1	626	
Tfg2	pGexTev	165-214	P203A, Y204A	BamH1/Not1	627	
Tfg2	pGexTev	87-400	P203A, Y204A	BamH1/Not1	-	
Tfg2	pGexTev	1-214	P203A, Y204A	BamH1/Not1		
Tfg2	pGexTev	165-400	P203A, Y204A	BamH1/Not1		
Tfg2	pRS314-endo. promoter-gene-myc	1-343	P203A, Y204A		681 PY	trp
Full length point mutants of Tfg2-GFP: NLS						
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	K179A, R180A, R181A	SpeI/SmaI	624	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	R184A, K185A, K186A	SpeI/SmaI	625	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	Y201A	SpeI/SmaI	626	leu
Tfg2	pRS415-ADH1 promoter-	full	P203A, Y204A	SpeI/SmaI	627	leu

Gene	Vector	Frag-ment	Mutations	Notes	KS#	-
	gene-GFP	length				
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	K356A, K357A, K359A	SpeI/SmaI	629	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	P203A, Y204A, K356A, K357A, K359A	SpeI/SmaI	627/629	leu
Full length point mutants of Tfg2-GFP: DNA binding domain						
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	R293A, K296A, K297A	SpeI/SmaI	672wt	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	R293A, K296A, K297A, P203A, Y204A		672PY	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	R293D, K296D		673wt	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	R293D, K296D, P203A, Y204A		673PY	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	S314A, K316A, K319A, R323A		674wt	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	S314A, K316A, K319A, R323A, P203A, Y204A		674PY	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	K335A, K341A, K342A		676wt	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	K335A, K341A, K342A, P203A, Y204A		676PY	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	K341D, K342D		677wt	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	K341D, K342D, P203A, Y204A		677PY	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	P344A, Y345A		678wt	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	P344A, Y345A, P203A, Y204A		678PY	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	E331A, D334A, K335A		679wt	leu
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	E331A, D334A, K335A, P203A, Y204A		679PY	leu
Tfg2	pRS314-endo. promoter-gene-myc	full length	S314A, K316A, K319A, R323A		674wt	trp
Tfg2	pRS314-endo. promoter-gene-myc	full length	S314A, K316A, K319A, R323A, P203A, Y204A		674PY	trp
Tfg2	pRS314-endo. promoter-gene-myc	full length	K335A, K341A, K342A		676wt	trp
Tfg2	pRS314-endo. promoter-gene-myc	full length	K335A, K341A, K342A, P203A, Y204A		676PY	trp
Tfg2	pRS314-endo. promoter-gene-myc	full length	E331A, D334A, K335A		679wt	trp
Tfg2	pRS314-endo. promoter-gene-myc	full length	E331A, D334A, K335A, P203A, Y204A		679PY	trp
Tfg2	pRS415-ADH1 promoter-gene-GFP	full length	Δ172-206	Ggsgg linker	ΔNLS	leu
Tfg2	pRS314-endo. promoter-gene-myc	full length	Δ172-206	Ggsgg linker	ΔNLS	trp

Table 7-6:Plasmids: putative Kap104p substrates

Gene	Vector	Coding length	5' enzyme	KS#
AFR1 bPH	pGexTev	150	BamH1	503
AFT1 bPF	pGexTev	150	BamH1	536nls
AFT1 bPF	pGexTev	150	BamH1	536nls
AFT1 bPF	pMalTev	150	BamH1	536nls
AIR2 fl	pGexTev	1035	Sal1	461
AOS1 fl	pGexTev	1044	Sal1	456
APC2 bPI	pGexTev	150	BamH1	496nls
APC2 bPI	pMalTev	150	BamH1	496nls
APC2 fl	pGexTev	2562	Sal1	496
ARP8 bPL	pGexTev	150	BamH1	450nls
ARP8 bPL	pGexTev	150	BamH1	450nls
ARP8 bPL	pMalTev	150	BamH1	450nls
ARP8 fl	pGexTev	2646	BamH1	450
ATG13 fl	pGexTev	2217	Sal1	473
BBP1 fl	pGexTev	1158	BamH1	468
BBP1 hPY	pGexTev	150	BamH1	488
BMS1 bPM	pGexTev	150	BamH1	507
BMS1 bPM	pMalTev	150	BamH1	507
CDC16 bPM	pGexTev	150	BamH1	476nls
CDC16 fl	pGexTev	2523	Sal1	476
CDC21 hPL	pGexTev	150	BamH1	491
CDC25 bPL	pMalTev	150	BamH1	544nls
CDC5 bPI	pGexTev	150	BamH1	498nls
CDC5 bPI	pMalTev	150	BamH1	498nls
CDC5 fl	pGexTev	2118	BamH1	498
CEG1 bPF	pGexTev	150	BamH1	480nls
CEG1 fl	pGexTev	1380	Sal1	480
CLG1 fl	pGexTev	1359	BamH1	469
CLG1 hPY	pGexTev	150	BamH1	489
CTF8 bPM	pGexTev	150	BamH1	508
CWC25 bPL	pGexTev	150	BamH1	451nls
CWC25 bPL	pGexTev	150	BamH1	451nls
CWC25 bPL	pMalTev	150	BamH1	451nls
CWC25 bPM	pGexTev	150	BamH1	534nls

Gene	Vector	Coding length	5' enzyme	KS#
CWC25 bPM	pGexTev	150	BamH1	534nls
CWC25 fl	pGexTev	540	BamH1	451
DIG1 bPL	pGexTev	150	BamH1	452nls
DIG1 bPL	pGexTev	150	BamH1	452nls
DIG1 bPL	pMalTev	150	BamH1	452nls
DIG1 fl	pGexTev	1359	BamH1	452
ENP1 bPL	pGexTev	150	BamH1	466nls
ENP1 bPL	pGexTev	150	BamH1	466nls
ENP1 bPL	pMalTev	150	BamH1	466nls
ENP1 fl	pGexTev	1452	BamH1	466
ESP1 bPL	pGexTev	150	BamH1	522nls
ESP1 bPL	pGexTev	150	BamH1	522nls
ESP1 bPL	pMalTev	150	BamH1	522nls
FOB1 bPM	pGexTev	150	BamH1	474nls
FOB1 fl	pGexTev	1701	BamH1	474
GBP2 bPI	pGexTev	150	BamH1	499nls
GBP2 bPI	pMalTev	150	BamH1	499nls
GBP2 fl	pGexTev	1284	BamH1	499
GCN5 bPM	pGexTev	150	BamH1	509
GCN5 bPM	pMalTev	150	BamH1	509
HAA1 bPL 1	pGexTev	150	BamH1	484
HAA1 bPL 2	pGexTev	150	BamH1	485
HAA1 bPM	pGexTev	150	BamH1	483
HAA1 bPM	pMalTev	150	BamH1	483
HAA1 fl	pGexTev	2085	BamH1	482
HRK1 bPI	pGexTev	150	BamH1	497nls
HRK1 PY/AA	pGexTev	QC		520
ICL1 bPI	pGexTev	150	BamH1	527nls
ICL1 bPI	pGexTev	150	BamH1	527nls
ICL1 bPI	pMalTev	150	BamH1	527nls
IME4 fl	pGexTev	1803	Sal1	470
IPP1 bPH	pGexTev	150	BamH1	477nls
IPP1 fl	pGexTev	864	BamH1	477
LRS4 bPI	pGexTev	150	BamH1	510
MIF2 bPL	pMalTev	150	BamH1	545nls
MMR1 bPI	pGexTev	150	BamH1	528nls
MMR1 bPI	pGexTev	150	BamH1	528nls

Gene	Vector	Coding length	5' enzyme	KS#
MMR1 bPI	pMalTev	150	BamH1	528nls
MRPS8 bPM	pGexTev	150	BamH1	535nls
MRPS8 bPM	pGexTev	150	BamH1	535nls
NAF1 bPL	pGexTev	150	BamH1	523nls
NAF1 bPL	pGexTev	150	BamH1	523nls
NAF1 bPL	pMalTev	150	BamH1	523nls
NAM8 fl	pGexTev	1572	Sal1	471
NAM8 hPY	pGexTev	150	BamH1	490
NDT80 bPF	pGexTev	150	BamH1	504
NDT80 bPF	pMalTev	150	BamH1	504
NFI1 fl	pGexTev	2181	Sal1	454
PGM2 fl	pGexTev	1710	BamH1	465
POL12 bPF	pGexTev	150	BamH1	505
POS5 hPL	pGexTev	150	BamH1	492
PRB1 bPL	pGexTev	150	BamH1	524nls
PRB1 bPL	pGexTev	150	BamH1	524nls
PRB1 bPL	pMalTev	150	BamH1	524nls
PSP2 fl	pGexTev	2144	BamH1	481
RGD1 bPF	pGexTev	150	BamH1	479nls
RGD1 fl	pGexTev	2001	Sal1	479
RML2 bPH	pGexTev	150	BamH1	487
RML2 bPY	pGexTev	150	BamH1	486
RML2 bPY	pMalTev	150	BamH1	486
RML2 fl	pGexTev	1182	BamH1	462
RPA135 bPL	pMalTev	150	BamH1	546nls
RPO41 bPY	pGexTev	150	BamH1	463nls
RPO41 bPY	pGexTev	150	BamH1	463nls
RPO41 bPY	pMalTev	150	BamH1	463nls
RPO41 fl	pGexTev	4056	Sal1	463
RPS1A bPF	pGexTev	<150	BamH1	530nls
RPS1A bPF	pGexTev	<150	BamH1	530nls
RPS1A bPF	pMalTev	<150	BamH1	530nls
RSC2 fl	pGexTev	2670	Sal1	457
SBP1 bPL	pMalTev	150	BamH1	547nls
SGD1 BPI	pGexTev	150	BamH1	529nls
SGD1 BPI	pGexTev	150	BamH1	529nls
SGV1 bPI	pGexTev	150	BamH1	511

Gene	Vector	Coding length	5' enzyme	KS#
SGV1 bPI	pMalTev	150	BamH1	511
SGV1 bPY	pGexTev	150	BamH1	513
SGV1 bPY	pMalTev	150	BamH1	513
SHE1 bPY	pMalTev	150	BamH1	550nls
SIN3 hPL	pGexTev	150	BamH1	493
SIP1 bPF	pGexTev	150	BamH1	532nls
SIP1 bPF	pGexTev	150	BamH1	532nls
SIZ1 bPH	pGexTev	150	BamH1	478nls
SIZ1 fl	pGexTev	2715	Sal1	478
Sko1 hPL	pGexTev	150	BamH1	494
SNP1 bPY	pGexTev	150	BamH1	472nls
SNP1 bPY	pGexTev	150	BamH1	472nls
SNP1 bPY	pMalTev	150	BamH1	472nls
SNP1 fl	pGexTev	903	BamH1	472
SOK1 fl	pGexTev	2706	BamH1	458
SPO13 bPI	pGexTev	150	BamH1	512
SPO13 bPI	pMalTev	150	BamH1	512
STE20 bPL	pGexTev	150	BamH1	525nls
STE20 bPL	pGexTev	150	BamH1	525nls
STE20 bPL	pMalTev	150	BamH1	525nls
TAF2 fl	pGexTev	4224	Sal1	459
TFB1 bPM	pGexTev	150	BamH1	475nls
TFB1 fl	pGexTev	1928	BamH1	475
TFC3 bPL	pMalTev	150	BamH1	548nls
TFG2 bPY	pGexTev	150	BamH1	449nls
TFG2 bPY	pGexTev	150	BamH1	449nls
TFG2 bPY	pMalTev	150	BamH1	449nls
TFG2 fl	pGexTev	1203	BamH1	449
TOF2 fl	pGexTev	2316	BamH1	464
TRM2 fl	pGexTev	1920	Sal1	460
TRM8 BPF	pGexTev	150	BamH1	506
TRM8 BPF	pMalTev	150	BamH1	506
TRP2 hPL	pGexTev	150	BamH1	495
VPS13 bPY	pMalTev	150	BamH1	549nls
YHI9 bPF	pGexTev	150	BamH1	533nls
YHI9 bPF	pGexTev	150	BamH1	533nls
YHI9 bPF	pMalTev	150	BamH1	533nls

Table 7-7: Xpo1 yeast plasmids

Gene	Vector	Fragment	Mutations	Notes	marker
vector	pRS415-ADH1-GFP-GFP-gene			BamH1/Not1 3' of last GFP. Ggsgg linker between every gene	leu
vector	pRS414-ADH1-GFP-GFP-gene			BamH1/Not1 3' of last GFP. Ggsgg linker between every gene	trp
vector	pRS415-ADH1-SV40nls-GFP-GFP-gene			BamH1/Not1 3' of last GFP. Ggsgg linker between every gene	leu
vector	pRS414-ADH1-SV40nls-GFP-GFP-gene			BamH1/Not1 3' of last GFP. Ggsgg linker between every gene	trp
Nmd3	pRS415-ADH1-SV40nls-GFP-GFP-gene	477-493	none	BamH1/Not1	leu
Nmd3	pRS414-ADH1-SV40nls-GFP-GFP-gene	477-493	none	BamH1/Not1	trp
Rev	pRS415-ADH1-SV40nls-GFP-GFP-gene	71-85	none	BamH1/Not1	leu
Rev	pRS414-ADH1-SV40nls-GFP-GFP-gene	71-85	none	BamH1/Not1	trp
Snp1	pRS415-ADH1-SV40nls-GFP-GFP-gene	1-15	none	BamH1/Not1	leu
Snp1	pRS414-ADH1-SV40nls-GFP-GFP-gene	1-15	none	BamH1/Not1	trp
Snp1	pRS415-ADH1-SV40nls-GFP-GFP-gene	1-15	L4A, L8A	BamH1/Not1	leu
Snp1	pRS415-ADH1-SV40nls-GFP-GFP-gene	1-341	none	BamH1/Not1	leu
Snp1	pRS415-ADH1-SV40nls-GFP-GFP-gene	1-341	L4A, L8A	BamH1/Not1	leu
Xpo1	pRS414-ADH1-gene	full length	none		leu
Xpo1	pRS414-gene	full length	none	has 0.5 kb upstream	leu

Tables of predicted Kapβ2 substrates

Table 7-8: Predicted Kapβ2 substrates with hydrophobic PY-NLSs

Accession number	Name	Localization[a]	N-Term. Residue	Sequences for Candidate Hydrophobic PY-NLS[b]	C-Term. Residue
Q8IZP0	Abl interactor 1	C, N	158	KHGNNQPARTGTLSRTNPPTQKPPSPP MSGRGTLG**R**NT**PY**KTLEPVKPPT	207
Q9UKA4	A-kinase anchor protein 11/AKAP 220	C, Centrosome	385	QRKGHKHGKSCMNPQKFKFDRPALPA NVRKPTP**R**KPES**PY**GNLCDAPDSP	434
P50995	Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50)	C, N	84	PVPPGGFGQPPSAQQPVPPYGMYPP PGGNPPS**R**MPSYP**PY**PGAPVPGQPM	133
Q13625	Apoptosis-stimulating of p53 protein 2	C, N	474	TLRKNQSSEDILRDAQVANKNVAKVP PPVPTKP**K**QINL**PY**FGQTNQPPSD	523
Q9BXP5	Arsenite-resistance protein 2[c]	not known	53	GEYRDYDRNRRERFSPPRHELSPP QKRMRRDWDE**H**SSD**PY**HSGYEMPYAG	102
Q92560	Ubiquitin carboxyl-terminal hydrolase BAP1(BRCA1-associated protein 1)[c]	N	685	EGMLANLVEQNISVRRRQGVSIGRL HKQRKPDRRK**R**SR**PY**KAKRQ	729
P48634	Large proline-rich protein BAT2 (HLA-B-associated transcript 2)	C, N	690	VPAPQAPPPPPKALYPGALGRPPPM PPMNFDP**R**WMMIP**PY**VDPRLLQGRP	739
O15178	Brachyury protein	N	251	TSTLCPPANPHPQFGGALSLPSTHS CDRYPTLRSH**R**SS**PY**PSPYAHRNNS	300
O60885	Bromodomain-containing protein 4 (HUNK1 protein)	N	1015	QGQQPPHPPPGQQPPPPQPAKPQQV IQHHHSP**R**HHKSD**PY**STGHLREAPSP	1064
Q14004	Cell division cycle 2-like protein kinase 5	not known	376	YERGGDVSPSPYSSSSWRRSRSPYSPVLRRSGKS**R**SRS**PY**SSRHSRSRSR	425
Q9NYV4	Cell division cycle 2-related protein kinase 7	N	256	SSNYDSYKKSPGSTSRRQSVSPPYK EPSAYQSST**R**SPS**PY**SRRQRSVSPY	305
Q5TGI0	Protein C6orf168	not known	94	IDSKDAIILHQFARPNNGVPSLSPF CLKMETYL**R**MADL**PY**QNYFGGKLSA	143
P49761	Dual specificity protein kinase CLK3 (CDC-like kinase 3/Clk3)[c]	N	18	YRWKRRRSYSREHEGRLRYPSRREP PPRRSRS**R**SHDRL**PY**QRRYRERRDS	67
P05997	Collagen alpha-2(V) chain precursor	not known	611	MGLPGPKGSNGDPGKPGEAGNPGVP GQRGAPGKDG**K**VG**PY**GPPGPPGLRG	660
Q03692	Collagen alpha-1(X) chain precursor	not known	84	GYGSPGLQGEPGLPGPPGPSAVGKP GVPGLPG**K**PGERG**PY**GPKGDVGPAG	133
Q8TBR5	Protein C19orf23[c]	not known	70	TWQTRNHTRTGHAYPRFTRPSFPSC NRNGKRRKL**R**LGL**PY**	119
Q96RT6	Protein cTAGE-2	not known	692	PPGTVFGASPDYFSPRDVPGPPRAP FAMRNVYLP**R**GFL**PY**RPPRPAFFPQ	741
Q9NSV4	Protein diaphanous homolog 3 (Diaphanous-related formin-3)	not known	1070	GAAFRDRRKRTPMPKDVRQSLSPM SQRPVLKVCN**H**GNK**PY**L	1110
P56177	Homeobox protein DLX-1	N	44	CLHSAGHSQPDGAYSSASSFSRPLG YPYVNSVSS**H**ASS**PY**ISSVQSYPGS	93
O95147	Dual specificity protein phosphatase 14/MAP kinase phosphatase 6	not known	156	RQLIDYERQLFGKSTVKMVQTPYGIV PDVYEKES**R**HLM**PY**WGI	200
Q9BUP0	EF-hand domain-containing protein 1 (Swiprosin-2)	not known	42	PPARAPTASADAELSAQLSRRLDINE GAARPRRC**R**VFN**PY**TEFPEFSRRL	91

Table 7-8 continued: Predicted Kapβ2 substrates with hydrophobic PY-NLSs

Q6ZV73	FYVE, RhoGEF and PH domain-containing protein 6 (Zinc finger FYVE domain-containing protein 24)	C	269	SSELEALENGKRSTLISSDGVSKKSE VKDLGPLEI**H**LV**PY**TPKFPTPKPR	318
Q92837	Proto-oncogene FRAT1	N	89	PAVPLLLPPALAETVGPAPPGVLRCA LGDRGRVRG**R**AA**PY**CVAELATGPS	138
Q96AE4	FUSE-binding protein 1/DNA helicase V	N	465	PGPHGPPGPPGPGTPMGPYNPAPYNP GPPGPAP**H**GPPA**PY**APQGWGNAYP	514
Q8NEA6	Zinc finger protein GLIS3	N	601	LTAVDAGAERFAPSAPSPHHISPRRV PAPSSILQ**R**TQP**PY**TQQPSGSHLK	650
Q8TEK3	Histone H3-K79 methyltransferase	N	775	SPAKIVLRRHLSQDHTVPGRPAASEL HSRAEHT**K**ENGL**PY**QSPSVPGSMK	824
P35452	Homeobox protein Hox-D12 (Hox-4H)	N	175	AGVASCLRPSLPDGKRCPCSPGRPAVG GGPGEA**R**KKRK**PY**TKQQIAELEN	224
Q13422	DNA-binding protein Ikaros (Lymphoid transcription factor LyF-1)	N	254	CKIGSERSLVLDRLASNVAKRKSSMPQ KFLGD**K**GLSDT**PY**DSSASYEKEN	303
O43474	Kruppel-like factor 4 (Epithelial zinc-finger protein EZF) (Gut-enriched Krueppel-like factor)	N	218	GKFVLKASLSAPGSEYGSPSVISVSKGS PDGS**H**PVVVA**PY**NGGPPRTCPK	267
Q8NEZ4	Histone-lysine N-methyltransferase, H3 lysine-4 specific MLL3	N	2427	NVNQAFTRPPPPYPGNIRSPVAPPLGPR YAVFP**K**DQRG**PY**PPDVASMGMR	2476
Q96G25	Mediator of RNA polymerase II transcription subunit 8 homolog (ARC32)	N	227	GAPSQQQPMLSGVQMAQAGQPGKMPSG IKTNI**K**SASMH**PY**QR	268
Q93074	Mediator of RNA polymerase II transcription subunit 12	N	1854	DLLHHPNPGSITHLNYRQGSIGLYTQN QPLPAGGP**R**VD**PY**RPVRLPMQKL	1903
O43312	Metastasis suppressor protein 1 (Metastasis suppressor YGL-1)	not known	379	LPRVTSVHLPDYAHYYTIGPGMFPSSQ IPSWKDWA**K**PG**PY**DQPLVNTLQR	428
Q13310	Polyadenylate-binding protein 4	C	484	GAAQQGLTDSCQSGGVPTAVQNLAPRA AVAAAAP**R**AVA**PY**KYASSVRSPH	533
Q9Y6V0	Piccolo protein (Aczonin)	C	2874	VVYKLPFGRSCTAQQPATTLPEDRFGYR DDHYQYD**R**SG**PY**GYRGIGGMKP	2923
Q8NFH8	RalBP1-associated Eps domain-containing protein 2 (RalBP1-interacting protein 2)	C	188	PTMSPLASPPSSPPHYQRVPLSHGYSKL RSSAEQM**H**PA**PY**EARQPLVQPE	237
O75177	SS18-like protein 1 (SYT homolog 1)	not known	196	SHYSSAQGGSQHYQGQSSIAMMGQGSQGSSMMGQ**R**PMA**PY**RPSQQGSSQQ	245
Q92922	SWI/SNF complex 155 kDa subunit (BRG1-associated factor 155)	C, N	960	QQQHGQNPQQAHQHSGGPGLAPLGAAGHPGMMP**H**QQPP**PY**PLMHHQMPPP	1009
P09012	U1 small nuclear ribonucleoprotein A (U1 snRNP protein A)	N	123	AVQGGGATPVVGAVQGPVPGMPPMTQAPRIMH**H**MPGQP**PY**MPPPGMIPPP	172
P18583	SON3/Negative regulatory element-binding protein/DBP-5	N	945	GQDPYRLGHDPYRLTPDPYRMSPRPYRI APRSY**R**IAPR**PY**RLAPRPLMLA	994

Table 7-8 continued: Predicted Kapβ2 substrates with hydrophobic PY-NLSs

Accession number	Name	Localization[a]	N-Term. Residue	Sequences for Candidate Hydrophobic PY-NLS[b]	C-Term. Residue
Q8IXZ3	Transcription factor Sp8 (Specificity protein 8)	N	164	GGSSAHSQDGSHQPVFISKVHTSVDGL QGIYP**R**VGMAH**PY**ESWFKPSHPG	213
Q15532	SSXT protein (SYT protein)	not known	214	QYNMPQGGGQHYQGQQPPMGMMGQVNQGNHMMGQ**R**QIP**PY**RPPQQGPPQQ	263
Q9UMS6	Synaptopodin-2 (Myopodin) (Genethonin 2)[c]	C, N	931	PSYPLAALKSQPSAAQPSKMGKKKGKK PLNALDVM**K**HQ**PY**QLNASLFTFQ	980
Q9Y5Q8	General transcription factor 3C polypeptide 5	N	31	GVVRDVAKMLPTLGGEEGVSRIYADPT KRLELYF**R**PKD**PY**CHPVCANRFS	80
Q04206	Transcription factor p65 (Nuclear factor NF-kappa-B p65 subunit)	C, N	310	KSIMKKSPFSGPTDPRPPPRRIAVPSR SSASVP**K**PAPQ**PY**PFTSSLSTIN	359
Q9NRE2	Teashirt homolog 2 (Zinc finger protein 218) (Ovarian cancer-related protein 10-2)	N	558	LPMGSRVLQIRPNLTNKLRPIAPKWKV MPLVSMPT**H**LA**PY**TQVKKESEDK	607
Q9UJT2	Testis-specific serine kinase substrate	not known	275	PAATSQGCPGPPGSPDKPSRPHGLVPA GWGMGP**R**AGEG**PY**VSEQELQKLF	324
Q8TAP9	TTD nonphotosensitive 1 protein	N	15	GPGGGGWGSGSSFRGTPGGGGPRPPSPRDGYGSPH**H**TP**PY**GPRSRPYGSS	64
Q96I51	Williams-Beuren syndrome chromosome region 16 protein (WBS16)	N	62	FVWGFSFSGALGVPSFVVPSSGPGPRAG ARPRR**R**IQPV**PY**RLELDQKISS	111
P19544	Wilms' tumor protein (WT33)	N	94	VHFSGQFTGTAGACRYGPFGPPPPSQAS SGQA**R**MFPNA**PY**LPSCLESQPA	143
P17861	X box-binding protein 1 (XBP-1) (Tax-responsive element-binding protein 5)	N	202	ISCWAFWTTWTQSCSSNALPQSLPAWRS SQRSTQ**K**DPV**PY**QPPFLCQWGR	251
Q8NAP3	Zinc finger and BTB domain-containing protein 38	N	539	HAIDHRLSISKKTANGGLKPSVYPYKLY RLLPMKC**K**RA**PY**KSYRNSSYEN	588
Q9C0A1	Zinc finger homeobox protein 2	N	784	VKPPATATPASLPKFNLLLGKVDDGTGR EAPK**RE**APAF**PY**PTATLASGPQ	833

[a] As annotated in the UniProtKB/Swiss-Prot entries. C represents cytoplasm and N represents nucleus.
[b] Central hydrophobic motifs are underlined and the R/K/H-PY motifs are in bold.
[c] Substrates also identified using bPY-NLS motif.

Table 7-9: Predicted Kapβ2 substrates with basic PY-NLSs

Accession number	Name	Localization[a]	N-Term. Residue	Sequences for Candidate Hydrophobic PY-NLS[b]	C-Term. Residue
Q13023	A-kinase anchor protein 6 (AKAP 100)	not available	1851	GSVKRVSENNGNGKNSSHTHELGT KRENKKTIFKVN**K**D**PY**VADMENGNIE	1900
Q9BXP5	Arsenite-resistance protein 2[c]	not available	61	NRRERFSPPRHELSPPQKRMRRDWD EHSSDPY**H**SGYEM**PY**AGGGGGPTYG	110
Q92560	BRCA1-associated protein 1[c]	N	685	EGMLANLVEQNISVRRRQGVSIG RLHKQRKPDRRK**R**SR**PY**KAKRQ	729
Q9NYF8	Bcl-2-associated transcription factor 1	C, N	32	KRYSSRSRSRTYSRSRSRDRMYSRD YRRDYRNNRGM**R**R**PY**GYRGRGRGYY	81
Q9ULD4	Bromodomain and PHD finger-containing protein 3	not available	1	MRKPRRKSRQNAEGR**R**SPS**PY**SLKC SPTRET	31
Q9UK58	Cyclin-L1	N	337	ASKPSSPREVKAEEKSPISINVKTVKK EPEDRQQAS**K**S**PY**NGVRKDSKRS	386
Q9NYF5	Protein C5orf5 (GAP-like protein N61)	not available	531	QRFLHDPEKLDSSSKALSFT RIRRSSFSSKDEKREDR**T**P**Y**QLVKKLQKKI	580
P49761	CDC-like kinase 3[c]	N	62	RERRDSDTYRCEERSPSFGEDYYGPS RSRHRRRSRE**R**G**PY**RTRKHAHHCH	111
Q8TBR5	Protein C19orf23[c]	not available	70	TWQTRNHTRTGHAYPRFTRPSFPS CNRNGKRRKL**R**LGL**PY**	109
Q92782	Zinc-finger protein neuro-d4	C, N	156	EDLEDDIPRRKNRAKGKAYGIGGLRKR QDTASLED**R**DK**PY**VCDKFYKELA	205
O00358	Forkhead box protein E1/Thyroid transcription factor 2	N	17	TVKEERGETAAGAGVPGEATGRGAGG RRRKRPLQ**R**GKP**PY**SYIALIAMAI	66
Q13461	Forkhead box protein E3 (FKHL12) (Forkhead-related transcription factor 8)	N	35	AEPGREPEEAAAGRGEAAPTPAPGPG RRRRRPLQ**R**GKP**PY**SYIALIAMAL	84
O75593	Forkhead box protein F1	N	1	MDPASSGPSKAKKTNAGIR**R**PEKP**PY**SYIALIVMAI	36
O75593	Forkhead box protein H1/Forkhead activin signal transducer 1	N	1	MGPCSGSRLGPPEAESPSQPPKRRKKR YL**R**HDKP**PY**TYLAMIALVI	46
Q9UPW0	Forkhead box protein J3	N	142	SKDDPGKGSYWAIDTNPKEDALPT RPKKRARSVE**R**AST**PY**SIDSDSLGME	191
P55317	Hepatocyte nuclear factor 3-alpha (Forkhead box protein A1).	N	135	MNPCMSPMAYAPSNLGRSRAGGGGDAKTFKRSYPHA**K**P**PY**SYISLITMAI	184
P55318	Hepatocyte nuclear factor 3-gamma (Forkhead box protein A3)	N	81	LGVSGGSSSSGYGAPGPGLVHGKEMP KGYRRPLAHA**K**P**PY**SYISLITMAI	130
Q9Y483	Metal-response element-binding transcription factor 2	N	370	HEFKIKGRKASKPISDSREVSNGIE KKGKKKSVG**R**PPG**PY**TRKMIQKTAE	419
O95644	NFAT transcription complex cytosolic component	C, N	238	PSTSPRASVTEESWLGARSSRPASP CNKRKYSLNG**R**QP**PY**SPHHSPTPSP	287
Q9ULL1	Pleckstrin homology domain-containing family G member 1	not available	1304	SKFVDADFSDNVCSGNTLHSLNSP RTPKKPVNS**K**LGLS**PY**LTPYNDSDKL	1353
Q99575	Ribonucleases P/MRP protein subunit POP1	N	372	QTELPDEKIGKKRKRKDDGENAKPIKK IIGDGT**R**DPCL**PY**SWISPTTGII	421
Q8NEY8	Periphilin 1/Gastric cancer antigen Ga50	C, N	84	YRWTRDDHSASRQPEYRDMRDGFRRKS FYSSHYARE**R**S**PY**KRDNTFFRES	133

Table 7-9 continued: Predicted Kapβ2 substrates with basic PY-NLSs

Q8NDT2	RNA-binding protein 15B	N	245	SRSGERWGADGDRGLPKPWEERRKRR SLSSDRG**R**TTHS**PY**EERSRTKGSG	294
Q14498	Splicing factor HCC1	N	60	DRERKKSKSRERKRSRSKERRRSRSRSRDRRFRG**R**YRS**PY**SGPKFNSAIR	109
P62241	40S ribosomal protein S8	N	1	GISRDNWHKRRKTGGK**R**K**PY**HKKRKYELGR	30
O95416	Transcription factor SOX-14	N	59	DEAKRLRAQHMKEHPDYKYRPRRKPKNLLKKD**R**YVFPL**PY**LGDTDPLKAA	108
Q9Y651	Transcription factor SOX-21 (SOX-A)	N	59	DEAKRLRAMHMKEHPDYKYRPRRKPKTLLKKD**K**FAFPV**PY**GLGGVADAEH	108
O00267	Transcription elongation factor SPT5	N	678	GGQRGGFGSPGGGSGGMSRGRGRRDNELIGQTV**R**ISQG**PY**KGYIGVVKDA	727
Q9UMS6	Synaptopodin-2 (Myopodin) (Genethonin 2)[c]	C, N	931	PSYPLAALKSQPSAAQPSKMGKKKGKKPLNALDVM**K**HQ**PY**QLNASLFTFQ	980
Q8IWR0	Zinc finger CCCH-type domain-containing protein 7A	N	464	ANIDHKCKKDILIGRIKNVEDKSWKKIRPRPT**K**TNYEG**PY**YICKDVAAEE	513
Q9H091	Zinc finger MYND domain-containing protein 15	not available	522	RDSLEVSVRPGSGISARPSSGTKEKGGRRDLQI**K**VSAR**PY**HLFQGPKPDL	571
Q9H116	Zinc finger protein 336	N	177	LTDSLDYPGERASNGMSSDLPP KKSKDKLDKKKEVV**K**P**PY**PKIRRASGRL	226
Q8N895	Zinc finger protein 366	N	49	RGPFSQFRYEPPPGDLDGFPGVFEGAGS RKRKSMPT**K**M**PY**NHPAEEVTLA	98

[a] As annotated in the UniProtKB/Swiss-Prot entries. C represents cytoplasm and N represents nucleus.
[b] Central basic-enriched regions are underlined and the R/K/H-PY motifs are in bold.
[c] Substrates also identified using hPY-NLS motif.

BIBLIOGRAPHY

Aitchison, J. D., G. Blobel, et al. (1996). "Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins." Science **274**(5287): 624-7.

Albertini, M., L. F. Pemberton, et al. (1998). "A novel nuclear import pathway for the transcription factor TFIIS." J Cell Biol **143**(6): 1447-55.

Allemand, E., S. Guil, et al. (2005). "Regulation of heterogenous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock." Proc Natl Acad Sci U S A **102**(10): 3605-10.

Anderson, J. T., S. M. Wilson, et al. (1993). "NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability." Mol Cell Biol **13**(5): 2730-41.

Apweiler, R., A. Bairoch, et al. (2004). "Protein sequence databases." Current opinion in chemical biology **8**: 76-80.

Arnaoutov, A., Y. Azuma, et al. (2005). "Crm1 is a mitotic effector of Ran-GTP in somatic cells." Nat Cell Biol **7**(6): 626-32.

Bader, A. G. and P. K. Vogt (2005). "Inhibition of protein synthesis by Y box-binding protein 1 blocks oncogenic cell transformation." Mol Cell Biol **25**(6): 2095-106.

Bairoch, A., B. Boeckmann, et al. (2004). "Swiss-Prot: Juggling between evolution and stability " Brief. Bioinform. **5**: 39-55.

Baker, D. (2006). "Prediction and design of macromolecular structures and interactions." Philos Trans R Soc Lond B Biol Sci **361**(1467): 459-63.

Bakhrat, A., K. Baranes, et al. (2006). "Nuclear import of ho endonuclease utilizes two nuclear localization signals and four importins of the ribosomal import system." J Biol Chem **281**(18): 12218-26.

Bear, J., W. Tan, et al. (1999). "Identification of novel import and export signals of human TAP, the protein that binds to the constitutive transport element of the type D retrovirus mRNAs." Mol Cell Biol **19**(9): 6306-17.

Belyanskaya, L. L., P. M. Gehrig, et al. (2001). "Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein." J Biol Chem **276**(22): 18681-7.

Benson, D. A., I. Karsch-Mizrachi, et al. (2006). "GenBank." Nucleic Acids Res **34**(Database issue): D16-20.

Bertolotti, A., T. Melot, et al. (1998). "EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes." Mol Cell Biol **18**(3): 1489-97.

Bischoff, F. R. and H. Ponstingl (1995). "Catalysis of guanine nucleotide exchange of Ran by RCC1 and stimulation of hydrolysis of Ran-bound GTP by Ran-GAP1." Methods Enzymol **257**: 135-44.

Blackwell, J. S., Jr., S. T. Wilkinson, et al. (2007). "Mutational analysis of H3 and H4 N termini reveals distinct roles in nuclear import." J Biol Chem **282**(28): 20142-50.

Bogerd, H. P., R. E. Benson, et al. (1999). "Definition of a consensus transportin-specific nucleocytoplasmic transport signal." J Biol Chem **274**(14): 9771-7.

Bogerd, H. P., R. A. Fridell, et al. (1996). "Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization-selection assay." Mol Cell Biol **16**(8): 4207-14.

Bonifaci, N., J. Moroianu, et al. (1997). "Karyopherin beta2 mediates nuclear import of a mRNA binding protein." Proc Natl Acad Sci U S A **94**(10): 5055-60.

Brachmann, C. B., A. Davies, et al. (1998). "Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications." Yeast **14**(2): 115-32.

Brocard, C. and A. Hartig (2006). "Peroxisome targeting signal 1: is it really a simple tripeptide?" Biochim Biophys Acta **1763**(12): 1565-73.

Brune, C., S. E. Munchel, et al. (2005). "Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export." RNA **11**(4): 517-31.

Brunger, A. T., A. P. D., et al. (1998). "Crystallography & NMR System: A new software suite for macromolecular structure determination." Acta Cryst. A **D54**: 905-921.

Budhu, A. S. and X. W. Wang (2005). "Loading and unloading: orchestrating centrosome duplication and spindle assembly by Ran/Crm1." Cell Cycle **4**(11): 1510-4.

Caesar, S., M. Greiner, et al. (2006). "Kap120 functions as a nuclear import receptor for ribosome assembly factor Rpf1 in yeast." Mol Cell Biol **26**(8): 3170-80.

Calado, A., U. Kutay, et al. (2000). "Deciphering the cellular pathway for transport of poly(A)-binding protein II." RNA **6**(2): 245-56.

Cansizoglu, A. E., B. J. Lee, et al. (2007). "Structure-based design of a pathway-specific nuclear import inhibitor." Nat Struct Mol Biol **14**(5): 452-4.

Cardoso, M. C. and H. Leonhardt (1999). "DNA methyltransferase is actively retained in the cytoplasm during early development." J Cell Biol **147**(1): 25-32.

Catimel, B., T. Teh, et al. (2001). "Biophysical characterization of interactions involving importin-alpha during nuclear import." J Biol Chem **276**(36): 34189-98.

CCP4 (1994). "The CCP4 suite: programs for X-ray crystallography." Acta Crystallogr. D **50**: 760-763.

Chan, C. K., S. Hubner, et al. (1998). "Mutual exclusivity of DNA binding and nuclear localization signal recognition by the yeast transcription factor GAL4: implications for nonviral DNA delivery." Gene Ther **5**(9): 1204-12.

Chaves, S. R. and G. Blobel (2001). "Nuclear import of spo12p, a protein essential for meiosis." J Biol Chem **276**(21): 17712-7.

Chen, H. T., L. Warfield, et al. (2007). "The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex." Nat Struct Mol Biol **14**(8): 696-703.

Chook, Y. and G. Blobel (1999). "Structure of the nuclear transport complex karyopherin-beta2-Ran.GppNHp." Nature **399**: 230-237.

Chook, Y. M. and G. Blobel (1999). "Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp." Nature **399**(6733): 230-7.

Chook, Y. M. and G. Blobel (2001). "Karyopherins and nuclear import." Current Opinions in Structural Biology **11**(6).

Chook, Y. M. and G. Blobel (2001). "Karyopherins and nuclear import." Curr Opin Struct Biol **11**(6): 703-15.

Chook, Y. M., A. Jung, et al. (2002). "Uncoupling Kapb2 substrate dissociation and Ran binding." Biochemistry **41**: 6955-6966.

Chung, W. H., J. L. Craighead, et al. (2003). "RNA polymerase II/TFIIF structure and conserved organization of the initiation complex." Mol Cell **12**(4): 1003-13.

Cingolani, G., J. Bednenko, et al. (2002). "Molecular basis for the recognition of a nonclassical nuclear localization signal by importin beta." Mol Cell **10**(6): 1345-53.

Cingolani, G., C. Petosa, et al. (1999). "Structure of importin-beta bound to the IBB domain of importin-alpha." Nature **399**(6733): 221-9.

Conti, E. and E. Izaurralde (2001). "Nucleocytoplasmic transport enters the atomic age." Curr Opin Cell Biol **13**(3): 310-9.

Conti, E. and J. Kuriyan (2000). "Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha." Structure **8**(3): 329-38.

Conti, E. and J. Kuriyan (2000). "Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha." Structure Fold Des **8**(3): 329-38.

Conti, E., M. Uy, et al. (1998). "Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha." Cell **94**(2): 193-204.

Damelin, M., P. A. Silver, et al. (2002). "Nuclear protein transport." Methods Enzymol **351**: 587-607.

Delahodde, A., R. Pandjaitan, et al. (2001). "Pse1/Kap121-dependent nuclear localization of the major yeast multidrug resistance (MDR) transcription factor Pdr1." Mol Microbiol **39**(2): 304-12.

DeLano, W. L. (2002). Pymol. (DeLano Scientific, San Carlos, CA).

DeLano, W. L. (2002). "The PyMOL User's Manual, DeLano Scientific, San Carlos, CA ".

Dingwall, C., S. V. Sharnick, et al. (1982). "A polypeptide domain that specifies migration of nucleoplasmin into the nucleus." Cell **30**(2): 449-58.

Dong, X., A. Biswas, et al. (2009). Structural basis for leucine-rich nuclear export signal recognition by CRM1.

Efthymiadis, A., H. Shao, et al. (1997). "Kinetic characterization of the human retinoblastoma protein bipartite nuclear localization sequence (NLS) in vivo and in vitro. A comparison with the SV40 large T-antigen NLS." J Biol Chem **272**(35): 22134-9.

Enenkel, C., G. Blobel, et al. (1995). "Identification of a yeast karyopherin heterodimer that targets import substrate to mammalian nuclear pore complexes." J Biol Chem **270**(28): 16499-502.

Fan, X. C. and J. A. Steitz (1998). "HNS, a nuclear-cytoplasmic shuttling sequence in HuR." Proc Natl Acad Sci U S A **95**(26): 15293-8.

Ferrigno, P., F. Posas, et al. (1998). "Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin beta homologs NMD5 and XPO1." Embo J **17**(19): 5606-14.

Fischer, U., J. Huber, et al. (1995). "The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs." Cell **82**(3): 475-83.

Fontes, M. R., T. Teh, et al. (2003). "Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha." J Biol Chem **278**(30): 27981-7.

Fontes, M. R., T. Teh, et al. (2000). "Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha." J Mol Biol **297**(5): 1183-94.

Fontes, M. R., T. Teh, et al. (2000). "Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha." J Mol Biol **297**(5): 1183-94.

Fornerod, M., M. Ohno, et al. (1997). "CRM1 is an export receptor for leucine-rich nuclear export signals [see comments]." Cell **90**(6): 1051-60.

Franke, J., B. Reimann, et al. (2001). "Evidence for a nuclear passage of nascent polypeptide-associated complex subunits in yeast." J Cell Sci **114**(Pt 14): 2641-8.

Freire-Picos, M. A., S. Krishnamurthy, et al. (2005). "Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex." Nucleic Acids Res **33**(16): 5045-52.

Fridell, R. A., R. Truant, et al. (1997). "Nuclear import of hnRNP A1 is mediated by a novel cellular cofactor related to karyopherin-beta." J Cell Sci **110**(Pt 11): 1325-31.

Fried, H. and U. Kutay (2003). "Nucleocytoplasmic transport: taking an inventory." Cell Mol Life Sci **60**(8): 1659-88.

Fries, T., C. Betz, et al. (2007). "A novel conserved nuclear localization signal is recognized by a group of yeast importins." J Biol Chem **282**(27): 19292-301.

Fukuda, M., S. Asano, et al. (1997). "CRM1 is responsible for intracellular transport meditted by the nuclear export signal." Nature **390**: 308-311.

Gaiser, F., S. Tan, et al. (2000). "Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution." J Mol Biol **302**(5): 1119-27.

Garreau, H., M. Geymonat, et al. (1996). "Membrane-anchoring domains of Cdc25p, a Saccharomyces cerevisiae ras exchange factor." Biol Cell **86**(2-3): 93-102.

Gattiker, A., E. Gasteiger, et al. (2002). "ScanProsite: a reference implementation of a PROSITE scanning tool." Applied Bioinformatics **1**: 107-108.

Ghazy, M. A., S. A. Brodie, et al. (2004). "Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II." Mol Cell Biol **24**(24): 10975-85.

Gilson, M. K. and H. X. Zhou (2007). "Calculation of protein-ligand binding affinities." Annu Rev Biophys Biomol Struct **36**: 21-42.

Gorlich, D. and U. Kutay (1999). "Transport between the cell nucleus and the cytoplasm." Annu Rev Cell Dev Biol **15**: 607-60.

Gorlich, D., M. J. Seewald, et al. (2003). "Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation." EMBO J **22**(5): 1088-100.

Green, D. M., K. A. Marfatia, et al. (2002). "Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p." J Biol Chem **277**(10): 7752-60.

Greiner, M., S. Caesar, et al. (2004). "The histones H2A/H2B and H3/H4 are imported into the yeast nucleus by different mechanisms." Eur J Cell Biol **83**(10): 511-20.

Groft, C. M., S. N. Uljon, et al. (1998). "Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly." Proc Natl Acad Sci U S A **95**(16): 9117-22.

Grosshans, H., K. Deinert, et al. (2001). "Biogenesis of the signal recognition particle (SRP) involves import of SRP proteins into the nucleolus, assembly with the SRP-RNA, and Xpo1p-mediated export." J Cell Biol **153**(4): 745-62.

Guttinger, S., P. Muhlhausser, et al. (2004). "Transportin2 functions as importin and mediates nuclear import of HuR." Proc Natl Acad Sci U S A **101**(9): 2918-23.

Hacker, S. and H. Krebber (2004). "Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins." J Biol Chem **279**(7): 5049-52.

Harel, A. and D. J. Forbes (2004). "Importin beta: conducting a much larger cellular symphony." Mol Cell **16**(3): 319-30.

Hegde, R. S. and H. D. Bernstein (2006). "The surprising complexity of signal sequences." Trends Biochem Sci **31**(10): 563-71.

Henry, M. F. and P. A. Silver (1996). "A novel methyltransferase (Hmt1p) modifies poly(A)+-RNA-binding proteins." Mol Cell Biol **16**(7): 3668-78.

Henry, N. L., A. M. Campbell, et al. (1994). "TFIIF-TAF-RNA polymerase II connection." Genes Dev **8**(23): 2868-78.

Ho, Y., A. Gruhler, et al. (2002). "Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry." Nature **415**(6868): 180-3.

Hodel, A. E., M. T. Harreman, et al. (2006). "Nuclear localization signal receptor affinity correlates with in vivo localization in Saccharomyces cerevisiae." J Biol Chem **281**(33): 23545-56.

Hodel, M. R., A. H. Corbett, et al. (2001). "Dissection of a nuclear localization signal." J Biol Chem **276**(2): 1317-25.

Hodel, M. R., A. H. Corbett, et al. (2001). "Dissection of a nuclear localization signal." J Biol Chem **276**(2): 1317-25.

Hodges, J. L., J. H. Leslie, et al. (2005). "Nuclear import of TFIIB is mediated by Kap114p, a karyopherin with multiple cargo-binding domains." Mol Biol Cell. **16**(7): 3200-10.

Horton, P., K. J. Park, et al. (2007). "WoLF PSORT: protein localization predictor." Nucleic Acids Res **35**(Web Server issue): W585-7.

Huber, J., U. Cronshagen, et al. (1998). "Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure." Embo J **17**(14): 4114-26.

Huh, W. K., J. V. Falvo, et al. (2003). "Global analysis of protein localization in budding yeast." Nature **425**(6959): 686-91.

Humphris, E. L. and T. Kortemme (2007). "Design of multi-specificity in protein interfaces." PLoS Comput Biol **3**(8): e164.

Iijima, M., M. Suzuki, et al. (2006). "Two motifs essential for nuclear import of the hnRNP A1 nucleocytoplasmic shuttling sequence M9 core." FEBS Lett **580**(5): 1365-70.

Imasaki, T., T. Shimizu, et al. (2007). "Structural Basis for Substrate Recognition and Dissociation by Human Transportin 1." Molecular Cell **28**: 57-67.

Ishidate, T., S. Yoshihara, et al. (1997). "Identification of a novel nuclear localization signal in Sam68." FEBS Lett **409**(2): 237-41.

Isoyama, T., A. Murayama, et al. (2001). "Nuclear import of the yeast ap-1-like transcription factor yap1p is mediated by transport receptor pse1p, and this import step is not affected by oxidative stress." J Biol Chem **276**(24): 21863-9.

Jones, T. A., S. W. Cowan, et al. (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr. A **47**: 110-119.

Kaffman, A., N. M. Rank, et al. (1998). "Phosphorylation regulates association of the transcription factor Pho4 with its import receptor Pse1/Kap121." Genes Dev **12**(17): 2673-83.

Kalderon, D., W. D. Richardson, et al. (1984). "Sequence requirements for nuclear location of simian virus 40 large-T antigen." Nature **311**(5981): 33-8.

Kalderon, D., W. D. Richardson, et al. (1984). "Sequence requirements for nuclear location of simian virus 40 large-T antigen." Nature. **311**(5981): 33-38.

Katahira, J., K. Strasser, et al. (1999). "The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human." Embo J **18**(9): 2593-609.

Kawamura, H., Y. Tomozoe, et al. (2002). "Identification of the nucleocytoplasmic shuttling sequence of heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its interaction with mRNA." J Biol Chem **277**(4): 2732-9.

Khaperskyy, D. A., M. L. Ammerman, et al. (2008). "Functions of Saccharomyces cerevisiae TFIIF during transcription start site utilization." Mol Cell Biol **28**(11): 3757-66.

Kortemme, T. and D. Baker (2002). "A simple physical model for binding energy hot spots in protein-protein complexes." Proc Natl Acad Sci U S A **99**(22): 14116-21.

Kortemme, T. and D. Baker (2004). "Computational design of protein-protein interactions." Curr Opin Chem Biol **8**(1): 91-7.

Kosugi, S., M. Hasebe, et al. (2008). "Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling." Chem Biol **15**(9): 940-9.

la Cour, T., L. Kiemer, et al. (2004). "Analysis and prediction of leucine-rich nuclear export signals." Protein Eng Des Sel **17**(6): 527-36.

Lanford, R. E. and J. S. Butel (1984). "Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen." Cell **37**(3): 801-13.

Lange, A., R. E. Mills, et al. (2008). "A PY-NLS nuclear targeting signal is required for nuclear localization and function of the Saccharomyces cerevisiae mRNA-binding protein Hrp1." J Biol Chem **283**(19): 12926-34.

Lange, A., R. E. Mills, et al. (2007). "Classical nuclear localization signals: definition, function, and interaction with importin alpha." J Biol Chem **282**(8): 5101-5.

Lee, B. J., A. E. Cansizoglu, et al. (2006). "Rules for nuclear localization sequence recognition by karyopherin beta 2." Cell **126**(3): 543-58.

Lee, D. C. and J. D. Aitchison (1999). "Kap104p-mediated nuclear import. Nuclear localization signals in mRNA-binding proteins and the role of Ran and Rna." J Biol Chem **274**(41): 29031-7.

Lee, S. J., T. Sekimoto, et al. (2003). "The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor." Science **302**(5650): 1571-5.

Leslie, D. M., B. Grill, et al. (2002). "Kap121p-mediated nuclear import is required for mating and cellular differentiation in yeast." Mol Cell Biol **22**(8): 2544-55.

Leslie, D. M., W. Zhang, et al. (2004). "Characterization of karyopherin cargoes reveals unique mechanisms of Kap121p-mediated nuclear import." Mol Cell Biol **24**(19): 8487-503.

Linding, R., L. J. Jensen, et al. (2003). "Protein disorder prediction: implications for structural proteomics." Structure (Camb) **11**(11): 1453-9.

Lukong, K. E., D. Larocque, et al. (2005). "Tyrosine phosphorylation of sam68 by breast tumor kinase regulates intranuclear localization and cell cycle progression." J Biol Chem **280**(46): 38639-47.

Lusk, C. P., T. Makhnevych, et al. (2002). "Karyopherins in nuclear pore biogenesis: a role for Kap121p in the assembly of Nup53p into nuclear pore complexes." J Cell Biol **159**(2): 267-78.

Ma, A. S., K. Moran-Jones, et al. (2002). "Heterogeneous nuclear ribonucleoprotein A3, a novel RNA trafficking response element-binding protein." J Biol Chem **277**(20): 18010-20.

Makhnevych, T., C. P. Lusk, et al. (2003). "Cell cycle regulated transport controlled by alterations in the nuclear pore complex." Cell **115**(7): 813-23.

Mancias, J. D. and J. Goldberg (2005). "Exiting the endoplasmic reticulum." Traffic **6**(4): 278-85.

Marelli, M., C. P. Lusk, et al. (2001). "A link between the synthesis of nucleoporins and the biogenesis of the nuclear envelope." J Cell Biol **153**(4): 709-24.

Marfatia, K. A., E. B. Crafton, et al. (2003). "Domain analysis of the Saccharomyces cerevisiae heterogeneous nuclear ribonucleoprotein, Nab2p. Dissecting the requirements for Nab2p-facilitated poly(A) RNA export." J Biol Chem **278**(9): 6731-40.

Matsuyama, A., R. Arai, et al. (2006). "ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe." Nat Biotechnol **24**(7): 841-7.

McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2005). "Likelihood-enhanced fast translation functions." Acta Cryst. **D61**
458-464.

Morehouse, H., R. M. Buratowski, et al. (1999). "The importin/karyopherin Kap114 mediates the nuclear import of TATA-binding protein." Proc Natl Acad Sci U S A **96**(22): 12542-7.

Mosammaparast, N., B. C. Del Rosario, et al. (2005). "Modulation of histone deposition by the karyopherin kap114." Mol Cell Biol **25**(5): 1764-78.

Mosammaparast, N., C. S. Ewart, et al. (2002). "A role for nucleosome assembly protein 1 in the nuclear transport of histones H2A and H2B." EMBO J **21**(23): 6527-38.

Mosammaparast, N., Y. Guo, et al. (2002). "Pathways mediating the nuclear import of histones H3 and H4 in yeast." J Biol Chem **277**(1): 862-8.

Mosammaparast, N., K. R. Jackson, et al. (2001). "Nuclear import of histone H2A and H2B is mediated by a network of karyopherins." J Cell Biol **153**(2): 251-62.

Mosammaparast, N., K. R. Jackson, et al. (2001). "Nuclear import of histone H2A and H2B is mediated by a network of karyopherins." J Cell Biol **153**(2): 251-62.

Mosammaparast, N. and L. F. Pemberton (2004). "Karyopherins: from nuclear-transport mediators to nuclear-function regulators." Trends Cell Biol. **14**(10): 547-56.

Mossessova, E., L. C. Bickford, et al. (2003). "SNARE selectivity of the COPII coat." Cell **114**(4): 483-95.

Nachury, M. V. and K. Weis (1999). "The direction of transport through the nuclear pore can be inverted." Proc Natl Acad Sci U S A **96**(17): 9622-7.

Nakielny, S., M. C. Siomi, et al. (1996). "Transportin: nuclear transport receptor of a novel nuclear protein import pathway." Exp Cell Res **229**(2): 261-6.

Nayeem, A., D. Sitkoff, et al. (2006). "A comparative study of available software for high-accuracy homology modeling: from sequence alignments to structural models." Protein Sci **15**(4): 808-24.

Neduva, V., R. Linding, et al. (2005). "Systematic discovery of new recognition peptides mediating protein interaction networks." PLoS Biol **3**(12): e405.

Neduva, V. and R. B. Russell (2005). "Linear motifs: evolutionary interaction switches." FEBS Lett **579**(15): 3342-5.

Nemergut, M. E., C. A. Mizzen, et al. (2001). "Chromatin docking and exchange activity enhancement of RCC1 by histones H2A and H2B." Science **292**(5521): 1540-3.

Nicholls, A., K. A. Sharp, et al. (1991). "Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons." Proteins: Struct. Funct. Genet. **11**: 281-296.

Nikolaev, I., M. F. Cochet, et al. (2003). "Nuclear import of zinc binuclear cluster proteins proceeds through multiple, overlapping transport pathways." Eukaryot Cell **2**(2): 209-21.

Nishi, K., M. Yoshida, et al. (1994). "Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression." J Biol Chem **269**(9): 6320-4.

Ossareh-Nazari, B., F. Bachelerie, et al. (1997). "Evidence for a role of CRM1 in signal-mediated nuclear protein export." Science **278**(5335): 141-4.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray Diffraction Data Collected in Oscillation Mode." Methods in Enzymology **276**: 307-326.

Paraskeva, E., E. Izaurralde, et al. (1999). "CRM1-mediated recycling of snurportin 1 to the cytoplasm." J Cell Biol **145**(2): 255-64.

Pemberton, L. F. and B. M. Paschal (2005). "Mechanisms of receptor-mediated nuclear import and nuclear export." Traffic **6**(3): 187-98.

Pemberton, L. F., J. S. Rosenblum, et al. (1997). "A distinct and parallel pathway for the nuclear import of an mRNA- binding protein." J Cell Biol **139**(7): 1645-53.

Pemberton, L. F., J. S. Rosenblum, et al. (1999). "Nuclear import of the TATA-binding protein: mediation by the karyopherin Kap114p and a possible mechanism for intranuclear targeting." J Cell Biol **145**(7): 1407-17.

Polizotto, R. S. and M. S. Cyert (2001). "Calcineurin-dependent nuclear import of the transcription factor Crz1p requires Nmd5p." J Cell Biol **154**(5): 951-60.

Pollard, V. W., W. M. Michael, et al. (1996). "A novel receptor-mediated nuclear protein import pathway." Cell **86**(6): 985-94.

Puntervoll, P., R. Linding, et al. (2003). "ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins." Nucleic Acids Res **31**(13): 3625-30.

Quan, X., R. Rassadi, et al. (2004). "Regulated nuclear accumulation of the yeast hsp70 Ssa4p in ethanol-stressed cells is mediated by the N-terminal domain, requires the nuclear carrier Nmd5p and protein kinase C." FASEB J **18**(7): 899-901.

Rapaport, D. (2003). "Finding the right organelle. Targeting signals in mitochondrial outer-membrane proteins." EMBO Rep **4**(10): 948-52.

Rappsilber, J., W. J. Friesen, et al. (2003). "Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode." Anal Chem **75**(13): 3107-14.

Rebane, A., A. Aab, et al. (2004). "Transportins 1 and 2 are redundant nuclear import factors for hnRNP A1 and HuR." Rna **10**(4): 590-9.

Richards, S. A., K. L. Carey, et al. (1997). "Requirement of guanosine triphosphate-bound ran for signal-mediated nuclear protein export." Science **276**(5320): 1842-4.

Riddick, G. and I. G. Macara (2005). "A systems analysis of importin-{alpha}-{beta} mediated nuclear protein import." J Cell Biol **168**(7): 1027-38.

Riddick, G. and I. G. Macara (2007). "The adapter importin-alpha provides flexible control of nuclear import at the expense of efficiency." Mol Syst Biol **3**: 118.

Robbins, J., S. M. Dilworth, et al. (1991). "Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence." Cell **64**(3): 615-23.

Rosenblum, J. S., L. F. Pemberton, et al. (1997). "A nuclear import pathway for a protein involved in tRNA maturation." J Cell Biol **139**(7): 1655-61.

Rosenblum, J. S., L. F. Pemberton, et al. (1998). "Nuclear import and the evolution of a multifunctional RNA-binding protein." J Cell Biol **143**(4): 887-99.

Rout, M. P., G. Blobel, et al. (1997). "A distinct nuclear import pathway used by ribosomal proteins." Cell **89**(5): 715-25.

Schaper, S., J. Franke, et al. (2005). "Nuclear import of the histone acetyltransferase complex SAS-I in Saccharomyces cerevisiae." J Cell Sci **118**(Pt 7): 1473-84.

Senger, B., G. Simos, et al. (1998). "Mtr10p functions as a nuclear import receptor for the mRNA-binding protein Npl3p." Embo J **17**(8): 2196-207.

Shaheen, H. H. and A. K. Hopper (2005). "Retrograde movement of tRNAs from the cytoplasm to the nucleus in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **102**(32): 11290-5.

Shen, E. C., M. F. Henry, et al. (1998). "Arginine methylation facilitates the nuclear export of hnRNP proteins." Genes Dev **12**(5): 679-91.

Sikorski, R. S. and P. Hieter (1989). "A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae." Genetics **122**(1): 19-27.

Siomi, H. and G. Dreyfuss (1995). "A nuclear localization domain in the hnRNP A1 protein." J Cell Biol **129**(3): 551-560.

Siomi, M. C., P. S. Eder, et al. (1997). "Transportin-mediated nuclear import of heterogeneous nuclear RNP proteins." J Cell Biol **138**(6): 1181-92.

Siomi, M. C., M. Fromont, et al. (1998). "Functional conservation of the transportin nuclear import pathway in divergent organisms." Mol Cell Biol **18**(7): 4141-8.

Smith, A. E., B. M. Slepchenko, et al. (2002). "Systems analysis of Ran transport." Science **295**(5554): 488-91.

Stade, K., C. S. Ford, et al. (1997). "Exportin 1 (Crm1p) is an essential nuclear export factor." Cell **90**(6): 1041-50.

Suel, K. E., H. Gu, et al. (2008). "Modular organization and combinatorial energetics of proline-tyrosine nuclear localization signals." PLoS Biol **6**(6): e137.

Suzuki, M., M. Iijima, et al. (2005). "Two separate regions essential for nuclear import of the hnRNP D nucleocytoplasmic shuttling sequence." Febs J **272**(15): 3975-87.

Swanton, E. and S. High (2006). "ER targeting signals: more than meets the eye?" Cell **127**(5): 877-9.

Tan, S., R. C. Conaway, et al. (1995). "Dissection of transcription factor TFIIF functional domains required for initiation and elongation." Proc Natl Acad Sci U S A **92**(13): 6042-6.

Timney, B. L., J. Tetenbaum-Novatt, et al. (2006). "Simple kinetic relationships and nonspecific competition govern nuclear import rates in vivo." J Cell Biol **175**(4): 579-93.

Titov, A. A. and G. Blobel (1999). "The karyopherin Kap122p/Pdr6p imports both subunits of the transcription factor IIA into the nucleus." J Cell Biol **147**(2): 235-46.

Tran, E. J., T. A. Bolger, et al. (2007). "SnapShot: nuclear transport." Cell **131**(2): 420.

Truant, R., R. A. Fridell, et al. (1998). "Identification and functional characterization of a novel nuclear localization signal present in the yeast Nab2 poly(A)+ RNA binding protein." Mol Cell Biol **18**(3): 1449-58.

Truant, R., Y. Kang, et al. (1999). "The human tap nuclear RNA export factor contains a novel transportin-dependent nuclear localization signal that lacks nuclear export signal function." J Biol Chem **274**(45): 32167-71.

Ueta, R., A. Fukunaka, et al. (2003). "Pse1p mediates the nuclear import of the iron-responsive transcription factor Aft1p in Saccharomyces cerevisiae." J Biol Chem **278**(50): 50120-7.

Van Ael, E. and M. Fransen (2006). "Targeting signals in peroxisomal membrane proteins." Biochim Biophys Acta **1763**(12): 1629-38.

Walsh, S. T., J. E. Sylvester, et al. (2004). "The high- and low-affinity receptor binding sites of growth hormone are allosterically coupled." Proc Natl Acad Sci U S A **101**(49): 17078-83.

Waragai, M., C. H. Lammers, et al. (1999). "PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival." Hum Mol Genet **8**(6): 977-87.

Weighardt, F., G. Biamonti, et al. (1995). "Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1." J Cell Sci **108 (Pt 2)**: 545-55.

Weis, K. (2003). "Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle." Cell **112**(4): 441-51.

Wen, W., J. L. Meinkoth, et al. (1995). "Identification of a signal for rapid export of proteins from the nucleus." Cell **82**(3): 463-73.

Windgassen, M. and H. Krebber (2003). "Identification of Gbp2 as a novel poly(A)+ RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast." EMBO Rep **4**(3): 278-83.

Wu, J., L. Zhou, et al. (1999). "The quaking I-5 protein (QKI-5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm." J Biol Chem **274**(41): 29202-10.

Xiao, C. Y., P. Jans, et al. (1998). "Negative charge at the protein kinase CK2 site enhances recognition of the SV40 large T-antigen NLS by importin: effect of conformation." FEBS Lett **440**(3): 297-301.

Xu, C. and M. F. Henry (2004). "Nuclear export of hnRNP Hrp1p and nuclear export of hnRNP Npl3p are linked and influenced by the methylation state of Npl3p." Mol Cell Biol **24**(24): 10742-56.

Yoshida, K. and G. Blobel (2001). "The karyopherin Kap142p/Msn5p mediates nuclear import and nuclear export of different cargo proteins." J Cell Biol **152**(4): 729-40.

Zakaryan, R. P. and H. Gehring (2006). "Identification and characterization of the nuclear localization/retention signal in the EWS proto-oncoprotein." J Mol Biol **363**(1): 27-38.

Zenklusen, D., P. Vinciguerra, et al. (2001). "The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p." Mol Cell Biol **21**(13): 4219-32.

Zhang, Z., X. An, et al. (2006). "Nuclear localization of the Saccharomyces cerevisiae ribonucleotide reductase small subunit requires a karyopherin and a WD40 repeat protein." Proc Natl Acad Sci U S A **103**(5): 1422-7.

Ziegler, E. C. and S. Ghosh (2005). "Regulating inducible transcription through controlled localization." Sci STKE **2005**(284): re6.