Modular Organization and Combinatorial Energetics of Proline–Tyrosine Nuclear Localization Signals

Katherine E. Süel, Hongmei Gu, Yuh Min Chook*

Department of Pharmacology, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, United States of America

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/KX$_{2-5}$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 the 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 the 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.

Introduction

Karyopherinβ proteins (Kapβs; Importins/Exportins) mediate the majority of nucleocytoplasmic protein transport. There are 19 known Kapβs in human and 14 in yeast [1,2]. Kapβs bind substrates through nuclear localization or export signals (NLSs or NESs) and transport them through the nuclear pore complex, and Ran GTPase regulates Kapβ–substrate interactions [3–6]. 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 [4], and this pathway is conserved functionally from human to yeast [7,8]. 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) [9–11] confirmed a previously determined consensus sequence of K(K/R)X(K/R) [8,12]. 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 [9]. 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 [13–15], thorough thermodynamic analysis of this subclass is not

available, and its consensus is less well-defined (one example is KRX$_{10-12}$KRRK) than that for the monopartite NLS. Furthermore, a nonfunctional SV40 NLS mutant was rescued by a bipartite-like addition of a two-residue N-terminal basic cluster [9], 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-NLSs [16]. 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 the N-terminal motifs divides PY-NLSs into hydrophobic and basic subclasses (hPY- and bPY-NLSs). The former contains four consecutive predominantly hydro-

Author Summary

To travel between the cytoplasm and nucleus, proteins rely on a family of transport proteins known as the karyopherinβ family. Karyopherinβ2, the human version of a family member, recognizes cargo proteins containing a class of nuclear localization signal known as the PY-NLS. The yeast homolog of Karyopherinβ2, Kap104p, also recognizes PY-NLSs, indicating that this pathway has been conserved between evolutionarily distant species. We mutated residues in the PY-NLSs of two Kap104p cargo proteins and analyzed how tightly these mutants bound Kap104p. These experiments revealed three PY-NLS regions, or epitopes, that are important for binding Kap104p. Each epitope is composed of amino acids that vary between cargoes. The epitopes are energetically independent and bind Kap104p with varying strengths in different PY-NLSs, such that mutating the epitope of one PY-NLS may mistakenly direct cargo to the cytoplasm, while a similar mutation in a different PY-NLS has little effect on cargo localization. This flexible, energetically modular, and combinatorial architecture of PY-NLSs may confer higher tolerance to mutations, but it also allows greater sequence diversity, making prediction of new PY-NLSs difficult. The characteristics of PY-NLSs reported here will assist in the identification of new Kap104p cargoes. And the approach used may be applicable to other biological recognition pathways.

phobic residues, while the equivalent region in bPY-NLSs is enriched in basic residues.

Approximately 100 different human proteins have been identified as potential Kapβ2 substrates [16–25]. Table 1 summarizes previously reported validated and potential PY-NLSs. Although many of these potential substrates were predicted by bioinformatics [16] and still need experimental testing, more than 20 have been validated for Kapβ2 binding (Table 1) [16–25]. Comparison of in vivo and in vitro validated PY-NLSs shows large sequence diversity, which is reflected in weak consensus sequences [16]. Structures of five different Kapβ2-bound PY-NLSs also show substantial variability, with structurally diverse linkers separating the convergent consensus regions [16,26,27]. 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, the 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 described necessarily 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 individ-

Table 1. Summary of Validated and Potential PY-NLSs

Description	Proteins	PY-NLSs	References
Human, validated in vivo and in vitro [a,d,e]	PQBP-1 [b,g]	161 RDRGYDKAD***REEGKERR***HHRREELA**PY** 187	[16[e], 19[b,e], 68[d]]
	YBP1 [b,c,g]	176 SAPEGQAQQ***RRPYRRRR***FPPYYMRR**PY** 196	[16[c,e], 19[b], 69[d]]
	PABP2 [b]	280 FYSGFNS***RPRGRVYRGR***ARATSWYS**PY** 306	[70[d]]
	EWS [b,c]	630 GG***RRGGRGGPGKMDKGEHR***QERRDR**PY** 656	[16[c,e], 19[b], 71[d]]
	FUS [b,c,f]	500 GGD***RGGFGPGKMDSRGEHR***QD**RR**ER**PY** 526	[16[c,e], 19[b], 71[f]]
	TAF$_{II}$68 [f]	566 YGGD***RGGYGGKMGGRNDYR***NDQRNR**PY** 592	[71[f]]
	SAM68 [b,c]	414 DDWNGT***RPSLKAPPARPVK***GAYREH**PY** 440	[16[c,e], 19[b], 55[d], 72[d]]
	ETLE [b]	317 GQEEWTNS***RHKAPSARTAK***GVYRDQ**PY** 343	[73 [d]]
	hnRNP M [b]	38 GEGERPAQNE***KRKEKNIKR***GGNRFE**PY** 64	[16[e], 19, 26[d,e]]
	hnRNP A1 [b]	263 FGNYNNQSSN**FGPM**KGGNFGG**R**SSG**PY** 289	[17[e], 21, 24[d], 25[d]]
	hnRNP A2 [f]	293 NYNQQPSN**YGPM**KSGNFGGS**R**NMGG**PY** 319	[24]
	hnRNP A3 [f]	334 NYSGQQQSN**YGPM**KGGSFGG**R**SSGS**PY** 360	[74]
	hnRNP A0 [f]	264 S**YGPM**KSGGGGGGGSSWGG**R**SNSG**PY** 290	
	hnRNP D [b]	329 YGDYSNQQSG**YGKV**SRRGG**H**QNSYK**PY** 355	[22[b], 37[d]]
	JKTBP-1 [b]	394 ADYSGQQSTYG**KASR**GGGN**H**QNNYQ**PY** 420	[20[d,e]]
	TAP [b]	49 SSRLEEDDGD**VAMS**DAQDGPRV**R**YNP**PY** 75	[23, 75, 76]
	HuR [b]	212 HHQAQRFR**FSPM**GVDH**MSGLSGVNVP**G 238	[18[d], 19[e], 77[e]]
Human, potential PY-NLSs validated in vitro [a,e]	HMBA-ind [b,c,g]	141 WGQQQRQLG***KKKHRRR***PSKKK**R**HWK**PY** 167	[16[c,e], 19[b]]
	RB15B [c,g]	258 GLPKPWEE***RRKRR***SLSSDRG**R**TTHS**PY** 284	[16[c,e]]
	HCC1 [c,g]	73 ***RSRSKERRRSRSRSRDRR***FRG**R**YRS**PY** 99	[16[c,e]]
	SOX14 [c,g]	72 HPDY***KYRPRRKPKNLLKK***DRYVFP**L**PY** 98	[16[c,e]]
	Clk3 [c,g]	75 RSPSFGEDYYGPS***RSRHRRRSR***ERG**PY** 101	[16[c,e]]
	hnRNP F [b]	184 FKSSQEEVRS**YSDP**PLKFMSVQ**R**PG**PY** 210	[16[e], 19[b], 78[b,e]]
	hnRNP H1 [b,f]	184 FKSSRAEVRS**YSDP**PLKFMSVQ**R**PG**PY** 210	[19[b]]
	CPSF6 [b,c]	364 PPTNSGMPTSDS**RGPP**PTDPYG**R**PP**PY** 390	[16[c,e], 19[b]]
	Cyclin T1 [b,c,g]	661 VNMLHSLLSA**QGVQ**PTQPTAFEFV**R**PY 687	[16[c,e], 19[b]]
	WBS16 [c]	75 PSFVVPSSG**PGPR**AGARP**RRR**IQP**PY** 101	[16[c,e]]
Yeast, validated in vivo and in vitro [a,d,e]	Hrp1p [b]	506 RSGGNH***RRNGRGGR***GGYNRR**NN**GYH**PY** 532	[34–37][b,d,e]
	Nab2p [b]	214 AVG***KNRRGGRGGGNRGGR***NNNST**R**FN**PL** 241	[34–37][b,d,e]

[a]Basic motifs are underlined, italicized, and in bold font. Hydrophobic motifs are underlined and in bold font. 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 NLSs 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, A3, and A0 are highly homologous to hnRNP A1; NLSs of hnRNPs H1 and F are almost identical.
[g]Predicted to have a classical NLS within this region by WoLF PSORT [55].
doi:10.1371/journal.pbio.0060137.t001

ually weak physical rules that together were able to provide substantial limits in sequence space for reasonable predictions of new Kapβ2 substrates [16]. However, the currently predicted substrates are most likely only a fraction of all PY-NLS-containing proteins because narrow sequence patterns were used in the initial search to achieve optimal accuracy. In fact, the sequence patterns used [16] 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 also will 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 [28–33].

The yeast homolog of Kapβ2 is Kap104p (32% sequence identity) [34]. 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 [35–37]. Little sequence homology was detected between NLSs recognized by Kapβ2 and Kap104p. Furthermore, substrate recognition by the two karyopherins appears nonanalogous, as Kap104p does not recognize human substrate hnRNP A1 [35,37]. 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 to those of 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.

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, respec-

tively (Figure 1A) [35–38]. 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 of 0.72 and 0.63 for Hrp1p and Nab2p, respectively [39]) in the full-length proteins (Figure 1A). ^{506}**RSGGNHRRNGRGGR**^{519} of Hrp1p and ^{216}**KNRRGGRGGNRGGR**^{229} of Nab2p contain many basic residues, like basic N-terminal motifs in human Kapβ2 substrates hnRNP M, PQBP-1, and YB-1 (Table 1) [16]. Closer to the C termini, the Hrp1p ^{525}**RNNGYHPY**^{532} and the Nab2p ^{235}**RFNPL**^{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 1B). Although it was previously reported that Ran could not dissociate substrate from Kap104p [36], 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 similar to other characterized Kapβ-mediated nuclear import processes in human [3–6]. Thermodynamic parameters for Kap104p binding to Hrp1p and Nab2p NLSs were obtained by isothermal titration calorimetry (ITC) (Figure S1). Both NLSs bound Kap104p with high affinity (K_D of 32 nM for Hrp1p and 37 nM for Nab2p (Tables 2 and 3)), and extensive mutagenesis of NLSs is discussed below. Thus, on the basis of their sequence characteristics, high affinity for karyopherin, and dissociation by RanGTP, yeast NLSs recognized by Kap104p resemble PY-NLSs.

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 2A). All four human PY-NLSs interacted with Kapβ2 [16], but only bPY-NLSs from hnRNP M and FUS bound yeast Kap104p in GST pull-down assays (Figure 2B). Both yeast Hrp1p and Nab2p NLSs bound equally well to Kap104p and Kapβ2 (Figure S2).

Hrp1p and Nab2p are the only two known Kap104p substrates [34–36]. We needed to identify additional yeast sequences to test the preference of Kap104p for bPY-NLS. Because Nab2p has a C-terminal PL instead of PY motif, suggesting that PL motifs also may be present in other functional PY-NLSs, we used the program ScanProsite [40] and sequence patterns Φ_1-G/A/S-Φ_3-Φ_4-X$_{7-12}$-R/K/H-X$_{2-5}$-P-Y/L (where Φ_1 is a hydrophobic residue and Φ_3 and Φ_4 are hydrophobic residues or R or K) [16] to search for potential hPY-NLSs within *Saccharomyces cerevisiae* proteins in the UniProtKB/Swiss-Prot protein database [41]. 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 [16] to accommodate additional validated human bPY-NLSs and NLSs in Nab2p and Hrp1p (Table 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
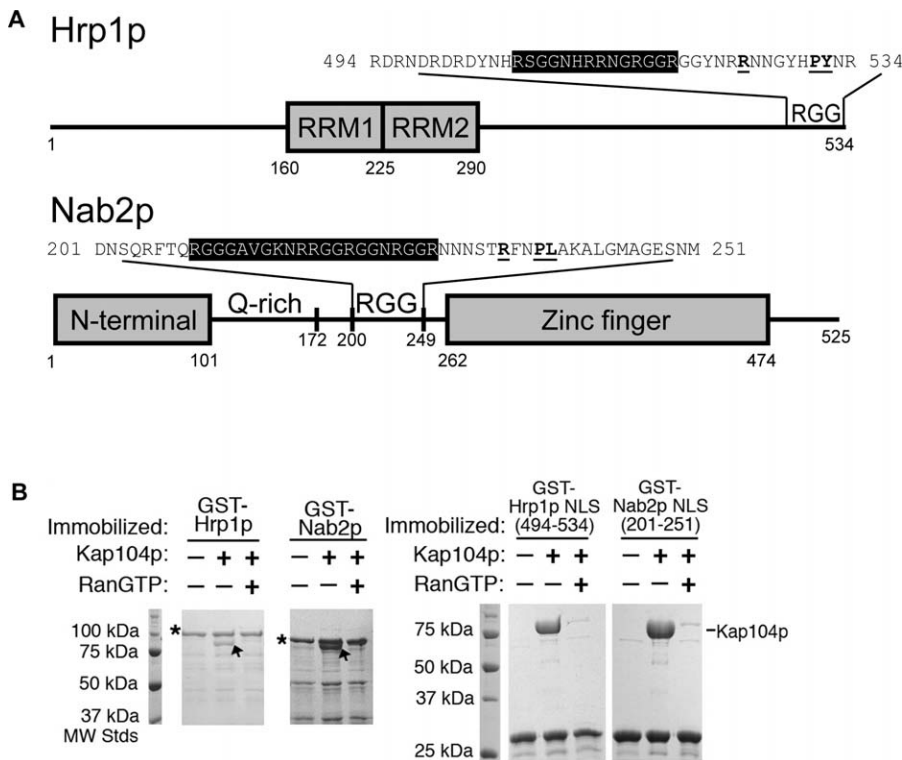
Figure 1. Interactions between Kap104p and Import Substrates Hrp1p and Nab2p

(A) Domain organization of Hrp1p and Nab2p. Domains are indicated by gray 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 $RX_{2-5}PY(L)$ 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 the absence of RanGTP. Bound proteins are stained with Coomassie blue.

doi:10.1371/journal.pbio.0060137.g001

potential yeast bPY-NLSs. The resulting lists were filtered for structural disorder [39] and overall basic character. Six hPY/L-containing fragments were tested, but none bound Kap104p (Figure 2A and 2C). However, 11 of 20 bPY/L-containing

fragments tested bound Kap104p and were dissociated by RanGTP (Figure 2A and 2D and Figure S3a and S3b). Two bPY/L-containing full-length substrates, Tfg2p and Rml2p, were tested, and both bound Kap104p and were dissociated by

Table 2. Summary of ITC Data for Kap104p Binding to Hrp1p Mutants

Hrp1p	Mutations	K_D [a] (nM)	ΔG [b] (kcal/mol)	ΔH (kcal/mol)	$T\Delta S$ [c] (kcal/mol/K)	$\Delta\Delta G$ [d] (kcal/mol)	$K_{D,mutant}/$ $K_{D,wild\ type}$
Wild type		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 ± 0.09
	[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 ± 0.02	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] $\Delta\Delta G = -RT \ln(K_{d,wild\ type}/K_{d,mutant})$.
n.d., not detectable; All binding experiments were performed 2–4 times (±standard deviation).
doi:10.1371/journal.pbio.0060137.t002

Table 3. Summary of ITC Data for Kap104p Binding to Nab2p Mutants

Nab2p	Mutations	K_D[a] (nM)	ΔG[b] (kcal/mol)	ΔH (kcal/mol)	$T\Delta S$[c] (kcal/mol/K)	$\Delta\Delta G$[d] (kcal/mol)	$K_{D,mutant}/$ $K_{D,wild\ type}$
Wild type		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	–0.012 ± 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 ± 81	–8.6 ± 0.07	–5.8 ± 1.3	2.8 ± 1.3	1.3 ± 0.07	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 ± 0.005	–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]$\Delta\Delta G = -RT \ln(K_{d,wild\ type}/K_{d,mutant})$.
All binding experiments were performed 2–4 times (±standard deviation).
doi:10.1371/journal.pbio.0060137.t003

RanGTP (Figure 2D). Of the 11 bPY/L-containing proteins in yeast that bound Kap104p (Table 1), 7 (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-NLSs.

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) [16,26]. 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 $RX_{2-5}PY$ motifs [26]. Correspondingly, both structures share many common Kapβ2 interface residues, especially those that contact the conserved C-terminal $RX_{2-5}PY$ motif (Figure 3A).

Approximately half of the Kapβ2–NLS interface residues are conserved in Kap104p. Interfaces with the $RX_{2-5}PY$ motifs (H8B–H12B) are mostly invariant, while differences occur at the structurally overlapping interfaces with the basic/hydrophobic N-terminal motifs (H15B–H17B) and at linker regions (H12B–H14B) (Figure 3A). Here, Kapβ2 residues I722, S723, N726, E734, T766, and I773 that contact the hnRNP A1 hydrophobic motif are replaced with T, P, I, L, S, and V, respectively, in yeast (Figure 3A and 3B) such that many

hydrophobic contacts with the FGPM N-terminal motif of hnRNP A1 are expected to be lost in yeast (detailed description in Text S1). In contrast, among Kapβ2 residues that contact basic side chains of bPY-NLSs, only E653 of Kapβ2 is different in yeast (Figure 3A and 3C), and several amino acids have been replaced by more electronegative amino acids in Kap104p (Figure 3C), further supporting bPY-NLS recognition in yeast.

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 3D). The B helices that line the interface are generally more conserved than the outer A helices. However, even in the former, sequence identities in H16B–H17B dipped significantly below 40% (Figure 3D). 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-NLSs. Results of these studies are discussed in Text S2 and shown in Figure S4A and S4B.

Distribution of Binding Energy along the Hrp1p NLS

We have performed scanning alanine mutagenesis covering residues 506–532 of the Hrp1p NLS (Figure 1A, Table 2, and Table S1). In the N-terminal region of the Hrp1p NLS, none of the four mutants ^{506}RSGG^{509}/AAAA, ^{512}RRNG^{515}/AAAA,
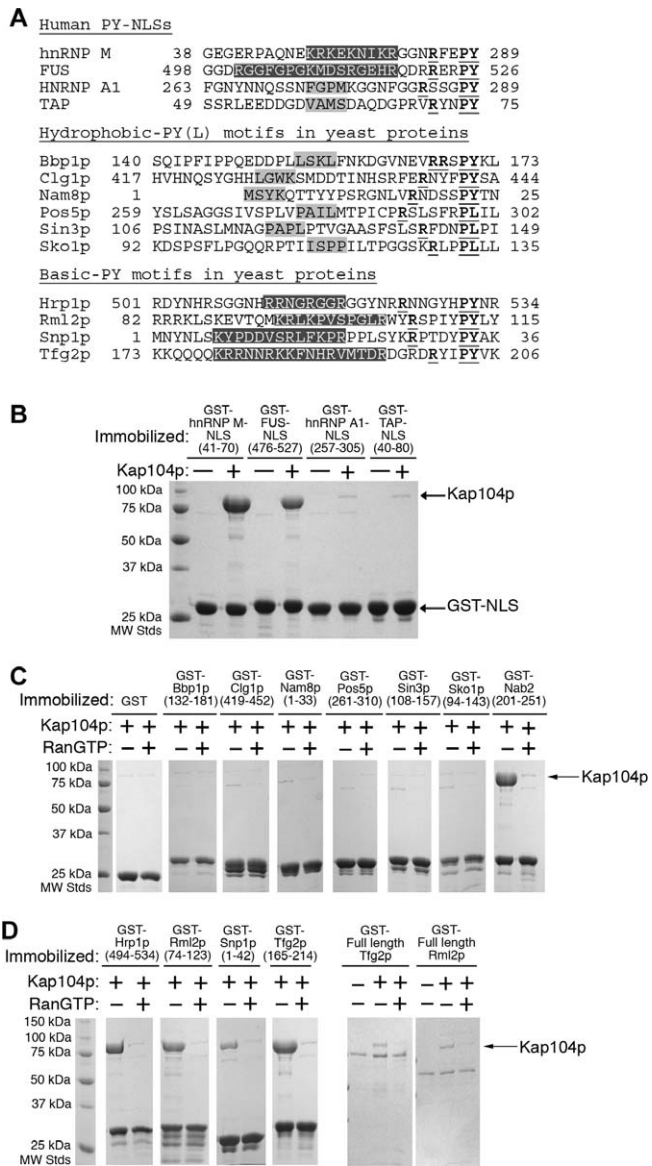
Figure 2. Kap104 Recognizes Basic PY-NLSs but Not Hydrophobic PY-NLSs

(A) Sequences of known PY-NLSs in human proteins and predicted hydrophobic-PY(L) (shaded in light gray) and basic PY motifs (shaded in dark gray) in yeast proteins. The $RX_{2-5}PY(L)$ 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 ~70 kDa 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 stained with Coomassie blue.

doi:10.1371/journal.pbio.0060137.g002

[516]RGG[518]/AAA, and [519]RGGYN[523]/AAAAA (Table 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 of the arginines

(R512, R513, R516, and R519) were mutated to alanines. This quadruple mutant decreased Kap104p binding by a marginal 5-fold (Figure 4A and Table 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 2), further suggesting that stereospecific interactions with arginine guanido groups are not important for Kap104p binding.

Kap104p binding was not affected significantly when both arginine residues, [524]RR[525], in the C-terminal $RX_{2-5}PY$ motif of the Hrp1p NLS were mutated to alanines ($K_{D,mutant}/K_{D,wild\ type} = 1.7$; Figure 4A and Table 2). In contrast, the C-terminal [531]PY[532]/AA mutation abolished detectable Kap104p binding (Table 2). The enthalpies of binding for all of the PY-NLSs that we have measured by ITC are similar, and the weakest measurable K_D in this series was 10 μM [26]. Therefore, we assume that the affinity of the Hrp1p [531]PY[532]/AA mutant is likely weaker than 10 μM and its $K_{D,mutant}/K_{D,wild\ type} > 200$ (Figure 4A). Thus, the Hrp1p NLS contains one strong binding hotspot at its PY motif, similar to the single significant hotspot at the C-terminal PY motif of the human substrate hnRNP M ($K_{D,mutantPY/AA}/K_{D,wild\ type} = 500$ for the hnRNP M NLS) [26]. Interestingly, we also located a modest binding hotspot at residue Y529 ($K_{D,mutant}/K_{D,wild\ type} = 4$ for Y529A; Figure 4A and Table 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 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 the Nab2p NLS (Figure 1A, Table 3, and Table S1). Binding energy along the Nab2p NLS appears quite distributed compared to that of the Hrp1p NLS, with no single binding hotspot that stands out above others (Figure 4B and Table 3). In its basic N-terminal region, [216]KNRR[219], [222]RGG[224], and [226]RGGRN[230] each were mutated to alanines, but only [216]KNRR[219]/AAAA showed a small 3-fold decrease in Kap104p affinity (Table 3). None of the single mutants K216A, R218A, R219A, R222A, R226A, or R229A decreased Kap104p binding (Table S1), and simultaneous mutation of all of 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 4B and Table 3), suggesting that the collective basic character of this region contributes significantly to the total binding energy of the NLS. Comparison of single arginine to alanine mutants ($K_{D,mutant}/K_{D,wild\ type} \approx 1.0$) to the pentamutant R218, R219, R222, R226, R229/AAAAA ($K_{D,mutant}/K_{D,wild\ type} = 60.8$) 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 5-fold (Figure 4B and Table 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 [16,26]. We also mutated R235 to lysine and histidine, but neither mutant
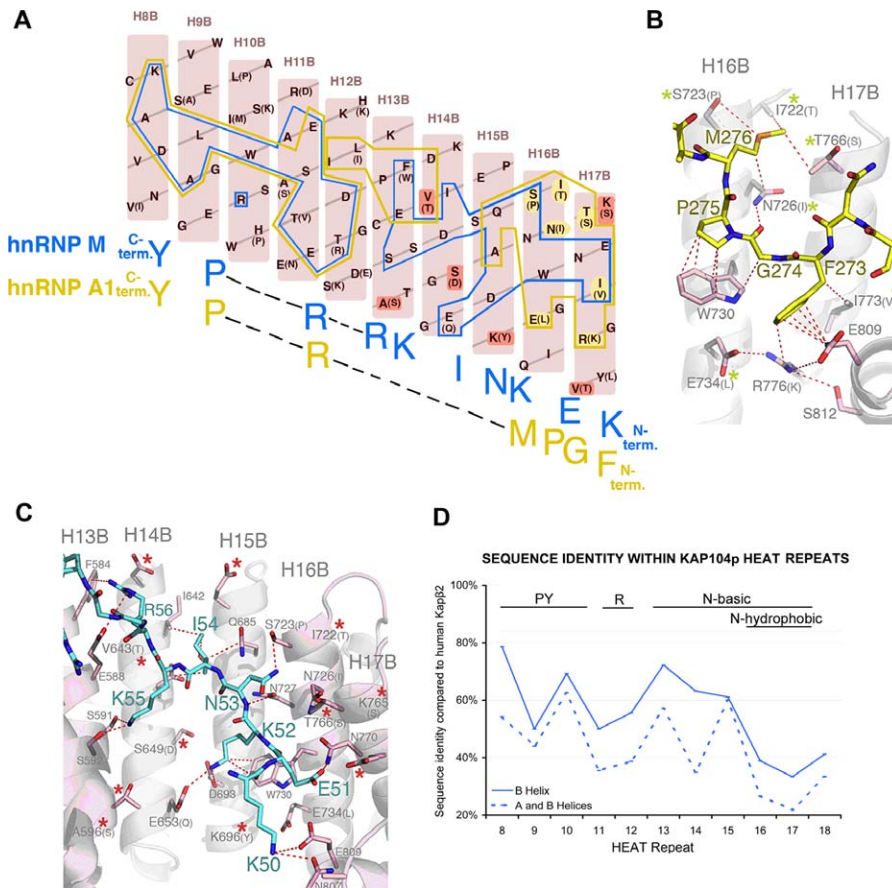
Figure 3. Kapβ2–NLS Structures and Kap104p Specificity

(A) Schematic representation of the Kapβ2–NLS interface showing B helices of Kapβ2 H8–H17 (pink). Residues that are different in Kap104p are in parentheses. Kapβ2 residues that contact the hnRNP A1 NLS (hPY-NLS) and the 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) (PDB ID 2H4M) drawn with PYMOL [67]. 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) (PDB ID 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.
doi:10.1371/journal.pbio.0060137.g003

affected Kap104p binding significantly ($K_{D,mutant}/K_{D,wild type}$ are 1.0 and 1.7, respectively; Table 3). The C-terminal $^{238}PL^{239}$/AA mutation in the Nab2p NLS decreased Kap104p binding by 10-fold (Figure 4B and Table 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 a slightly higher affinity at a 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 the 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; $K_{D,mutant}/K_{D,wild type}$ = 11.1; Table 3) show almost perfect additivity when compared to a single R222A mutant (did not affect Kap104p binding; Table S1) and double mutant $^{238}PL^{239}$/AA ($K_{D,mutant}/K_{D,wild type}$ = 10.2; Table 3). A second Nab2p triple mutant R235A, $^{238}PL^{239}$/AA (K_D = 544 nM; $K_{D,mutant}/K_{D,wild type}$ = 14.7; Table 3) also was compared to a

single R235A mutant ($K_{D,mutant}/K_{D,wild type}$ = 5.5; Table S1) and double $^{238}PL^{239}$/AA mutant ($K_{D,mutant}/K_{D,wild type}$ = 10.2; Table 3). Strict additivity between the R and the PL sites would give a calculated $K_{D,mutant}/K_{D,wild type}$ value of 56.1 for the triple mutant. Thus, the experimental $K_{D,mutant}/K_{D,wild type}$ value of 14.7 for the triple mutant indicated 3.8-fold cooperativity between the two epitopes. Similarly, Hrp1p triple mutant R512A, $^{524}RR^{525}$/AA and double mutant R512A, Y529A showed cooperativity of approximately 1.4- and 2-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 (>60-fold cooperativity).

We also located a new binding hotspot at F236 in Nab2p ($K_{D,mutant}/K_{D,wild type}$ = 8 for F236A; Figure 4B and Table 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
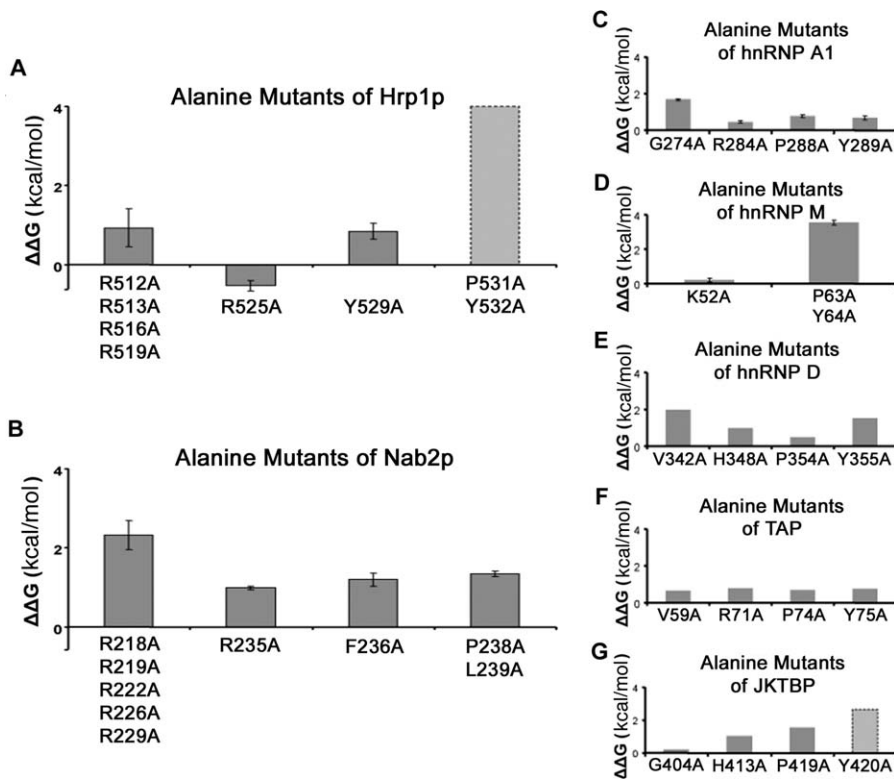
Figure 4. Mutagenic Analyses of the Hrp1p NLS, Nab2p NLS, and Human PY-NLSs

(A and B) Loss of Kap104p binding energy in alanine mutants of (A) Hrp1p and (B) Nab2p ($\Delta\Delta G = -RT \ln(K_{D,\text{wild type}}/K_{D,\text{mutant}})$).

(C–G) Loss of Kapβ2 binding energy in alanine mutants of PY-NLSs from (C) hnRNP A1, (D) hnRNP M, (E) hnRNP D, (F) TAP, and (G) JKTBP ($\Delta\Delta G = -RT \ln(K_{D,\text{wild type}}/K_{D,\text{mutant}})$). $K_{D,\text{wild type}}$ and $K_{D,\text{mutant}}$ values for hnRNP A1 were obtained from [16], hnRNP M from [26], hnRNP D, TAP, and JKTBP from [27]. K_D values for Hrp1p, Nab2p, and hnRNPs A1 and M were obtained by ITC whereas those for hnRNP D, TAP, and JKTBP were obtained by surface plasmon resonance.

doi:10.1371/journal.pbio.0060137.g004

the Hrp1p NLS Y529L mutant, the F236L mutation in Nab2p did not affect Kap104p binding (Table 3). Aromatic or hydrophobic residues occur at this position in many human PY-NLSs, including hnRNPs M, D, and F, JKTBP, TAP, HMBA-inducible protein, PABP2, PQBP-1, RB15B, and WBS-16 [16,22,23,27]. Aromatic side chains at this position overlap in the crystal structures of Kapβ2 bound to the NLSs of hnRNPs M and D and TAP [26,27]. The F61 of the hnRNP M NLS, Y352 of the hnRNP D NLS, and Y72 of the TAP NLS 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 a 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 the 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 described roughly 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 4A and 4B). Similarly, in previously characterized PY-NLSs of hnRNPs A1 and D, TAP, and JKTBP [16,27], energetic distributions at the three epitopes also are quite varied, with rough patterns of strong–weak–weak, strong–medium–medium, weak–weak–weak, and weak–medium–strong, respectively (Figure 4C–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 $RX_{2-5}PY$ motif is medium to weakly energetically significant.

Degeneracy of Tyrosine in the C-Terminal PY Motif Of the more than 20 sequences that bind Kapβ2 and Kap104p (Table 1) [16], 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 5A and Table S2). Only Y532F, Y532H, and Y532M showed measurable Kap104p binding by ITC. Y532F best resembles the wild type, with only a 4-fold decrease in Kap104p affinity. Both Y532H and Y532M in Hrp1p bound significantly weaker with K_D values of 1 and 2 μM, respectively.

We also mutated L239 in the Nab2p PL motif to the other 19 amino acids (Figure 5B and Table S3). Binding energy along the Nab2p NLS is distributed very evenly compared to that of the Hr1p1p NLS with the Nab2p $^{238}PL^{239}$/AA mutation

A

Mutations in the PY motif of Hrp1p

B

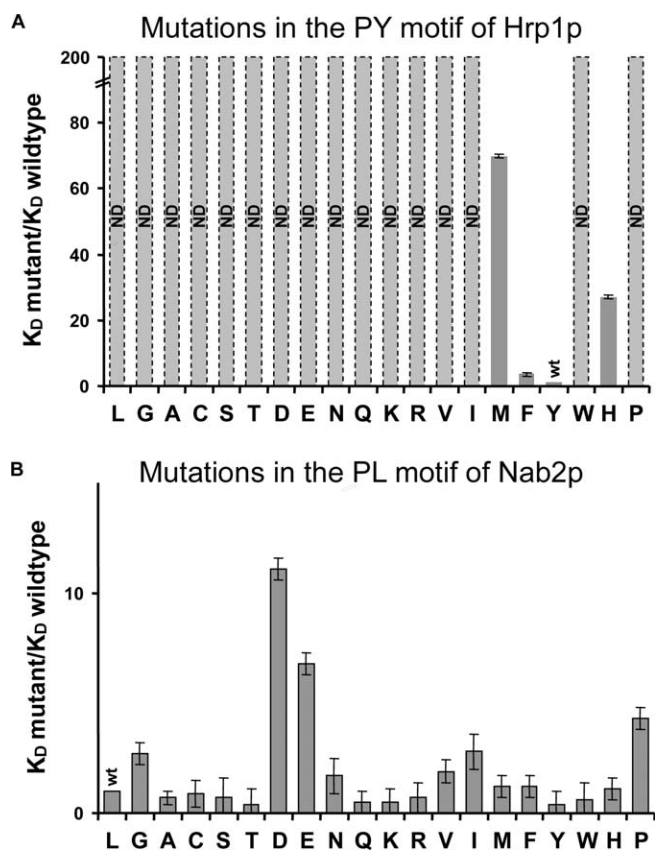**Mutations in the PL motif of Nab2p**

Figure 5. 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, and 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.

doi:10.1371/journal.pbio.0060137.g005

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 because only L239D and L239E showed significant affinity decreases of 11- and 7-fold, respectively. L239G, L239I, and L230P showed a 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 Hr1p1 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 1). It appears that if the PY site is energetically very significant, such as that 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 that in Nab2p and hnRNP A1 [16], the distribution of allowed amino acids in the terminal position is likely much wider, with only 2–5 residues disallowed.

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 wild-type and mutant proteins in yeast. Wild-type Hrp1–GFP and Nab2p–GFP are localized in the nucleus as has been reported previously (Figure 6A–D) [34–36]. 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 6A and 6C). 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 5-fold, also is mislocalized (Figure 6A and 6C). Xu and Henry have shown previously that substitutions of R516 and R519 with glutamines mislocalized Hrp1p, but proteins with lysine substitutions are properly localized [38,42]. 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 (pentamutant 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 6B and 6D). 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 affinity for Kap104p [38,43]. Combined mutations of both the N- and the C-terminal motifs resulted in diffuse localization of the fusion protein, consistent with further affinity reduction for Kap104p (Table 3 and Figure 6B and 6D). 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 [16,26]. 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. On the basis of all of this and the thermodynamic data from five human PY-NLSs [16,26,27], we propose the following physical properties that govern the affinity of PY-NLS recognition by Kapβ2:

1. PY-NLSs Contain at Least Three Energetically Significant Binding Epitopes

Structures of PY-NLSs from hnRNPs A1, M, and 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
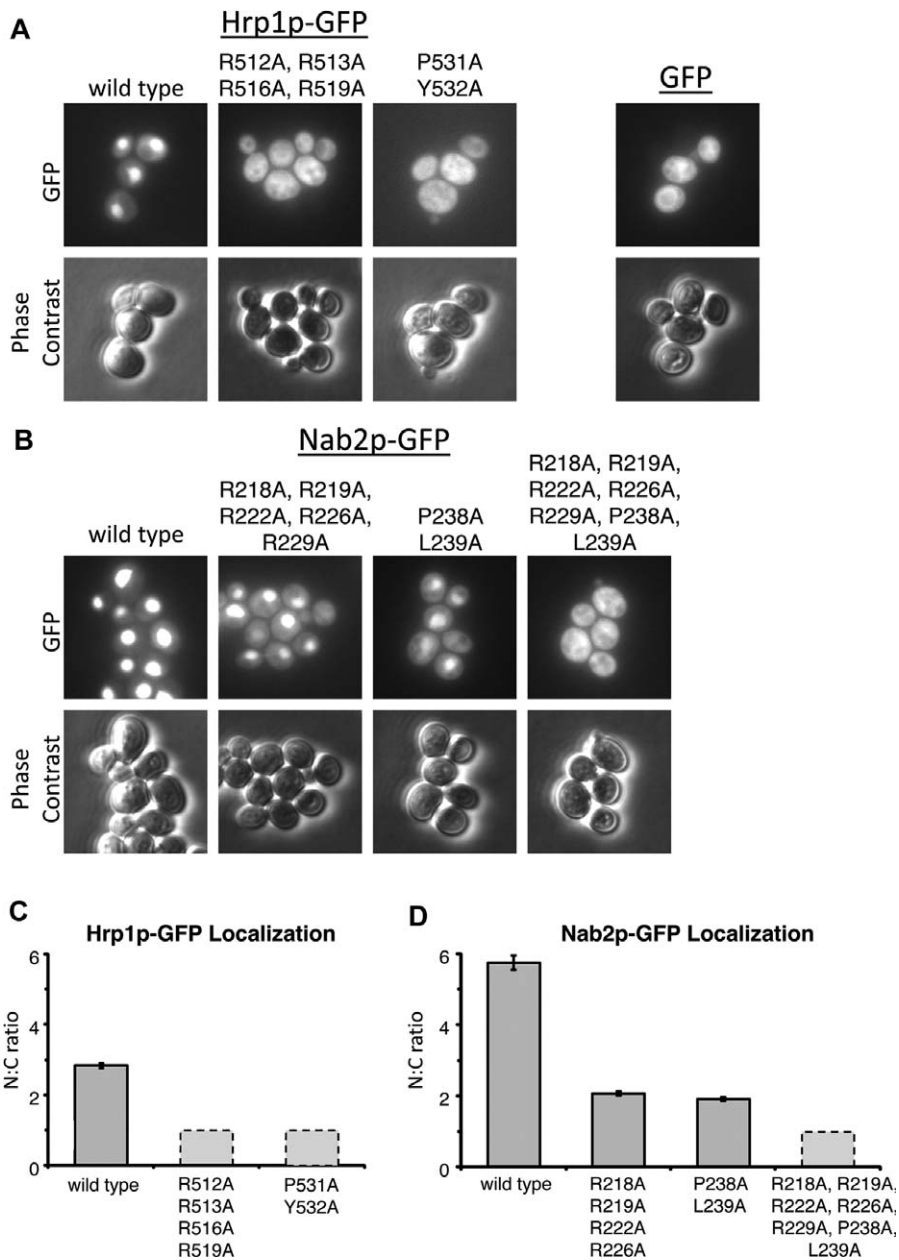
Figure 6. Hrp1p and Nab2p Mutants Are Mislocalized In Vivo

(A) *S. cerevisiae* cells expressing either wild-type or mutant full-length Hrp1p–GFP fusion proteins were analyzed by fluorescence microscopy and phase contrast. GFP is displayed in the same fluorescence scale in each panel.

(B) Cells expressing either wild-type or mutant full-length Nab2p–GFP fusion proteins were analyzed as in (A).

(C and D) Mean pixel values were used to determine the nuclear/cytoplasmic (N/C) ratio of fluorescence intensity for either (C) Hrp1p–GFP or (D) Nab2p–GFP fusion proteins (±standard error of the mean). Dashed lines indicate an estimated N/C ratio of 1:1 due to the diffuse nuclear and cytoplasmic localization of the fusion protein.

doi:10.1371/journal.pbio.0060137.g006

of the C-terminal $RX_{2-5}PY$ sequence motif, and (3) the PY of the C-terminal $RX_{2-5}PY$ motif [16,26,27]. 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 of 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 the hnRNP A1 NLS and

the equivalent region of the hnRNP D NLS that contains both hydrophobic and basic residues are also energetically significant, with mutations decreasing binding energy by ~2 kcal/mol [26].

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 [16,26]. 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 4B, 4E, and 4G).

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 1.3–4 kcal/mol in Hrp1p, Nab2p, hnRNPs M and D, and JKTBP. However, exceptions are seen in hnRNP A1 and TAP, where PY mutations decreased binding modestly by only ~0.7 kcal/mol.

Because 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 [44–46]. 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 4 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 1 [16,26] 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 the exact positions of basic groups may not be important (Tables 2 and 3 and Table S1). 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 1) shows that arginine is most prevalent in this position, although histidines are found in this position in hnRNP D, JKTBP, and HuR and lysines in potential yeast NLSs of Naf1p, Sbp1p, Arp8p, and Ste20p (Figure S3A). 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 1) has a PG dipeptide, and 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, then the terminal site tends to be phenylalanine, histidine, and methionine. If the dipeptide motif is fairly silent energetically, then 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 2 and Table S1). Mutations within the N-terminal basic region of the hnRNP M NLS also show cooperativity, in a similar regime, of ~40-fold [26]. In contrast, seven examples of simultaneous mutations between different linear epitopes in Hrp1p, Nab2p (Tables 2 and 3), and hnRNPs A1 and M [16,26] 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 [47]. 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 [48]. 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 [16,26,27]. 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 4A and 4B). Similarly, PY in hnRNP A1 contributes only weakly to Kapβ2 binding, while PY motifs in hnRNP M and Hrp1p are the sole binding hotspots in the NLSs (Figure 4A, 4C, and 4D). We previously had taken advantage of the energetic variability of PY-NLS epitopes by harnessing the avidity effect of the NLS hotspot at epitope 1 of hnRNP A1 fused to the 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 [26]. 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 [26] 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 "multifaceted" interactions, where different ligands make energetically significant interactions with different subsets of interface residues, were recently studied in a theoretical context [49] and also suggested to be more tolerant to mutations and are therefore quite evolvable.

Multiple functions have been identified in fact in several PY-NLSs. In Nab2p, the RGG region that overlaps NLS epitope 1 is a putative RNA binding region [50]. The PY-NLSs in Nab2p, Hrp1p, EWS, and FUS interact with and are methylated by arginine methyltransferases [43,51–54]. Phosphorylation sites also have evolved within PY-NLSs to regulate nucleocytoplasmic localization. Serine phosphorylation in the hnRNP A2 NLS and tyrosine phosphorylation in the SAM68 NLS [55] both alter subcellular localization of the proteins. A PY-NLS also may 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 a potential classical NLS [56] in the N-terminal basic motifs of eight human bPY-NLSs in Table 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, then the sequence tolerance for this motif is small, and sequence content of the other two epitopes will likely not matter. Thus, this subset of the PY-NLSs should be identified easily 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 allow 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-NLS 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 likely will vary depending on cellular concentrations of substrates but should not be unbounded [57]. For example, a designed peptide with a K_D of 100 pM binds Kapβ2 too tightly for in vivo nuclear import [26], 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 predictions of binding energies for protein–protein interactions are even further behind [58]. 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 [59] 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 [16], and the method was refined iteratively. Binding energy calculation remains problematic. We expect that 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 [60–62].

Conclusions

PY-NLSs are very diverse in sequence and structure and thus cannot be described sufficiently 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 [16]. Here, we examined the energetic organization of PY-NLSs 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 describes how functional determinants of the 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.

Materials and Methods

Plasmids and strains. The *Kap104p* gene (gift from J. Aitchison) was subcloned into the pGEX-Tev vector [34]. Yeast substrate genes were obtained by PCR from a *S. cerevisiae* genomic DNA library (Novagen) and subcloned into the BamHI and NotI sites of the pGEX-Tev and/or pMAL-Tev vectors [63,64]. Site-directed mutagenesis of Nab2p 201–251 and Hrp1p 494–534 were performed using the QuikChange method (Stratagene) and confirmed by nucleotide sequencing.

Full-length Nab2p and Hrp1p wild-type and mutant genes were subcloned into the SpeI and SmaI sites of a modified pRS415 (*CEN6*, *ARS*, *LEU2*, and *AP^R*) shuttle vector containing a C-terminal GFP gene [65].

Cell culture and microscopy. BY4741 (*MAT**a** his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) cells harboring pRS415 plasmids were grown at 30 °C in SC-Leu media to mid-logarithmic phase [66]. Cells were transferred to a 1.5% low-melting-point agarose pad made with SC-leu in a coverslip bottom Wilco dish. Cells were observed on an Olympus IX-81 inverted microscope (60× objective), and images were acquired with a Hamamatsu ORCA-ER camera. All images were analyzed in Image-Pro Plus software (Media Cybernetics). To obtain the 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.

Protein expression and purification. The GST–Kap104p protein was expressed in *Escherichia coli* Rosetta(DE3)pLysS cells (Novagen). Cells were lysed using an EmulsiFlex-C5 homogenizer (Avestin). The supernatant was applied to glutathione sepharose (GE Healthcare) and washed extensively with Tris buffer (50 mM Tris, pH 7.5, 100 mM NaCl, 1 mM EDTA, 2 mM DTT, and 20% glycerol). GST–Kap104p was eluted with Tris Buffer plus 20 mM glutathione, pH 8.1. The GST tag was cleaved using 0.5 ml of TEV protease in a total volume of 10 ml and separated from Kap104p using an anion exchange column (GE Healthcare). Kap104p was purified further by gel filtration chromatography in TB buffer (20 mM HEPES, pH 7.3, 110 mM KAc, 2 mM MgAc, 1 mM EGTA, 2 mM DTT, and 20% glycerol).

Yeast substrates and NLSs were expressed in *E. coli* BL21(DE3). The maltose-binding protein (MBP) NLSs were lysed as above and purified by affinity chromatography using amylose resin (New England Biolabs). After extensive washing with Tris buffer, protein was eluted with Tris buffer plus 10 mM maltose. The protein was purified further by cation exchange chromatography.

The GST substrates were lysed by sonication and immobilized on glutathione sepharose. The protein was washed with TB buffer and left on the beads for binding assays. Human substrates were expressed and purified as previously reported [16].

Bioinformatics search for new Kap104p substrates. Potential Kap104p substrates were identified as described in Lee et al. [16]. 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 were

used in ScanProsite [40] to screen *S. cerevisiae* proteins in the UniProtKB/Swiss-Prot database [41].

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). Proteins were dialyzed against buffer containing 20 mM Tris, pH 7.5, 100 mM NaCl, 2 mM β-mercaptoethanol, and 10% glycerol. The 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-site binding model using MicroCal Origin software (version 7.0).

Supporting Information

Figure S1. Isothermal Titration Calorimetry Measurements of Kap104p Binding to (A) MBP–Hrp1p NLS and (B) MBP–Nab2p NLS.

Found at doi:10.1371/journal.pbio.0060137.sg001 (89 KB PDF).

Figure S2. Binding Assays of Kapβ2 with Immobilized Nab2p NLS and Hrp1p NLS in the Presence and the Absence of RanGTP

Bound proteins are stained with Coomassie blue.

Found at doi:10.1371/journal.pbio.0060137.sg002 (70 KB PDF).

Figure S3. Kap104 Recognizes Basic PL-NLSs

(A) The sequences of predicted basic PL motifs in yeast proteins. Basic motifs are shaded in dark gray, and the $RX_{2–5}PY(L)$ motif is in bold and underlined.
(B) Experimental validation 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.

Found at doi:10.1371/journal.pbio.0060137.sg003 (180 KB PDF).

Figure S4. 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 (gray asterisks). Contact residues that differ in *Homo sapiens* (yellow) and *S. cerevisiae* (gray) 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.

Found at doi:10.1371/journal.pbio.0060137.sg004 (187 KB PDF).

Table S1. Summary of ITC Data for Kap104p Binding to Additional Hrp1p and Nab2p Mutants

Found at doi:10.1371/journal.pbio.0060137.st001 (91 KB PDF).

Table S2. Summary of ITC Data for Mutations of the PY Motif of Hrp1p

Found at doi:10.1371/journal.pbio.0060137.st002 (78 KB PDF).

Table S3. Summary of ITC Data for Mutations of the PL Motif of Nab2p

Found at doi:10.1371/journal.pbio.0060137.st003 (82 KB PDF).

Text S1. Differences in Interface Residues between Kap104p and Kapβ2

Found at doi:10.1371/journal.pbio.0060137.sd001 (10 KB PDF).

Text S2. Prediction of Kapβ2 Homologs That Recognize hPY-NLSs

Found at doi:10.1371/journal.pbio.0060137.sd002 (79 KB PDF).

Accession Numbers

Accession numbers for genes mentioned in this paper from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) are: *Bbp1p* (855820), *Clg1p* (852657), *hnRNP A1* (3178), *Hrp1p* (853997), *Kap104p* (852305), *Kapβ2* (3842), *Nab2p* (852755), *Nam8p* (856486), *Pos5p* (855913), *Rml2p* (856660), *Sin3p* (854158), *Sko1p* (855554), *Snp1p* (854749), *TAP* (10482), and *Tfg2p* (852888).

Acknowledgments

References

1. Mosammaparast N, Pemberton LF (2004) Karyopherins: from nuclear-transport mediators to nuclear-function regulators. Trends Cell Biol 14: 547–556.
2. Fried H, Kutay U (2003) Nucleocytoplasmic transport: taking an inventory. Cell Mol Life Sci 60: 1659–1688.
3. Chook YM, Blobel G (2001) Karyopherins and nuclear import. Curr Opin Struct Biol 11: 703–715.
4. Conti E, Izaurralde E (2001) Nucleocytoplasmic transport enters the atomic age. Curr Opin Cell Biol 13: 310–319.
5. Gorlich D, Kutay U (1999) Transport between the cell nucleus and the cytoplasm. Annu Rev Cell Dev Biol 15: 607–660.
6. Weis K (2003) Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. Cell 112: 441–451.
7. Enenkel C, Blobel G, Rexach M (1995) Identification of a yeast karyopherin heterodimer that targets import substrate to mammalian nuclear pore complexes. J Biol Chem 270: 16499–16502.
8. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J (1998) Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α. Cell 94: 193–204.
9. Hodel MR, Corbett AH, Hodel AE (2001) Dissection of a nuclear localization signal. J Biol Chem 276: 1317–1325.
10. Catimel B, Teh T, Fontes MR, Jennings IG, Jans DA, et al. (2001) Biophysical characterization of interactions involving importin-alpha during nuclear import. J Biol Chem 276: 34189–34198.
11. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, et al. (2007) Classical nuclear localization signals: definition, function, and interaction with importin α. J Biol Chem 282: 5101–5105.
12. Kalderon D, Richardson WD, Markham AF, Smith AE (1984) Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature 311: 33–38.
13. Fontes MR, Teh T, Jans D, Brinkworth RI, Kobe B (2003) Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-α. J Biol Chem 278: 27981–27987.
14. Fontes MR, Teh T, Kobe B (2000) Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-α. J Mol Biol 297: 1183–1194.
15. Conti E, Kuriyan J (2000) Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin α. Structure Fold Des 8: 329–338.
16. Lee BJ, Cansizoglu AE, Suel KE, Louis TH, Zhang Z, et al. (2006) Rules for nuclear localization sequence recognition by karyopherin β2. Cell 126: 543–558.
17. Bonifaci N, Moroianu J, Radu A, Blobel G (1997) Karyopherin β2 mediates nuclear import of a mRNA binding protein. Proc Natl Acad Sci U S A 94: 5055–5060.
18. Fan XC, Steitz JA (1998) HNS, a nuclear-cytoplasmic shuttling sequence in HuR. Proc Natl Acad Sci U S A 95: 15293–15298.
19. Guttinger S, Muhlhausser P, Koller-Eichhorn R, Brennecke J, Kutay U (2004) Transportin2 functions as importin and mediates nuclear import of HuR. Proc Natl Acad Sci U S A 101: 2918–2923.
20. Kawamura H, Tomozoe Y, Akagi T, Kamei D, Ochiai M, 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: 2732–2739.
21. Pollard VW, Michael WM, Nakielny S, Siomi MC, Wang F, et al. (1996) A novel receptor-mediated nuclear protein import pathway. Cell 86: 985–994.
22. Suzuki M, Iijima M, Nishimura A, Tomozoe Y, Kamei D, et al. (2005) Two separate regions essential for nuclear import of the hnRNP D nucleocytoplasmic shuttling sequence. FEBS J 272: 3975–3987.
23. Truant R, Kang Y, Cullen BR (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: 32167–32171.
24. Siomi H, Dreyfuss G (1995) A nuclear localization domain in the hnRNP A1 protein. J Cell Biol 129: 551–560.
25. Weighardt F, Biamonti G, Riva S (1995) Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. J Cell Sci 108: 545–555.
26. Cansizoglu AE, Lee BJ, Zhang ZC, Fontoura BM, Chook YM (2007) Structure-based design of a pathway-specific nuclear import inhibitor. Nat Struct Mol Biol 14: 452–454.
27. Imasaki T, Shimizu T, Hashimoto H, Hidaka Y, Kose S, et al. (2007) Structural basis for substrate recognition and dissociation by human transportin 1. Mol Cell 28: 57–67.
28. Rapaport D (2003) Finding the right organelle. Targeting signals in mitochondrial outer-membrane proteins. EMBO Rep 4: 948–952.
29. Van Ael E, Fransen M (2006) Targeting signals in peroxisomal membrane proteins. Biochim Biophys Acta 1763: 1629–1638.
30. Brocard C, Hartig A (2006) Peroxisome targeting signal 1: is it really a simple tripeptide? Biochim Biophys Acta 1763: 1565–1573.
31. Swanton E, High S (2006) ER targeting signals: more than meets the eye? Cell 127: 877–879.
32. Hegde RS, Bernstein HD (2006) The surprising complexity of signal sequences. Trends Biochem Sci 31: 563–571.
33. Mancias JD, Goldberg J (2005) Exiting the endoplasmic reticulum. Traffic 6: 278–285.
34. Aitchison JD, Blobel G, Rout MP (1996) Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins. Science 274: 624–627.
35. Truant R, Fridell RA, Benson RE, Bogerd H, Cullen BR (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: 1449–1458.
36. Lee DC, Aitchison JD (1999) Kap104p-mediated nuclear import. Nuclear localization signals in mRNA-binding proteins and the role of Ran and RNA. J Biol Chem 274: 29031–29037.
37. Siomi MC, Fromont M, Rain JC, Wan L, Wang F, et al. (1998) Functional conservation of the transportin nuclear import pathway in divergent organisms. Mol Cell Biol 18: 4141–4148.
38. Marfatia KA, Crafton EB, Green DM, Corbett AH (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: 6731–6740.
39. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, et al. (2003) Protein disorder prediction: implications for structural proteomics. Structure (Cambridge) 11: 1453–1459.
40. Gattiker A, Gasteiger E, Bairoch A (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. Appl Bioinformatics 1: 107–108.
41. Bairoch A, Boeckmann B, Ferro S E. G (2004) Swiss-Prot: juggling between evolution and stability. Brief Bioinform 5: 39–55.
42. Xu C, Henry MF (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: 10742–10756.
43. Green DM, Marfatia KA, Crafton EB, Zhang X, Cheng X, 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: 7752–7760.
44. Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, et al. (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol 3: e405.
45. Neduva V, Russell RB (2005) Linear motifs: evolutionary interaction switches. FEBS Lett 579: 3342–3345.
46. Puntervoll P, Linding R, Gemund C, Chabanis-Davidson S, Mattingsdal M, et al. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. Nucleic Acids Res 31: 3625–3630.
47. Walsh ST, Sylvester JE, Kossiakoff AA (2004) The high- and low-affinity receptor binding sites of growth hormone are allosterically coupled. Proc Natl Acad Sci U S A 101: 17078–17083.
48. Mossessova E, Bickford LC, Goldberg J (2003) SNARE selectivity of the COPII coat. Cell 114: 483–495.
49. Humphris EL, Kortemme T (2007) Design of multi-specificity in protein interfaces. PLoS Comput Biol 3: e164.
50. Anderson JT, Wilson SM, Datar KV, Swanson MS (1993) NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. Mol Cell Biol 13: 2730–2741.
51. Henry MF, Silver PA (1996) A novel methyltransferase (Hmt1p) modifies poly(A)+-RNA-binding proteins. Mol Cell Biol 16: 3668–3678.
52. Shen EC, Henry MF, Weiss VH, Valentini SR, Silver PA, et al. (1998) Arginine methylation facilitates the nuclear export of hnRNP proteins. Genes Dev 12: 679–691.
53. Belyanskaya LL, Gehrig PM, Gehring H (2001) Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein. J Biol Chem 276: 18681–18687.
54. Rappsilber J, Friesen WJ, Paushkin S, Dreyfuss G, Mann M (2003) Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode. Anal Chem 75: 3107–3114.
55. Lukong KE, Larocque D, Tyner AL, Richard S (2005) Tyrosine phosphorylation of sam68 by breast tumor kinase regulates intranuclear localization and cell cycle progression. J Biol Chem 280: 38639–38647.

56. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, et al. (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35: W585–W587.

57. Hodel AE, Harreman MT, Pulliam KF, Harben ME, Holmes JS, et al. (2006) Nuclear localization signal receptor affinity correlates with in vivo localization in Saccharomyces cerevisiae. J Biol Chem 281: 23545–23556.

58. Gilson MK, Zhou HX (2007) Calculation of protein–ligand binding affinities. Annu Rev Biophys Biomol Struct 36: 21–42.

59. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein–protein complexes. Proc Natl Acad Sci U S A 99: 14116–14121.

60. Kortemme T, Baker D (2004) Computational design of protein–protein interactions. Curr Opin Chem Biol 8: 91–97.

61. Baker D (2006) Prediction and design of macromolecular structures and interactions. Philos Trans R Soc Lond B Biol Sci 361: 459–463.

62. Nayeem A, Sitkoff D, Krystek S Jr. (2006) A comparative study of available software for high-accuracy homology modeling: from sequence alignments to structural models. Protein Sci 15: 808–824.

63. Chook YM, Blobel G (1999) Structure of the nuclear transport complex karyopherin-β2-Ran x GppNHp. Nature 399: 230–237.

64. Chook YM, Jung A, Rosen MK, Blobel G (2002) Uncoupling Kapβ2 substrate dissociation and ran binding. Biochemistry 41: 6955–6966.

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

66. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, 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: 115–132.

67. DeLano WL (2002) Pymol. San Carlos (California): DeLano Scientific.

68. Waragai M, Lammers CH, Takeuchi S, Imafuku I, Udagawa Y, 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: 977–987.

69. Bader AG, Vogt PK (2005) Inhibition of protein synthesis by Y box-binding protein 1 blocks oncogenic cell transformation. Mol Cell Biol 25: 2095–2106.

70. Calado A, Kutay U, Kuhn U, Wahle E, Carmo-Fonseca M (2000) Deciphering the cellular pathway for transport of poly(A)-binding protein II. RNA 6: 245–256.

71. Zakaryan RP, Gehring H (2006) Identification and characterization of the nuclear localization/retention signal in the EWS proto-oncoprotein. J Mol Biol 363: 27–38.

72. Ishidate T, Yoshihara S, Kawasaki Y, Roy BC, Toyoshima K, et al. (1997) Identification of a novel nuclear localization signal in Sam68. FEBS Lett 409: 237–241.

73. Wu J, Zhou L, Tonissen K, Tee R, Artzt K (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: 29202–29210.

74. Ma AS, Moran-Jones K, Shan J, Munro TP, Snee MJ, et al. (2002) Heterogeneous nuclear ribonucleoprotein A3, a novel RNA trafficking response element-binding protein. J Biol Chem 277: 18010–18020.

75. Bear J, Tan W, Zolotukhin AS, Tabernero C, Hudson EA, 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: 6306–6317.

76. Katahira J, Strasser K, Podtelejnikov A, Mann M, Jung JU, et al. (1999) The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. EMBO J 18: 2593–2609.

77. Rebane A, Aab A, Steitz JA (2004) Transportins 1 and 2 are redundant nuclear import factors for hnRNP A1 and HuR. RNA 10: 590–599.

78. Siomi MC, Eder PS, Kataoka N, Wan L, Liu Q, et al. (1997) Transportin-mediated nuclear import of heterogeneous nuclear RNP proteins. J Cell Biol 138: 1181–1192.