Published in final edited form as:
Traffic. 2010 March ; 11(3): 311–323. doi:10.1111/j.1600-0854.2009.01028.x.

Expanding the Definition of the Classical Bipartite Nuclear Localization Signal

Allison Lange[*,1], **Laura M. McLane**[*,2], **Ryan E. Mills**[3], **Scott E. Devine**[4], and **Anita H. Corbett**[1]
[1]Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road NE, Atlanta, GA 30322

Abstract

Nuclear localization signals (NLSs) are amino acid sequences that target cargo proteins into the nucleus. Rigorous characterization of NLS motifs is essential to understanding and predicting pathways for nuclear import. The best-characterized NLS is the classical NLS (cNLS), which is recognized by the cNLS receptor, importin-α. cNLSs are conventionally defined as having one (monopartite) or two clusters of basic amino acids separated by a 9-12 amino acid linker (bipartite). Motivated by the finding that Ty1 integrase, which contains an unconventional putative bipartite cNLS with a 29 amino acid linker, exploits the classical nuclear import machinery, we assessed the functional boundaries for linker length within a bipartite cNLS. We confirmed that the integrase cNLS is a *bona fide* bipartite cNLS, then carried out a systematic analysis of linker length in an obligate bipartite cNLS cargo, which revealed that some linkers longer than conventionally defined can function in nuclear import. Linker function is dependent on the sequence and likely the inherent flexibility of the linker. Subsequently, we interrogated the *Saccharomyces cerevisiae* proteome to identify cellular proteins containing putative long bipartite cNLSs. We experimentally confirmed that Rrp4 contains a bipartite cNLS with a 25 amino acid linker. Our studies reveal that the traditional definition of bipartite cNLSs is too restrictive and linker length can vary depending on amino acid composition

Keywords

nuclear protein import; importin-α; classical nuclear localization signal (cNLS); bipartite cNLS; Ty1 integrase; nucleome; Rrp4

A hallmark of the eukaryotic cell is the physical separation of the genetic material and associated proteins in the nucleus from the translational machinery in the cytoplasm. The division of the nuclear and cytoplasmic compartments by the nuclear envelope necessitates a mechanism by which proteins and other macromolecules can move into and out of the nucleus as is required for cellular processes. This macromolecular transport occurs through large, proteinaceous structures that perforate the nuclear envelope called nuclear pore complexes (NPCs) (1-4). NPCs permit passive diffusion of small molecules, but proteins over ~40 kDa undergo active receptor-mediated transport to enter the nucleus (5).

Corresponding Author: Anita H. Corbett, Ph.D., Emory University School of Medicine, Department of Biochemistry, 1510 Clifton Rd., NE, Atlanta, GA 30322; Tel: (404) 727-4546; Fax: (404) 727-2738; acorbe2@emory.edu.
[2]Current Address: Department of Microbiology, University of Pennsylvania School of Medicine, 433 South University Avenue, Philadelphia, PA 19104
[3]Current Address: Brigham and Women's Hospital, Boston, MA 02115
[4]Current Address: Institute for Genome Sciences, University of Maryland School of Medicine, 801 W. Baltimore Street Baltimore, MD 21201
[*]These authors contributed equally to the work.

Nuclear transport receptors recognize diverse targeting signals within macromolecular cargoes to mediate transport. A large family of transport receptors, termed the karyopherin-β (kapβ) family, is responsible for most active protein transport through NPCs (6). In general, kapβ receptors recognize and specifically bind to a set of sequences called nuclear localization signals (NLSs) within the primary structure of cargo proteins (7). Following binding of the NLS by a kapβ receptor, kapβ transiently interacts with components of the NPC called nucleoporins to mediate translocation into the nucleus (8). Although the majority of nuclear import occurs via direct binding of a kapβ receptor to a cargo protein, the best-studied system for protein import is the classical protein nuclear import pathway, which requires an adaptor protein, importin-α, to recognize and bind to a specific class of NLS motifs, termed classical NLSs (cNLSs). Importin-α escorts the cargo into the nucleus in complex with the kapβ family member, importin-β. Once inside the nucleus, importin-β is bound by a small GTPase, RanGTP, which causes a conformational change within importin-β, resulting in dissociation of the import complex and release of the cargo protein in the nucleus (9).

Classical NLS motifs are defined as either monopartite, consisting of a single stretch of basic amino acids, or bipartite, consisting of two stretches of basic amino acids separated by a linker region (10-12). The simian virus 40 (SV40) T-antigen NLS (^{126}P**K**<u>**K**</u>**KRK**V^{132}) is the canonical monopartite cNLS. Co-crystal structures of importin-α and monopartite SV40 cNLS peptides reveal that monopartite cNLSs bind in an extended conformation along a major binding pocket of importin-α (13,14). Furthermore, functional studies of this cNLS reveal that the lysine in the third position (underlined above) is critical for productive nuclear targeting of the SV40 NLS (15,16). The best-studied bipartite cNLS is found in the *Xenopus laevis* protein, nucleoplasmin (^{150}**KR**PAATKKAGQA**KKK**^{170}) (12,17). In a bipartite cNLS, upstream basic residues bind to a minor binding pocket on importin-α and a downstream monopartite-like sequence within the cNLS interacts with the major binding pocket (13,14,18,19). Structural studies of mouse importin-α co-crystallized with bipartite cNLS peptides show that the region between the major and minor binding pockets of importin-α can make contacts with both the backbone and the side chains of the amino acids in the bipartite cNLS linker (20).

The linker of the bipartite cNLS has traditionally been limited to ten amino acids based on historical characterization of the nucleoplasmin cNLS (12,17) and, importantly, freely available NLS prediction programs such as PSORT II and PredictNLS restrict the bipartite linker to between 9 and 12 residues (21,22). Bipartite cNLSs with linkers as long as twelve residues can bind to importin-α *in vitro* (20) and studies of artificial peptide-based nuclear targeting signals have revealed atypical longer sequences that may contain linkers of up to 20 amino acids (23). Bipartite cNLSs with much longer linkers have also been proposed in cellular proteins such as Smad4 (24) and topoisomerase II (25), though no rigorous studies have shown that these sequences are true bipartite cNLSs that mediate import *in vivo*. Furthermore, the Ty1 integrase protein, an essential protein of the Ty1 retrotransposon in the yeast *Saccharomyces cerevisiae*, contains an unconventional putative bipartite cNLS at its C-terminus (26,27). The proposed bipartite cNLS of integrase contains two lysine-lysine-arginine (KKR) motifs, termed basic region 1 and 2 (BR1 and BR2) separated by a 29 amino acid linker (see Figure 1A). Nuclear import of Ty1 integrase is dependent upon the classical import receptor, importin-α (28); however, the mechanism by which the integrase cNLS interacts with importin-α has not been defined.

The presence of a 29 amino acid linker within the Ty1 integrase cNLS raises the question of whether the integrase cNLS sequence constitutes a *bona fide* bipartite cNLS motif or whether the two basic regions function as two independent monopartite signals. To address this question, we examined mutants of integrase cNLS BR1 and BR2. In addition, we took

advantage of an artificial obligate bipartite cNLS (bpSV40 T3 NLS) to independently analyze the 29 amino acid linker of the integrase cNLS. Taken together with previous studies (26,29), our results reveal that the integrase cNLS is, in fact, a classical bipartite NLS with an unconventionally long linker sequence.

Based on the finding that the integrase cNLS contains a functional linker region that is nearly three times longer than the conventional definition for cNLSs, we sought to reevaluate the current consensus sequence for classical bipartite NLS motifs. To this end, we manipulated the length of the linker region between the basic amino acid regions of the artificial bpSV40 T3 NLS to investigate the effect of linker length on nuclear targeting of bipartite cNLS sequences via importin-α. We then queried the *S. cerevisiae* proteome using an expanded version of the PSORT II algorithm (21) for bipartite cNLSs to identify candidate cargoes that contain putative long bipartite cNLS motifs. Through a preliminary study of several candidate proteins, we demonstrated that Rrp4, a subunit of the exosome (30) contains a classical bipartite NLS motif with a linker length of 25 amino acids that is both necessary and sufficient for importin-α-mediated import *in vivo*. Taken together, these findings suggest that the function of the linker region within a classical bipartite NLS is sequence specific and that the cNLS consensus should be expanded to include a linker region consisting of up to 29 amino acids. This revision of the cNLS definition greatly increases the pool of potential bipartite cNLS-bearing cargo proteins.

Results

Functional analysis of the putative bipartite cNLS within Ty1 integrase

The Ty1 integrase cNLS, which interacts with and depends on the classical protein nuclear import machinery (importin-α/β) to access the nucleus (28), has been proposed to function through a bipartite mechanism despite its exceptionally long linker region (26,29). Previous studies have shown that amino acid substitutions within BR1 and BR2 (KKK→GGG) of integrase individually do not result in a complete loss of nuclear accumulation, which suggests that the integrase cNLS may function as two individual monopartite cNLS motifs (26). However, due to the single proton R-group on glycine residues, it is difficult to determine if the observed localization defect is due to an alteration of the NLS sequence itself or an overall structural change caused by the glycine residue substitutions.

To examine whether the two basic regions in Ty1 integrase constitute a bipartite cNLS, we exploited a previously characterized GFP-GFP fusion to the integrase cNLS region (amino acids 595-630) termed GFP_2-IN NLS (28) and localized various GFP_2-IN NLS reporter plasmids containing alanine substitutions at both BR1 and BR2 (Figure 1A and 1B). Simultaneous amino acid substitution within both BR1 and BR2 (GFP_2-IN NLS_{br1br2}) caused a total loss of nuclear localization of the reporter. Amino acid substitutions within BR1 or BR2 that create either GFP_2-IN NLS_{BR1br2} or GFP_2-IN NLS_{br1BR2}, respectively, caused mislocalization of the GFP_2-IN NLS reporter to the cytoplasm; however, some nuclear signal remained.

To assess the functional consequences of changing each BR sequence, we performed a previously described retrotransposition assay (31). Alanine substitutions within integrase BR1 and BR2 each severely compromised Ty1 retrotransposition (Figure 1C), consistent with previous studies where basic residues were changed to glycine (26,29). Taken together with localization studies, our data suggest that both BR1 and BR2 are necessary for wildtype levels of retrotransposition. In addition, although BR1 and BR2 might have the capacity to function as weak individual cNLS motifs, our data support a model where the integrase cNLS functions as a bipartite cNLS sequence with a 29 amino acid linker region since the

presence of both BR regions is required for significant levels of both nuclear import and retrotransposition.

Expanding the classical bipartite NLS linker

This analysis suggests that the Ty1 retrotransposon integrase protein contains a bipartite cNLS with an unconventionally long 29 amino acid linker. To more generally investigate the boundaries of the length of bipartite cNLS linker sequences as compared to the established consensus linker length of 9-12 amino acids (21,22), we created a series of GFP_2 reporter proteins with variable linker lengths. For these experiments, it was critical to employ a reporter that is dependent on both regions of basic amino acids and, hence, a bipartite mode of binding to importin-α for nuclear import. Therefore, we took advantage of a modified version of the artificial bpSV40 NLS sequence (bpSV40 T3 NLS), which contains a lysine to threonine amino acid substitution (underlined) in the second cluster of amino acids (**KR**TADGSEFESP**K<u>T</u>KRKV**) (15,16). As a monopartite signal, the modified PKTKRKV motif does not interact with importin-α with sufficient affinity (K_D ~3000 nM) to mediate efficient nuclear import; however, with the addition of the upstream KR motif, the bpSV40 T3 NLS binds tightly to importin-α *in vitro* (K_D ~13.5 nM) and efficiently localizes a reporter to the nucleus *in vivo* (15,32). The bpSV40 T3 NLS provides a tool that can be used to study variation within the linker region of a bipartite cNLS sequence while ensuring that all nuclear import that occurs is mediated through an importin-α-dependent bipartite mechanism.

GFP_2 reporter proteins that contain bipartite cNLS motifs with artificial linkers of 8, 10, 13, 15, 20, or 30 amino acids were localized in wildtype cells by direct fluorescence microscopy (Figure 2A). The sequences of the linker regions within this series of reporters (Table 1) were derived from the artificial linker region within bpSV40, which has been previously studied both *in vivo* and *in vitro* (15, 32). As expected, the bpSV40 T3 NLS-GFP_2 reporter with a wildtype linker of 10 amino acids, bpSV40 T3 NLS (10 aa), was predominantly localized to the nucleus. The bpSV40 T3 NLS (8 aa) and (13 aa) reporters were also primarily nuclear. The bpSV40 T3 NLS (15 aa) reporter showed slightly more cytoplasmic signal than the bpSV40 T3 NLS (13 aa) reporter. The bpSV40 T3 NLS (20 aa) reporter had some nuclear accumulation with a strong cytoplasmic signal and the bpSV40 T3 NLS (30 aa) reporter showed virtually no nuclear concentration. These data suggest that classical bipartite NLS motifs with this artificial linker sequence ranging from 8 to at least 20 amino acids can mediate nuclear import *in vivo*.

To verify that the bpSV40 T3 NLS-GFP_2 reporter proteins are imported by the classical, importin-α-mediated, nuclear protein import system, these reporter proteins were localized in a mutant of importin-α, *srp1-54*, which mislocalizes importin-α-dependent cargoes (Figure 2B) (33, 34). All of the bpSV40 T3 NLS-GFP_2 linker variants displayed an increase in cytoplasmic localization in the *srp1-54* cells as compared to wildtype cells. As a control, Nab2, a protein which uses a pathway independent of importin-α for nuclear import (35), remained nuclear. These data confirm that import of the bpSV40 T3 NLS reporters is dependent on the classical nuclear import machinery.

The linker region of bipartite cNLSs can make contact with importin-α (20) and the amino acid composition of the linker region can affect import of cNLS-containing cargo proteins (23). To verify that functional longer cNLS linkers are not limited to a single sequence, we created two additional series of GFP_2 reporter proteins that contain either a flexible and energetically stable linker, consisting of a serine/glycine repeat (36,37) or a more rigid and charge neutral linker, consisting of repeating alanine residues (38). Each series of reporters was then localized in wildtype cells (Figures 2C and 2D). The bpSV40 T3 NLS-GFP_2 reporters with serine/glycine linkers yielded similar results to the artificial bpSV40 T3 linker

(compare Figure 2A and 2C). Both series of reporters showed the most nuclear localization with a 10 aa linker, with more cytoplasmic localization as the linker length approached 20 aa until the reporter containing the 30 aa linker showed virtually no nuclear concentration. However, the bpSV40 T3 NLS-GFP_2 reporters with the inflexible alanine linkers showed significant nuclear localization with linkers of 8 aa and 10 aa and only slight nuclear localization with linkers of 12 aa and 13 aa. All alanine linker reporters with linkers longer than 13 aa were not concentrated to the nucleus. These results strongly suggest that the flexibility of the linker is critical, particularly for linkers longer than the conventional length.

The integrase cNLS linker region can contribute to a long classical bipartite NLS motif

The bpSV40 T3 NLS-GFP_2 fusion proteins with artificial linker sequences or serine/glycine linker sequences showed steady-state nuclear localization only when the linker length is between 8 and 20 amino acids; however, in Figure 1, we provided evidence that the integrase cNLS functions as a bipartite cNLS even though it contains a 29 aa linker region. To examine whether the linker region from the integrase cNLS can contribute to the nuclear localization of the obligate bipartite cNLS reporter, we inserted the 29 aa integrase linker or a variant of the integrase linker (mutIN) containing four amino acid substitutions (E601Q/D602N/E604Q/E606Q), which have previously been shown to prevent nuclear accumulation of integrase (26), into the bpSV40 T3 NLS-GFP_2 reporter (Table 1) and localized the reporter proteins in wildtype cells (Figure 3A). Despite the exceptionally long 29 amino acid linker, the bpSV40 T3 NLS-GFP_2 (IN) reporter localized exclusively to the nucleus. The bpSV40 T3 NLS-GFP_2 reporter containing the mutIN linker showed increased cytoplasmic localization as compared to the reporter containing the wildtype linker. As expected, the bpSV40 T3 NLS-GFP_2 (10 aa) reporter localized primarily to the nucleus, while GFP_2 and the bpSV40 T3 NLS-GFP_2 (30 aa) reporter localized throughout the cell. These data reveal that the 29 aa integrase linker can function in the context of a bipartite cNLS and provide further evidence that the specific sequence within the linker is important for function.

To ensure that the nuclear accumulation observed with the bpSV40 T3 NLS (IN) reporter depends upon importin-α, the localization of bpSV40 T3 NLS-GFP_2 (10 aa), bpSV40 T3 NLS-GFP_2 (IN), or control Nab2-GFP was examined by direct fluorescence microscopy in *srp1-54* cells (Figure 3B). Both the bpSV40 T3 NLS-GFP_2 (10 aa) and the bpSV40 T3 NLS-GFP_2 (IN) reporters showed a dramatic reduction in nuclear localization as compared to wildtype cells, while Nab2, a control protein imported into the nucleus in an importin-α-independent manner (35), remained nuclear.

To confirm that the bpSV40 T3 NLS containing the 29 aa integrase cNLS linker binds directly to importin-α, we used a quantitative solid phase plate binding assay (39). We assayed binding between a truncated form of importin-α (ΔIBB) and purified recombinant bpSV40 T3 NLS-GFP protein containing the artificial 10 amino acid linker (10 aa), the long 30 amino acid linker (30 aa), or the linker from Ty1 integrase (IN). ΔIBB-importin-α lacks the N-terminal importin-β binding (IBB) domain, which is auto-inhibitory and competes with cNLS-containing cargoes for binding to the cNLS-binding pockets of importin-α (40). Therefore, ΔIBB-importin-α simulates the state of importin-α while in an import complex where the IBB domain is sequestered away from the cNLS-binding pockets by importin-α, allowing cNLS-containing proteins to interact with the binding pockets on importin-α. As shown in Figure 4 and Table 2, bpSV40 T3 NLS-GFP (IN) bound importin-α with a binding affinity of ~25 nM and bpSV40 T3 NLS-GFP (10 aa) bound importin-α with a similar binding affinity of ~27 nM. No interaction was detected between either GFP alone or bpSV40 T3 NLS-GFP (30 aa) and importin-α, which is consistent with the observation that an obligate bipartite cNLS containing this artificial 30 aa linker cannot target cargo to the nucleus.

Together, these data confirm that the atypically long linker within the integrase cNLS can function as part of a classical bipartite NLS. Therefore, not only can the linker region of bipartite cNLSs vary from the canonical 9-12 amino acids, but the sequence of the linker region can also influence the ability of the cNLS to function and interact with importin-α since the 29 aa sequence from integrase functions significantly more efficiently than the 30 aa artificial linker (see Table 1).

Candidate bipartite cNLS-containing proteins in the S. cerevisiae *proteome*

The finding that classical bipartite NLS motifs with linker lengths other than the conventional 9-12 amino acids can both function *in vivo* and mediate high affinity binding to importin-α has the potential to greatly expand the pool of cargo proteins that could be targeted to the nucleus by a bipartite cNLS mechanism. Putative bipartite cNLS-cargoes currently are predicted to represent about 16.5% of the budding yeast proteome (41). To assess how expanding the definition of the bipartite cNLS impacts the number of putative cargoes that could enter the nucleus via the classical import pathway, the PSORT II algorithm for classical bipartite NLS motifs (21) was modified to include linker lengths ranging from 8 to 30 residues. The revised prediction of new putative bipartite cargoes in the *S. cerevisiae* proteome is presented in Table 3. This updated predicted prevalence of classical bipartite-mediated nuclear import is likely greatly overestimated since an individual protein may contain more than one length of bipartite cNLS and, therefore, be present in multiple categories. However, this analysis provides a library of candidate proteins that contain putative bipartite cNLS motifs with extended linker regions for further investigation. Each of these putative bipartite cNLS sequences with an atypical linker must be experimentally evaluated before it can be deemed a functional targeting sequence. To identify those proteins most likely to contain an atypical bipartite cNLS motif, we searched our library for putative bipartite-containing proteins that interact with importin-α according to the BioGRID database (42) and that are localized at steady-state to the nucleus or nucleolus according to a comprehensive subcellular localization study performed with a global yeast GFP-fusion library (43). Candidate cargoes that emerged from this search are presented in Supplemental Data, Table S1.

The candidate proteins chosen for further analysis (Csl4, Rrp4, and Mft1) contain exactly one possible long bipartite cNLS and have linker regions of 16 (Csl4), 25 (Rrp4), and 28 (Mft1) amino acids (Table 4). To ensure that the candidate proteins show steady-state nuclear localization, we examined the localization of each protein as an integrated C-terminal GFP fusion protein (Figure 5A). As reported (43), Csl4, Rrp4, and Mft1 are primarily localized to the nucleus in wildtype cells.

The predicted Rrp4 and Mft1 cNLSs are sufficient for importin-α-mediated nuclear import

To determine if the putative bipartite cNLS motifs in Csl4, Rrp4, and Mft1 are sufficient for nuclear import, GFP_2-reporters were created that contain each of the potential targeting sequences (Csl4 aa 97-125, Rrp4 aa 111-159, or Mft1 aa 244-283) and were localized in both wildtype and *srp1-54* cells (Figure 5B). Csl4 NLS-GFP_2 localized to the cytoplasm in both wildtype and *srp1-54* cells. GFP_2-Rrp4 NLS showed some nuclear accumulation in wildtype cells and was more cytoplasmic in *srp1-54* cells. Mft1 NLS-GFP_2 mainly localized to the nucleus in wildtype cells and was more cytoplasmic in *srp1-54* cells. As controls, bpSV40 T3 NLS-GFP_2 (10 aa) and Nab2-GFP were both localized to the nucleus in wildtype cells, but only bpSV40 T3 NLS-GFP_2 (10 aa) became more cytoplasmic when examined in *srp1-54* cells. These results show that both the Rrp4 and Mft1 predicted cNLS motifs are sufficient to mediate importin-α-mediated nuclear import *in vivo*.

To determine if the predicted bipartite cNLS linkers within Rrp4 and Mft1 can contribute to a functional bipartite cNLS sequence, we created bpSV40 T3 NLS-GFP_2 reporters containing the putative linkers from these two proteins (Table 4). Localization of these bpSV40 T3 NLS-GFP_2 reporters was analyzed in wildtype cells by direct fluorescence microscopy (Figure 6A). Wildtype cells expressing bpSV40 T3 NLS-GFP_2 (Rrp4), which contains a 25 aa linker, displayed a low level of nuclear accumulation of the reporter. In contrast, the bpSV40 T3 NLS-GFP_2 (Mft1) reporter did not localize to the nucleus. As controls, the bpSV40 T3 NLS-GFP_2 (10 aa) and bpSV40 T3 NLS-GFP_2 (IN) reporters both localized to the nucleus and the bpSV40 T3 NLS-GFP_2 (30 aa) reporter was primarily cytoplasmic. This result further confirms the importance of specific sequences within the linker region of bipartite cNLS motifs.

To investigate whether the localization of bpSV40 T3 NLS-GFP_2 (Rrp4) is dependent upon importin-α, we localized the Rrp4 linker reporter in *srp1-54* cells (Figure 6B). Nuclear localization of the bpSV40 T3 NLS-GFP_2 (Rrp4) reporter was decreased in *srp1-54* cells as compared to wildtype cells. Taken together, these data show that the Rrp4 linker can contribute to a functional bipartite cNLS that is dependent on importin-α and suggest that Rrp4 may indeed contain a bipartite cNLS motif with a linker longer than conventionally defined.

The putative bipartite cNLS within Rrp4 is necessary for proper nuclear import

Since the long linker of Rrp4 can contribute to a functional bipartite cNLS both in the context of the bpSV40 T3 reporter and in the context of the Rrp4 cNLS itself, we sought to confirm the bipartite nature of the Rrp4 cNLS. To this end, Rrp4 and two variants of Rrp4 containing amino acid substitutions within either of the basic regions of the putative Rrp4 cNLS (NLSmut1, K122A/R123A or NLSmut2, R149A/R150A/K151A) were fused to GFP and localized in wildtype cells (Figure 6C). This plasmid-encoded Rrp4-GFP localized mainly to the nucleus as was observed previously for integrated Rrp4-GFP (See Figure 5). The Rrp4 variants, Rrp4 NLSmut1-GFP and Rrp4-NLSmut2-GFP, both localized mainly to the cytoplasm with some nuclear signal. Since both basic regions of Rrp4 are necessary for proper localization of Rrp4-GFP, these data strongly suggest that these amino acid clusters cooperate to form a bipartite nuclear targeting sequence with a long linker.

Discussion

In recent years, there have been an increasing number of studies designed to define novel nuclear localization signals within various cargo proteins. Classical NLS motifs that interact with importin-α are the best-studied NLS sequences; however, classes of non-canonical classical NLS motifs that interact with importin-α have recently begun to emerge (23,26-28,44-46). The finding, for example, that the Ty1 integrase protein utilizes the classical nuclear protein import machinery (28) despite containing a putative bipartite cNLS with a linker nearly three times the size of traditional linker sequences (26,27) prompted us to reinvestigate what truly defines a classical bipartite NLS.

Previous sequence analysis of the Ty1 integrase cNLS suggested that it is a bipartite cNLS motif (26,27). Our data confirm that the integrase cNLS can function as a classical bipartite NLS, suggesting that bipartite cNLS linkers can, in fact, be significantly longer than what had previously been defined (10-12). The finding that the bpSV40 T3 NLS reporter containing the integrase linker, but not random sequence, tandem serine/glycine residues, or tandem alanine residues of nearly the same length, localizes to the nucleus, provides compelling evidence that the sequence of the linker is important. This dependence on linker sequence likely reflects a need to adopt a particular conformation to allow the two basic regions of the cNLS to assume an orientation that facilitates interaction with the two cNLS-

binding pockets on importin-α. A further consideration beyond flexibility is the possibility that the linker region may interact with the body of importin-α, which may require specific sequence contacts. In support of this hypothesis, previous studies have shown and we have confirmed that deletion of, or amino acid substitutions within, a series of acidic amino acids in the linker region of the Ty1 integrase cNLS abolishes nuclear import (26), again suggesting that the sequence of the linker region is key for facilitating interactions with importin-α and subsequent nuclear localization of bipartite cNLS-bearing cargo. Interestingly, it seems that the combination of particular cNLS basic residues with particular linkers may be required for proper nuclear targeting. Our finding that the Mft1 linker can contribute to a functional cNLS in the context of its own upstream and downstream basic residues, but not in the context of the bpSV40 T3 NLS, suggests that there may be a requisite collaboration between the basic residues and the linker region in certain bipartite cNLSs.

Based on the surprising result that the long 29 amino acid linker region in the Ty1 integrase NLS can contribute to a functional cNLS, we modified the PSORT II algorithm to allow for substantial variation in the length of the bipartite cNLS linker and identified cellular proteins that contain predicted bipartite cNLS sequences whose linkers range between eight and thirty amino acids. Of the predicted proteins, we analyzed the linker sequences of three candidate nuclear proteins (Csl4, Rrp4, and Mft1) and found that the predicted Rrp4 and Mft1 bipartite cNLSs are indeed sufficient for classical nuclear import. These data are important because they suggest that the phenomenon of long bipartite cNLS motifs is not limited to atypical cargoes such as viral proteins, which may have developed unusual or at least unrepresentative mechanisms to exploit the existing host machinery. Rather, cellular proteins such as Rrp4 also bear long bipartite cNLSs and are targeted to the nucleus by the classical nuclear protein import system. However, it is important to keep in mind that not all putative classical NLSs with long linkers will prove to be true targeting signals, as evidenced by the finding that only one of the long linkers tested was able to contribute to a functional cNLS. Our results with the alanine linker also highlight the concept that even predicted bipartite cNLSs with linkers of conventional length may not all function *in vivo* and reinforce the need to experimentally test not only potential long bipartite cNLS sequences, but all putative cNLS motifs before declaring them functional nuclear targeting signals.

The expansion of the consensus for the classical bipartite NLS vastly increases the number of putative classical nuclear protein import cargoes in the cell and provides a starting point for researchers in varied fields to begin to investigate potential mechanisms for nuclear import of their favorite protein. These studies may shed light on the function of the protein or suggest novel methods of temporal-spatial regulation. For example, Rrp4 is a component of both the nuclear and cytoplasmic exosomes (30). Previously, it had been proposed that Rrp6, a protein found in the nuclear exosome but not the cytoplasmic exosome, is the component that is responsible for nuclear import of the exosome (30) even though no NLS within Rrp6 has been functionally defined. Recent studies have also implicated the Rrp6-associated protein, Dis3, in the nuclear localization of the nuclear exosome components (47). Our studies based on the expanded cNLS archetype suggest that Rrp4 may mediate, either individually or in combination with Rrp6 and Dis3, the import of the nuclear exosome components. Interestingly, Rrp4 contains a phosphorylation site directly after the long bipartite cNLS (serine residue 152) that is conserved in the human ortholog (48). Phosphorylation within or near NLS sequences can control nuclear import of cargo proteins by modulating the affinity of the NLS for its receptor (49-51). Therefore, it is possible that import of the nuclear exosome components is managed by regulating the binding affinity of the novel long bipartite cNLS within Rrp4 for importin-α. This possibility underscores the hypothesis that longer linkers may provide enhanced opportunity for regulation through

post-translational modifications that directly impact receptor binding or influence conformation.

In conclusion, we provide evidence that the traditional consensus for a bipartite cNLS is too restrictive. Our studies uncovered importin-α-dependent bipartite cNLS motifs within Ty1 integrase and Rrp4 that contain linkers of 29 and 25 amino acids, respectively, which allowed us to expand the bipartite cNLS consensus beyond the previous limitations of a 9-12 amino acid linker. Furthermore, even some eight amino acid linkers were found to be functional in this study. After refining the cNLS definition, we interrogated the *S. cerevisiae* proteome and discovered that the prevalence of classical nuclear transport may be much greater than previously appreciated (41). Overall, our studies extend the definition of the classical bipartite NLS and provide evidence that both linker length and sequence influence nuclear targeting function.

Materials and Methods

Strains and plasmids

All DNA manipulations were carried out according to standard methods (52) and all media were prepared by standard procedures (53). *Saccharomyces cerevisiae* strains and plasmids for this study are listed in Table 5. For some localization studies, two tandem GFP molecules (GFP_2) were fused to either the N-terminus or C-terminus of candidate NLS motifs to generate a reporter of greater than 55 kDa, which prevents passive diffusion through the NPC due to the size limitations of the pore. All chemicals were obtained from Ambion (Austin, TX), Sigma, U.S. Biologicals (Swampscott, MA) or Fisher Scientific unless otherwise specified.

Construction of IN NLS variants

All variants of the integrase NLS were created using a QuikChange site-directed mutagenesis kit (Stratagene). Briefly, alanine residues (AAA) were substituted for the KKR motifs in basic region 1 (BR1) and/or basic region 2 (BR2) in GFP_2-IN NLS (pAC1804) to create the reporters (GFP_2-IN NLS_{BR1br2}, GFP_2-IN NLS_{br1BR2}, GFP_2-IN NLS_{br1br2}) illustrated in Figure 1. A complementary set of integrase NLS BR1 and BR2 variants was also created in the context of the intact integrase open reading frame of the pAR100 retrotransposition test plasmid.

Direct fluorescence microscopy

Localization of GFP-fusion proteins was analyzed by direct fluorescence microscopy in living cells. Cells expressing GFP-fusion proteins expressed from their own promoters were grown overnight in YEPD at 30°C, diluted in fresh media, and grown for three hours prior to localization studies. Cells expressing constructs under control of the *MET25* promoter were grown overnight in selective media at 30°C, pelleted, washed once in dH_2O, resuspended in selective media, and grown for five hours at 30°C to induce expression prior to localization studies. The GFP signal was visualized using a GFP-optimized filter (Chroma Technology) on an Olympus BX60 epifluorescence microscope equipped with a Photometrics Quantix digital camera. In some cases, cells were incubated with 4.5 nM Hoechst dye (Sigma) to visualize chromatin and reveal the location of the nucleus.

Ty1 retrotransposition assay

The Ty1 retrotransposition assay was performed as previously described (31). Because the Ty1 retroelement is flanked by direct repeat sequences, homologous recombination can occur between the plasmid-borne Ty1 element and genomic Ty1 elements integrated throughout the genome. To eliminate the false positives that could be detected due to such

recombination events, we utilized *Δrad52* cells where the levels of homologous recombination between the plasmid-borne elements and the genomic Ty1 elements are significantly reduced, decreasing the number of false positives detected in our assay (54).

Briefly, *Δrad52* cells were transformed with the pAR100 (*HIS3*) test plasmid containing either wildtype Ty1 integrase sequence or the Ty1 integrase variants analogous to the GFP_2-IN NLS reporter plasmids (Figure 1A). Transformants were selected on SC ura^- glu plates. To initiate the retrotransposition assay, nine transformants were patched onto SC ura^- glu plates. These plates were then replica plated to SC ura^- gal and grown for two days at 25°C to induce transposition. Patches were then replica plated sequentially to: a) yeast peptone dextrose (YEPD); b) SC medium containing 1.2 g/liter 5-fluoroorotic acid (5-FOA) and 2% glucose; and c) SC his^- glu. Growth on the final selection plate was compared to controls: *Δrad52* cells containing the WT pAR100 test plasmid (positive control) or *Δrad52* cells containing a *URA3* vector control plasmid (pRS316) (negative control). Results were quantified by counting the number of His^+ colonies for each Ty1 variant and plotting the average of the counts as a percentage of the total number of colonies observed with the wildtype pAR100 test plasmid.

Protein purification

Purified recombinant proteins used in these studies (His_6-tagged NLS-binding fragment of importin-α consisting of residues 89-530 (ΔIBB) and GFP-fusion bpSV40 T3 NLSs with various linker regions) were expressed in *E. coli* BL21 (DE3) cells and purified by nickel affinity chromatography as previously described (40). The ΔIBB-importin-α lacks the auto-inhibitory domain of importin-α that competes with NLS-cargoes for binding to the binding pockets of importin-α. ΔIBB-importin-α has a similar affinity for cNLS cargo as importin-α in the context of the importin-α/β import complex (40).

Quantitative binding assay

Solid phase binding assays were performed essentially as previously described (39). Microtiter Immulux HB plates (Dynex) were coated with 100 μl/well of 50 nM GFP-proteins [GFP, bpSV40 T3 NLS-GFP (10 aa), bpSV40 T3 NLS-GFP (30 aa), bpSV40 T3 NLS-GFP (IN)] in coating buffer (1×PBS supplemented with 2 mM DTT and 0.2 mM phenylmethylsulfonyl fluoride) overnight at 4°C on a rocker. Plates were then washed three times by in 1×PBS and incubated in 100 μl of binding buffer (coating buffer supplemented with 3% BSA and 0.1% Tween 20) for 3.5 hours at 4°C on a rocker. Two hundred μl of 0-200 nM S-tagged ΔIBB-importin-α protein in binding buffer was added to each well and incubated overnight at 4°C to allow binding to S-tagged ΔIBB-importin-α. Plates were washed three times in binding buffer lacking BSA. Proteins were then cross-linked for 15 min at room temperature in 1 mg/ml 1-ethyl-3-(3-dimethylaminopropyl) carboiimide (Pierce) in binding buffer lacking BSA. The wells were then subjected to a series of washes: 1) 20 min in PBS-T (1×PBS supplemented with 0.2% Tween 20); 2) 10 min with PBS-T containing 100 mM ethanolamine; and finally, 3) 10 min with PBS-T containing 3% BSA. The bound S-tagged ΔIBB-importin-α was then detected by incubation with S-protein-horseradish peroxidase conjugate (Novagen) in coating buffer containing 1% BSA and 0.1% Tween 20 for 1 hr at 4°C on a rocker. The plates were washed three times in 1×PBS. Following the washes, 100 μl/well of horseradish peroxidase substrate [1-Step™ Turbo TMB (3,3'5,5'-tetramethylbenzidine)-ELISA (Pierce)] was incubated in each well for 20 min at room temperature. The reaction was stopped by addition of 100 μl 2 M H_2SO_4. The absorbance of the samples was measured at 450 nm with an EL_X808 Ultra Microplate Reader with KCjunior software (Bio-Tek Instruments, Inc.). Average absorbance values at OD_{450} were determined for GFP fusion proteins at each S-tagged ΔIBB-importin-α concentration. Background GFP absorbance values were subtracted from those of the

bpSV40 T3 NLS-GFP proteins. The absorbance values for the bpSV40 T3 NLS-GFP proteins at each S-tagged ΔIBB-importin-α concentration were used to generate binding curves by non-linear regression using Prism5 software (GraphPad Software, Inc.). Using Prism5, we calculated the K_D of ΔIBB-importin-α binding to each of the bpSV40 T3 NLS-GFP proteins.

Bioinformatics

The algorithm from PSORT II (21) for classical bipartite NLS motifs was used with ScanProsite to search three sets of data: the 5850 proteins in the *S. cerevisiae* GenBank (55), the 1515 proteins localized to either the nucleus or the nucleolus in a comprehensive subcellular localization study performed with a global yeast GFP-fusion library (43), and the 224 proteins that interact with importin-α according to the BioGRID database (42). To determine the prevalence of putative cargos that contain a putative classical bipartite NLS with a linker of non-standard length, the PSORT II algorithm for bipartite NLSs was modified to allow for the following linkers: 8-9, 11-13, 14-15, 16-20, or 21-30 residues ([K/R][K/R]--$X_{8\text{-}30}$--[3 out of 5 K/R]).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Corbett and Devine laboratories for stimulating discussions. This work was supported by grants from the NIH to AHC and SED.

References

1. D'Angelo M, Hetzer M. Structure, dynamics and function of nuclear pore complexes. Trends Cell Biol. 2008; 18(10):456–466. [PubMed: 18786826]
2. Lim R, Ullman K, Fahrenkrog B. Biology and biophysics of the nuclear pore complex and its components. Int Rev Cell Mol Biol. 2008; 267:299–342. [PubMed: 18544502]
3. Allen TD, Cronshaw JM, Bagley S, Kiseleva E, Goldberg MW. The nuclear pore complex: mediator of translocation between nucleus and cytoplasm. J Cell Sci. 2000; 113(2000):1651–1659. [PubMed: 10769196]
4. Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT. The yeast nuclear pore complex. Composition, architecture, and transport mechanism. J Cell Biol. 2000; 148:635–652. [PubMed: 10684247]
5. Sorokin A, Kim E, Ovchinnikov L. Nucleocytoplasmic transport of proteins. Biochemistry (Mosc). 2007; 72(13):1439–1457. [PubMed: 18282135]
6. Weis K. Importins and exportins: how to get in and out of the nucleus. Trends Biochem Sci. 1998; 23:185–189. [PubMed: 9612083]
7. Wozniak RW, Rout MP, Aitchison JD. Karyopherins and kissing cousins. Trends Cell Biol. 1998; 8:184–188. [PubMed: 9695836]
8. Iovine MK, Watkins JL, Wente SR. The GLFG repetitive region of the nucleoporin Nup116p interacts with Kap95p, an essential yeast nuclear nmport factor. J Cell Biol. 1995; 131:1699–1713. [PubMed: 8557738]
9. Lee SJ, Matsuura Y, Liu SM, Stewart MS. Structural basis for nuclear import complex dissociation by RanGTP. Nature. 2005; 435(7042):693–696. [PubMed: 15864302]
10. Kalderon D, Richardson WD, Markham AF, Smith AE. Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature. 1984; 311(5981):33–38. [PubMed: 6088992]
11. Dingwall C, Laskey RA. Nuclear targeting sequences--a consensus? Trends Biochem Sci. 1991; 16(12):478–481. [PubMed: 1664152]

12. Robbins J, Dilworth SM, Laskey RA, Dingwall C. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell. 1991; 64:615–623. [PubMed: 1991323]

13. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell. 1998; 94(1998): 193–204. [PubMed: 9695948]

14. Kobe B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. Nat Struct Biol. 1999; 6:301–304. [PubMed: 10201390]

15. Hodel MR, Corbett AH, Hodel AE. Dissection of a nuclear localization signal. J Biol Chem. 2001; 276(2):1317–1325. [PubMed: 11038364]

16. Kalderon D, Roberts BL, Richardson WD, Smith AE. A short amino acid sequence able to specify nuclear location. Cell. 1984; 39:499–509. [PubMed: 6096007]

17. Dingwall C, Sharnick SV, Laskey RA. A polypeptide domain that specifies migration of nucleoplasmin into the nucleus. Cell. 30(1982):449–458. [PubMed: 6814762]

18. Fontes MR, Teh T, Kobe B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. J Mol Biol. 2000; 297(5):1183–1194. [PubMed: 10764582]

19. Conti E, Kuriyan J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure Fold Des. 2000; 8(3):329–338. [PubMed: 10745017]

20. Fontes MR, Teh T, Jans D, Brinkworth RI, Kobe B. Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha. J Biol Chem. 2003; 278(30):27981–27987. [PubMed: 12695505]

21. Nakai K, Horton P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci. 1999; 24(1):34–36. [PubMed: 10087920]

22. Cokol M, Nair R, Rost B. Finding nuclear localization signals. EMBO Rep. 2000; 1(5):411–415. [PubMed: 11258480]

23. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. JBC. 2009; 284(1):478–485.

24. Xiao Z, Latek R, Lodish HF. An extended bipartite nuclear localization signal in Smad4 is required for its nuclear import and transcriptional activity. Oncogene. 2003; 22:1057–1069. [PubMed: 12592392]

25. Kim KH, Kanbe T, Akashi T, Mizuguchi I, Kikuchi A. Identification of a single nuclear localization signal in the C-terminal domain of an Aspergillus DNA topoisomerase II. Mol Genet Genomics. 2002; 268(3):287–297. [PubMed: 12436251]

26. Moore SP, Rinckel LA, Garfinkel DJ. A Ty1 integrase nuclear localization signal required for retrotransposition. Molecular and Cellular Biology. 1998; 18(2):1105–1114. [PubMed: 9448008]

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

28. McLane L, Pulliam K, Devine S, Corbett A. The Ty1 integrase protein can exploit the classical nuclear protein import machinery for entry into the nucleus. Nucleic Acids Res. 2008; 36(13): 4317–4326. [PubMed: 18586821]

29. Boeke JD, Devine SE. Yeast retrotransposons: finding a nice quiet neighborhood. Cell. 1998; 93(7):1087–1089. [PubMed: 9657139]

30. Mitchell P, Petfalski E, Tollervey D. The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. Genes Dev. 1996; 10(4):502–513. [PubMed: 8600032]

31. Griffith JL, Coleman LE, Raymond AS, Goodson SG, Pittard WS, Tsui C, Devine SE. Functional genomics reveals relationships between the retrovirus-like Ty1 element and its host Saccharomyces cerevisiae. Genetics. 2003; 164(3):867–879. [PubMed: 12871900]

32. Hodel AE, Harreman MT, Pulliam KF, Harben ME, Holmes JS, Hodel MR, Berland KM, Corbett AH. Nuclear localization signal receptor affinity correlates with in vivo localization in *Saccharomyces cerevisiae*. J Biol Chem. 2006; 281(33):23545–23556. [PubMed: 16785238]

33. Yano R, Oakes ML, Tabb MM, Nomura M. Yeast Srp1p has homology to armadillo/plakoglobin/b-catenin and participates in apparently multiple nuclear functions including the maintenance of the nucleolar structure. Proc Natl Acad Sci USA. 1994; 91:6880–6884. [PubMed: 8041713]
34. Hahn S, Maurer P, Caesar S, Schlenstedt G. Classical NLS proteins from Saccharomyces cerevisiae. JMB. 2008; 379:678–694.
35. Lee DC, Aitchison JD. Kap104p-mediated nuclear import. Nuclear localization signals in mRNA-binding proteins and the role of Ran and Rna. J Biol Chem. 1999; 274(41):29031–29037. [PubMed: 10506153]
36. Loughran ST, Loughran NB, Ryan BJ, D'Souza BN, Walls D. Modified His-tag fusion vector for enhanced protein purification by immobilized metal affinity chromatography. Analytical Biochemistry. 2006; 255:148–150. [PubMed: 16814244]
37. Argos P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. JMB. 1990; 211:943–958.
38. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. Protein Science. 2003; 12:1313–1322. [PubMed: 12824478]
39. Bayliss R, Littlewood T, Strawn LA, Wente SR, Stewart M. GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta. J Biol Chem. 2002; 277(52):50597–50606. [PubMed: 12372823]
40. Fanara P, Hodel MR, Corbett AH, Hodel AE. Quantitative analysis of nuclear localization signal (NLS)-importin alpha interaction through fluorescence depolarization. Evidence for auto-inhibitory regulation of NLS binding. J Biol Chem. 2000; 275(28):21218–21223. [PubMed: 10806202]
41. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH. Classical nuclear localization signals: definition, function, and interaction with importin alpha. J Biol Chem. 2007; 282(8):5101–5105. [PubMed: 17170104]
42. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006; 34(Database issue):D535–539. [PubMed: 16381927]
43. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. Global analysis of protein localization in budding yeast. Nature. 2003; 425(6959):686–691. [PubMed: 14562095]
44. McLane L, Corbett A. Nuclear localization signals and human disease. IUBMB. 2009 In Press.
45. Wolff T, Unterstab G, Heins G, Richt J, Kann M. Characterization of an unusual importin alpha binding motif in the borna disease virus p10 protein that directs nuclear import. J Biol Chem. 2002; 277(14):12151–12157. [PubMed: 11796712]
46. Chen M, Ben-Efraim I, Mitrousis G, Walker-Kopp N, Sims P, Cingolani G. Phospholipid scramblase 1 contains a nonclassical nuclear localization signal with unique binding site in importin alpha. J Biol Chem. 2005; 280(11):10599–10606. [PubMed: 15611084]
47. Graham A, Davis S, Andrulis E. Interdependent nucleocytoplasmic trafficking and interactions of Dis3 with Rrp6, the core exosome, and importin-alpha3. Traffic. 2009 Epub ahead of print.
48. Synowsky S, van den Heuvel R, Mohammed S, Pijnappel P, Heck A. Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. Mol Cell Proteomics. 2006; 5(9):1581–1592. [PubMed: 16829593]
49. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. PNAS. 2009; 106(25):10171–10176. [PubMed: 19520826]
50. Beals C, Clipstone N, Ho S, Crabtree G. Nuclear localization of NF-ATc by a calcineurin-dependent, cyclosporin-sensitive intramolecular interaction. Genes Dev. 1997; 11(7):824–834. [PubMed: 9106655]
51. Harreman MT, Kline TM, Milford HG, Harben MB, Hodel AE, Corbett AH. Phosphorylation of nuclear import by phosphorylation adjacent to nuclear localization signals. J Biol Chem. 2004; 279(20):20613–20621. [PubMed: 14998990]
52. Sambrook, J.; Fritsch, EF.; Maniatis, T. Molecular Cloning: A Laboratory Manual. Second ed. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, New York: 1989.

53. Adams, A.; Gottschling, DE.; Kaiser, CA.; Stearns, T. Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press; Cold Spring Harbor: 1997.
54. Yamaguchi-Iwai Y, Sonoda E, Buerstedde JM, Bezzubova O, Morrison C, Takata M, Shinohara A, Takeda S. Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in RAD52. Mol Cell Biol. 1998; 18(11):6430–6435. [PubMed: 9774659]
55. Benson D, Karsch-Mizrachi I, Lipman D, Ostell J, Sayers E. GenBank. Nucleic Acids Res. 2009; 37(Database Issue):D26–31. [PubMed: 18940867]
56. Boeke J, Garfinkel D, Styles C, Fink G. Ty elements transpose through an RNA intermediate. Cell. 1985; 40(3):491–500. [PubMed: 2982495]
57. Truant R, Fridell RA, Benson RE, Bogerd H, Cullen BR. Identification and functional characterization of a novel nuclear localization signal present in the yeast Nab2 poly(A)+ RNA binding protein. Mol Cell Biol. 1998; 18(3):1449–1458. [PubMed: 9488461]
58. Winston F, Dollard C, Ricupero-Hovasse SL. Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. Yeast. 1995; 11(11):53–55. [PubMed: 7762301]
59. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics. 1989; 122:19–27. [PubMed: 2659436]
60. Depienne C, Mousnier A, Leh H, Le Rouzic E, Dormont D, Benichou S, Dargemont C. Characterization of the nuclear import pathway for HIV-1 integrase. J Biol Chem. 2001; 276(21): 18102–18107. [PubMed: 11278458]
61. Green DM, Marfatia KA, Crafton EB, Zhang X, Cheng X, Corbett AH. Nab2p Is required for poly(A) RNA export in *Saccharomyces cerevisiae* and is regulated by arginine methylation via Hmt1p. J Biol Chem. 2002; 277(10):7752–7760. [PubMed: 11779864]
62. Leung SW, Harreman MT, Hodel MR, Hodel AE, Corbett AH. Dissection of the karyopherin alpha nuclear localization signal (NLS)-binding groove: functional requirements for NLS binding. J Biol Chem. 2003; 278(43):41947–41953. [PubMed: 12917403]

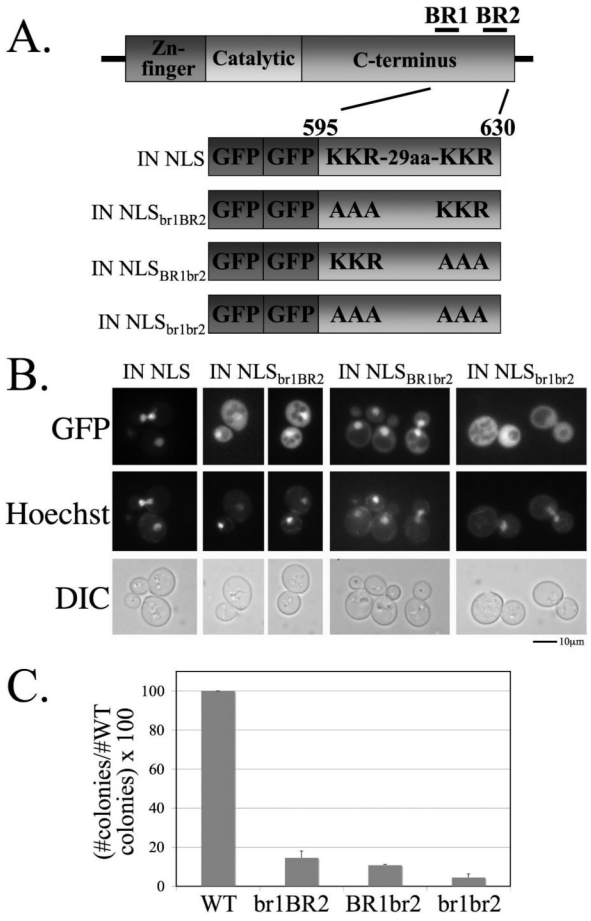

Figure 1. Both basic regions, BR1 and BR2, are required for proper nuclear localization and retrotransposition of Ty1 integrase

A) Tools for investigating the requirements for nuclear localization of Ty1 IN NLS. The last 55 amino acids of Ty1 IN containing both BR1 and BR2 were fused to two tandem GFP molecules as shown. Using site-directed mutagenesis, alanine residues were substituted for the KKR sequence of BR1 (IN NLS_{br1BR2}), BR2 (IN NLS_{BR1br2}), or in combination (IN NLS_{br1br2}). B) Localization of GFP_2-IN NLS variants. Wildtype cells expressing GFP_2-IN NLS variants were analyzed by direct fluorescence microscopy. Hoechst dye was used to visualize the location of the nucleus. Corresponding DIC images are shown. C) As assessed using a previously described retrotransposition assay (56), retrotranposition is significantly decreased in Ty1 NLS br1BR2 and Ty1 NLS BR1br2 mutants. Recombination deficient cells (*Δrad52)* were transformed with the pAR100 test plasmid, which contains the either the wildtype Ty1 retrotransposon or Ty1 NLS variants (Figure 1A). The number of retrotransposition events obtained for wildtype integrase was set to 100%. Results were quantitated as described in Materials and Methods.

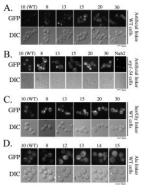

Figure 2. Localization of bipartite SV40 T3 NLS reporters containing various linker lengths

A) Wildtype cells expressing bpSV40 T3 NLS reporters containing artificial linkers of various lengths (number of amino acids in linker) were analyzed by direct fluorescence microscopy. Corresponding DIC images are shown. B) Importin-α mutant *srp1-54* cells expressing bpSV40 T3 NLS reporters containing various linker lengths were analyzed by direct fluorescence microscopy. Nab2-GFP was localized as a control that is imported into the nucleus through a mechanism that does not depend on importin-α (35,57). Corresponding DIC images are shown. C) Wildtype cells expressing bpSV40 T3 NLS reporters containing serine/glycine linkers of various lengths (number of amino acids in linker) were analyzed by direct fluorescence microscopy. Corresponding DIC images are shown. D) Wildtype cells expressing bpSV40 T3 NLS reporters containing alanine linkers of various lengths (number of amino acids in linker) were analyzed by direct fluorescence microscopy. Corresponding DIC images are shown.

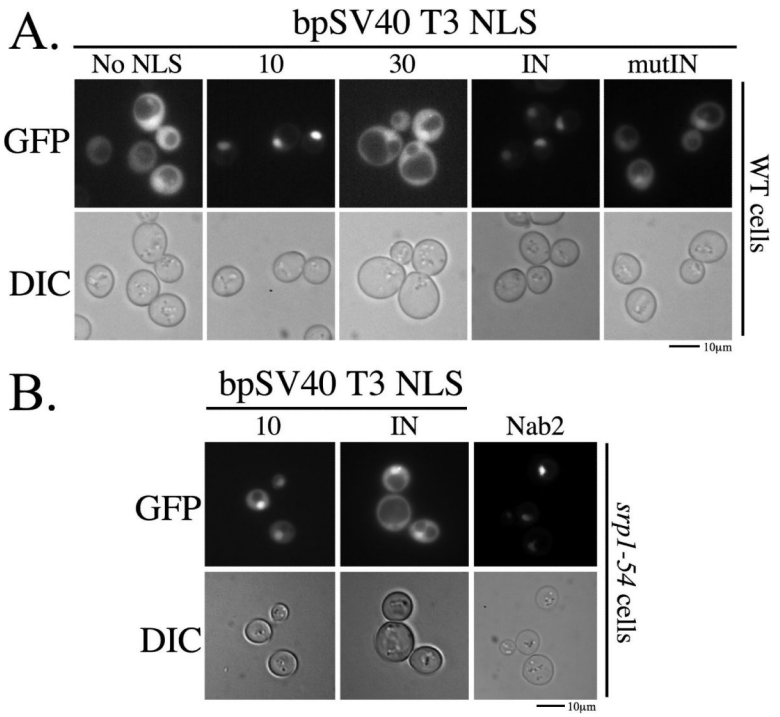

Figure 3. The 29 amino acid linker from integrase is a functional bipartite linker sequence

A) Wildtype cells expressing GFP_2 (control), bpSV40 T3 NLS with the conventional 10 aa linker (10), a long 30 aa linker (30), the 29 aa integrase linker (IN), or a variant of the integrase linker (mutIN) were analyzed using direct fluorescence microscopy. Corresponding DIC images are shown. B) Importin-α mutant *srp1-54* cells expressing bpSV40 T3 NLS (10), bpSV40 T3 NLS (IN), or control Nab2-GFP were analyzed using direct fluorescence microscopy. Corresponding DIC images are shown.

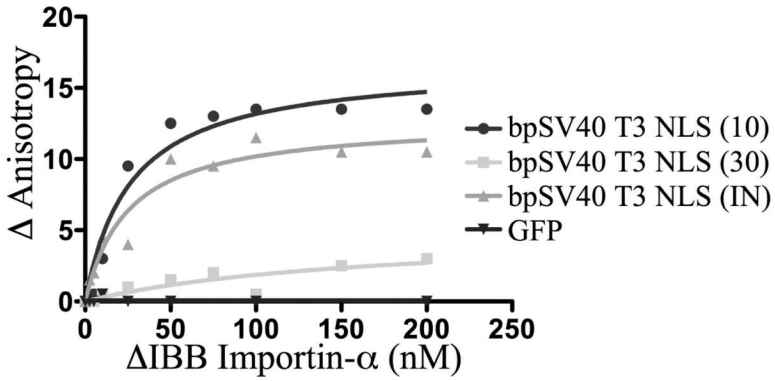

Figure 4. The 29 amino acid linker of integrase mediates binding to importin-α through a bipartite mechanism

Binding of purified ΔIBB-importin-α to bpSV40 T3 NLS cargoes was assessed using a quantitative *in vitro* binding assay (39). Purified GFP, bpSV40 T3 NLS-GFP (10), (30) or (IN) proteins were incubated with increasing concentrations of ΔIBB-importin-α and the change in absorbance (Δ Anisotropy) of the samples at 450 nm was measured. Using Prism5 software (GraphPad Software, Inc.), we generated non-linear regression binding curves to yield the K_D values provided in Table 2.

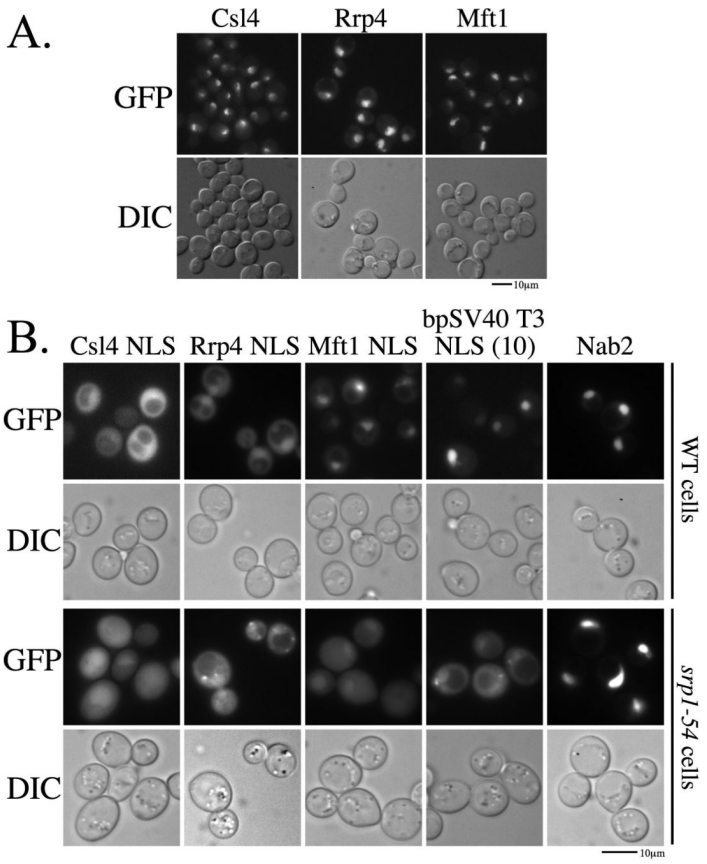

Figure 5. Analysis of putative bipartite cNLSs within candidate cellular proteins

A) Wildtype cells expressing integrated Csl4-GFP, Rrp4-GFP, or Mft1-GFP were analyzed by direct fluorescence microscopy. Corresponding DIC images are shown. B) Wildtype or *srp1-54* cells expressing Csl4 NLS-GFP_2, GFP_2-Rrp4 NLS, Mft1 NLS-GFP_2, bpSV40 T3 NLS-GFP_2 (10), or Nab2-GFP were analyzed using direct fluorescence microscopy. Corresponding DIC images are shown.

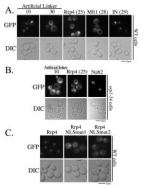

Figure 6. The bipartite cNLS within Rrp4 is necessary for nuclear import

A) Wildtype cells expressing bpSV40 T3 NLS (10), bpSV40 T3 NLS (30), bpSV40 T3 NLS (IN), or bpSV40 T3 NLS reporters containing the putative long bipartite cNLS linkers from Rrp4 or Mft1 were analyzed using direct fluorescence microscopy. Corresponding DIC images are shown. B) Importin-α mutant *srp1-54* cells expressing bpSV40 T3 NLS (10), bpSV40 T3 NLS (Rrp4), or Nab2-GFP (control) were analyzed using direct fluorescence microscopy. Corresponding DIC images are shown. C) Wildtype Rrp4 or Rrp4 containing amino acid substitutions within the putative bipartite cNLS (K122A/R123A, NLSmut1 or R149A/R150A/K151A, NLSmut2) was localized as a GFP-fusion protein in wildtype cells using direct fluorescence microscopy. Corresponding DIC images are shown.

Table 1

Artificial linker sequences contained within the in vivo *bpSV40 T3 NLS-GFP$_2$ reporter proteins*

Linker Length (aa)	Linker Sequence (KR—linker—KTKRKV)
8	TADEFESP
10 (WT)	TADGSEFESP
13	TADGSGSSEFESP
15	TADGSTADGSEFESP
20	TADGSEFESATADGSEFESP
30	TADGSEFESATADGSEFESPTADESEFESP
IN linker (29 aa)	SLEDNETEIKVSRDTWNTKNMRSLEPPRS
mutIN linker (29 aa)	SL**QN**N**Q**T**Q**IKVSRDTWNTKNMRSLEPPRS

Table 2

Dissociation constants (K_D) for various NLS-GFP cargoes for importin-α

NLS-GFP cargo	K_D (nM)
bpSV40 T3 NLS (10 aa)	27
bpSV40 T3 NLS (30 aa)	Too weak to measure
bpSV40 T3 NLS (IN)	25

Table 3

PSORTII results from the S. cerevisiae *proteome of predicted bipartite cNLS-containing proteins of various linker lengths*

Predicted Bipartite Linker Length	Proteomic Prevalence	
8-9	553	9.1%
10	968	16.5%
11-13	674	11.5%
14-15	285	4.9%
16-20	1018	17.4%
20-30	1481	25.3%

Table 4

Putative bipartite cNLS sequences within PSORTII-identified candidates

Protein	Putative cNLS Sequence	Predicted Linker Length (aa)
Csl4	^{100}**RR**TVKNILVSVLPGTEKG**RKTNK**^{122}	16
Rrp4	^{122}**KR**WKVDIGGKQHAVLMLGSVNLPGGIL**RRK**^{131}	25
Mft1	^{247}**KR**DGLLNEAEGDNIDEDYESDEDEERKERF**KRQR**^{280}	28

Table 5

Strains and plasmids used

Strains/Plasmids	Description	Reference
ACY192	*mat a ura3-52, leu2-1, trp1-63*	(58)
BY4741	*mat a ura3-52, leu2-1, trp1-63, his3-1*	
ACY1563 (NOY672)	*mat a ura3-1, leu2-3, trp1-1, his3-11, can1-100 srp1-54*	(33)
ACY1956	*mat a ura3-52, leu2-1, trp1-63, his3-1, RAD52::KANMX*	
pRS316	*URA3 CEN AMP*	(59)
pAR100 (pAC1735)	*pGAL-Ty1-HIS3 URA3 CEN AMP*	(60)
pAC719	Nab2-GFP *URA3 2μ AMP*	(61)
pAC781	c-fus-GFP *KAN* pET28a-based	(40)
pAC1059	*pMET25*-bpSV40 T3 NLS-GFP_2 *URA3 CEN AMP*	(62)
pAC1069	*pMET25*-GFP_2 *URA3 CEN AMP*	(32)
pAC1481	bpSV40 T3 NLS (10)-GFP *KAN* pET28a-based	(62)
pAC1804	*pMET25*-GFP_2-*IN NLS URA3 CEN AMP*	(28)
pAC2355	*pMET25*-GFP_2-*IN* NLS_{br1BR2} *URA3 CEN AMP*	This study
pAC2357	*pMET25*-GFP_2-*IN* NLS_{BR1br2} *URA3 CEN AMP*	This study
pAC2406	bpSV40 T3 NLS (30)-GFP *KAN* pET28a-based	This study
pAC2407	*pMET25*-bpSV40 T3 NLS (8)-GFP_2 *URA3 CEN AMP*	This study
pAC2408	*pMET25*-bpSV40 T3 NLS (13)-GFP_2 *URA3 CEN AMP*	This study
pAC2409	*pMET25*-bpSV40 T3 NLS (15)-GFP_2 *URA3 CEN AMP*	This study
pAC2410	*pMET25*-bpSV40 T3 NLS (20)-GFP_2 *URA3 CEN AMP*	This study
pAC2411	*pMET25*-bpSV40 T3 NLS (30)-GFP_2 *URA3 CEN AMP*	This study
pAC2429	*pGAL-Ty1* IN_{br1BR2}-*HIS3 URA3 CEN AMP*	This study
pAC2431	*pMET25-*GFP_2-*IN* NLS_{br1br2} *URA3 CEN AMP*	(28)
pAC2446	ΔIBB-Srp1 *KAN* pET30a(+)-based	This study
pAC2489	*pGAL-Ty1* IN_{BR1br2}-*HIS3 URA3 CEN AMP*	This study
pAC2534	*pGAL-Ty1* IN_{br1br2}-*HIS3 URA3 CEN AMP*	This study
pAC2577	*pMET25*-bpSV40 T3 NLS (Rrp4)-GFP_2 *URA3 CEN AMP*	This study
pAC2578	*pMET25*-bpSV40 T3 NLS (Mft1)-GFP_2 *URA3 CEN AMP*	This study
pAC2657	bpSV40 T3 NLS (IN)-GFP *KAN* pET28a-based	This study
pAC2667	*pMET25*-bpSV40 T3 NLS (IN)-GFP_2 *URA3 CEN AMP*	This study
pAC2774	*pMET25*-bpSV40 T3 NLS (8 Ser/Gly)-GFP_2 *URA3 CEN AMP*	This study
pAC2775	*pMET25*-bpSV40 T3 NLS (10 Ser/Gly)-GFP_2 *URA3 CEN AMP*	This study
pAC2776	*pMET25*-bpSV40 T3 NLS (13 Ser/Gly)-GFP_2 *URA3 CEN AMP*	This study
pAC2777	*pMET25*-bpSV40 T3 NLS (15 Ser/Gly)-GFP_2 *URA3 CEN AMP*	This study
pAC2778	*pMET25*-bpSV40 T3 NLS (20 Ser/Gly)-GFP_2 *URA3 CEN AMP*	This study
pAC2779	*pMET25*-bpSV40 T3 NLS (30 Ser/Gly)-GFP_2 *URA3 CEN AMP*	This study
pAC2780	*pMET25*-bpSV40 T3 NLS (8 Ala)-GFP_2 *URA3 CEN AMP*	This study
pAC2781	*pMET25*-bpSV40 T3 NLS (10 Ala)-GFP_2 *URA3 CEN AMP*	This study

Strains/Plasmids	Description	Reference
pAC2782	*pMET25*-bpSV40 T3 NLS (13 Ala)-GFP_2 *URA3 CEN AMP*	This study
pAC2783	*pMET25*-bpSV40 T3 NLS (15 Ala)-GFP_2 *URA3 CEN AMP*	This study
pAC2787	*pMET25*-Csl4 NLS (aa 97-125)-GFP_2 *URA3 CEN AMP*	This study
pAC2788	*pMET25*-Mft1 NLS (aa 244-283)-GFP_2 *URA3 CEN AMP*	This study
pAC2789	*pMET25*-bpSV40 T3 NLS (mutIN)-GFP_2 (E601Q/D602N/E604Q/E606Q) *URA3 CEN AMP*	This study
pAC2790	*pMET25*-bpSV40 T3 NLS (14 Ala)-GFP_2 *URA3 CEN AMP*	This study
pAC2793	Rrp4-GFP *URA3 CEN AMP*	This study
pAC2797	Rrp4-GFP NLSmut1 (K122A/R123A) *URA3 CEN AMP*	This study
pAC2798	Rrp4-GFP NLSmut2 (R149A/R150A/K151A) *URA3 CEN AMP*	This study
pAC2809	*pMET25*-bpSV40 T3 NLS (12 Ala)-GFP_2 *URA3 CEN AMP*	This study
pAC2815	*pMET25*-GFP_2-Rrp4 NLS (aa 111-159) *URA3 CEN AMP*	This study