Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin α*

Received for publication, September 10, 2008, and in revised form, November 10, 2008 Published, JBC Papers in Press, November 10, 2008 DOI 10.1074/jbc.M807017200

Shunichi Kosugi‡§1, Masako Hasebe‡, Nobutaka Matsumura‡, Hideaki Takashima‡, Etsuko Miyamoto-Sato‡, Masaru Tomita§, and Hiroshi Yanagawa‡§2

From the §Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan and the ¶Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University, Yokohama 223-8522, Japan

The importin α/β pathway mediates nuclear import of proteins containing the classical nuclear localization signals (NLSs). Although the consensus sequences of the classical NLSs have been defined, there are still many NLSs that do not match the consensus rule and many nonfunctional sequences that match the consensus. We report here six different NLS classes that specifically bind to distinct binding pockets of importin α. By screening of random peptide libraries using an mRNA display, we selected peptides bound by importin α and identified six classes of NLSs, including three novel classes. Two noncanonical classes (class 3 and class 4) specifically bound the minor binding pocket of importin α, whereas the classical monopartite NLSs (class 1 and class 2) bound to the major binding pocket. Using a newly developed universal green fluorescent protein expression system, we found that these NLS classes, including plant-specific class 5 NLSs and bipartite NLSs, fundamentally require the regions outside the core basic residues for their activity and have specific residues or patterns that confer the activities differently between yeast, plants, and mammals. Furthermore, amino acid replacement analyses revealed that the consensus basic patterns of the classical NLSs are not essential for activity, thereby generating more unconventional patterns, including redox-sensitive NLSs. These results explain the causes of the NLS diversity. The defined consensus patterns and properties of importin α-dependent NLSs provide useful information for identifying NLSs.

Nuclear localization signals (NLSs) are short peptide motifs that mediate the nuclear import of proteins by binding to their receptors, known as importins (karyopherins), and are one of the most characterized motifs (1–5). Nuclear import of proteins is generally initiated by the formation of a ternary complex with importin α, importin β1, and a cargo, where importin β1 docks the complex to the NPC to release the cargo into the nucleus through the binding of Ran-GTP to importin β1 (6,7). An asymmetric distribution of Ran-GTP and Ran-GDP between the nucleus and the cytoplasm controls the cargo import and export, and this gradient is maintained by various Ran-associated regulatory factors (8).

In the importin α/β pathway, importin α serves as an adaptor that links cargos and importin β1 and recognizes NLSs within the cargos. Importin α recognizes two classes of NLSs, known as classical NLSs: monopartite NLSs having a single cluster of basic amino acid residues and bipartite NLSs having two clusters of basic amino acids separated by a 10–12-amino acid linker (5). Further, there are two types of monopartite NLSs; one has at least four consecutive basic amino acids, exemplified by the SV40 large T antigen NLS (PKKKRKV) (9), whereas the other has only three basic amino acids and is represented by X(K/R)3/5(K/R) as a putative consensus sequence. The latter is exemplified by the c-Myc NLS (PAAKR/KLD) (10). In this study, the two types of the monopartite NLS have been designated as “class 1” and “class 2” NLSs, respectively. A putative consensus sequence of the bipartite NLS has been defined as (K/R)(K/R)X10–12(K/R)3/5 where (K/R)3/5 represents at least three of either lysine or arginine of five consecutive amino acids, in which the linker region has been found to be tolerant to amino acid conversion (11,12).

Although the putative consensus sequences of the classical NLSs have been defined, there are a number of experimentally defined NLSs that do not match the consensus sequences, and furthermore, there is likely to be a great number of nonfunctional sequences that match the consensus. Importin α possesses two distinct NLS binding sites, and the major NLS binding site binds to the classical monopartite NLSs whereas both sites are required for binding to bipartite NLS (13–16). It is likely that the presence of two distinct binding sites in a single importin α molecule generates a number of NLS patterns. In addition, monopartite NLSs require specific residues flanking the core basic cluster for their complete activity (17–20), which further causes the sequence diversity of NLSs.

In this study, we have collected various types of NLS by screening random peptide libraries and identified six different classes of NLSs, including noncanonical ones that bind specifically to the minor binding site of importin α. The results dem-
 demonstrate that a single importin α can bind more diverse patterns of NLS than previously recognized.

**EXPERIMENTAL PROCEDURES**

**Plasmid Construction**—The cDNA fragments of importins were PCR-amplified and cloned to pGEX-6p-1 (Amersham Biosciences) or pMAL-c2. Universal green fluorescent protein (GFP) expression constructs were generated using DNA fragments encoding an enhanced GFP from pQBI25-fC1 (Qbiogene, CA) and a tTA transactivator from pTet-off (Clontech). For more details, see the supplemental material.

**Production of Recombinant Proteins**—Recombinant importin proteins fused with glutathione S-transferase (GST) or maltose-binding protein were expressed in Escherichia coli BL21 and purified with glutathione-Sepharose 4B (Amersham Biosciences) or amylose resin (New England Biolabs). For more details, see the supplemental material.

**Construction of Random Peptide-mRNA/cDNA Fusion Libraries**—IVV libraries containing randomized peptides were prepared with oligonucleotides (supplemental Table S1). Double-stranded oligonucleotides containing SP6 promoter were *in vitro* transcribed, ligated to a polyethylene glycol-puromycin spacer (21, 22), and *in vitro* translated. The resulting peptide-mRNA fusions were converted to peptide-mRNA/cDNA by reverse transcription. For more details, see supplemental material.

**Library Screening**—The random peptide IVV library (40 pmol) of cDNA/mRNA-peptide fusion was diluted in 60 µl of binding buffer (20 mM Hepes-NaOH (pH 7.8), 0.15 mM NaCl, 1 mM EDTA, 0.2% Nonidet P-40, 0.5% skim milk) and incubated for 30 min with the rice, human, and yeast importin α (20 µg) fused with GST, which had been bound to 20–25 µl of glutathione beads packed in a minicolumn made of a 200-µl pipette tip. For the rice maltose-binding protein-importin α (10 µg) the incubation with the library was conducted in solution for 30 min at room temperature, and the solution was repeatedly passed through a 200-µl pipette tip column containing 25 µl of amylose resin. These columns were sequentially washed with 200 µl of binding buffer, 200 µl of washing buffer (20 mM Hepes-NaOH (pH 7.8), 0.1 mM NaCl, 1 mM EDTA, 0.2% Nonidet P-40), and 200 µl of washing buffer lacking NaCl and eluted with 30 µl of 20 mM glutathione and 20 mM maltose. The volume of the washing buffer was increased to 600–800 µl with each cycle of the screening. The selected DNAs in the eluate were PCR-amplified with SP6-SD2 and Flag2 primers (supplemental Table S1) and HotStarTaq DNA polymerase (Qiagen). The amplified DNA fragments were purified with a QIAquick PCR purification kit (Qiagen) and used as a library for the subsequent round of screening. As the screening cycle was increased, the amount of library and the amplification cycles of PCR were decreased to 10–20 pmol and 23–25 cycles, respectively. Finally, the amplified DNAs were digested with BamHI and XbaI and cloned into the pTUE-GFP3 vector for sequencing and transfection assay. At least 10 clones from each screening round (except for the first round) were sequenced and assayed for nuclear import activity. The optimal round was selected to satisfy both high sequence diversity and high rate of functional NLS in the pool from the respective rounds.

**Cell Culture and Transfection and GFP Observation**—The mouse fibroblast cell line NIH3T3 or the HeLa cell line was grown at 37 °C under 5% CO₂ in Dulbecco’s modified Eagle’s medium for NIH3T3 cells and Eagle’s minimum essential medium for HeLa cells, each supplemented with 10% calf serum and 2 mM l-glutamine. Transfections were performed in 5 × 10⁴ cells/ml in 24-well plates or 35-mm culture dishes, and the cells were cotransfected with ~0.5 µg of a GFP reporter plasmid and 0.3 µg of a pTet-Off plasmid (Clontech) using 2 µl of jet-PEI reagent (PolyPlus-transfection, Strasbourg, France) according to the manufacturer’s instruction and cultured for 36–48 h.

For plant cells, suspension-cultured tobacco cells were established from calli of *Nicotiana tabacum* cv. Sunspan NN and maintained as previously described (23). The cultured tobacco cells were blotted on 5-cm filter paper discs and transfected with plasmid DNA by microprojectile bombardment (23), except that 1.2 mg of ~1-µm tungsten particles were coated with 2 µg each of a GFP reporter and 4 µg of a p35S-tTA plasmid. Bombarded cells were cultured at 28 °C for 20 h in the dark.

Yeast manipulations, including culture and transformation, were performed as described (23), and strain SFY526 (Clontech) was used in most experiments. SFY526 was cotransformed with 0.5 µg each of a GFP reporter and a pGBK-tTA plasmid and cultured with the SD synthetic medium lacking leucine and tryptophan in 48-well plates at 30 °C. The yeast expressing tTA grew very slowly, so the transformed cells were cultured for at least 3 days. GFP expressed in these transfected cells was observed using an epifluorescence microscope, model BX51 (Olympus), with an excitation filter specific to 460–490 nm.

**Importin Binding Assay**—The purified GST-importin fusion proteins (2.0 or 3.0 µg) were incubated with the thioredoxin (Trx)-GFP-NLS proteins (2.0 or 3.0 µg each of a GFP reporter and a pGBK-tTA plasmid and cultured with the SD synthetic medium lacking leucine and tryptophan in 48-well plates at 30 °C. The yeast expressing tTA grew very slowly, so the transformed cells were cultured for at least 3 days. GFP expressed in these transfected cells was observed using an epifluorescence microscope, model BX51 (Olympus), with an excitation filter specific to 460–490 nm.

**RESULTS**

**Five Classes of Monopartite NLSs Identified by mRNA Display Screening with Importin α**—To identify many examples of functional monopartite NLSs, we screened random peptide libraries by an *in vitro* virus (IVV) method with different importin α variants as bait (Fig. 1, A and B). IVV is an mRNA display technique for the production of a protein covalently linked to its encoding mRNA, and it is more suitable for screening a large library as compared with other conventional methods (22, 24–26). Using importin α variants (Osimp α1 and Hsimp α3) lacking the N-terminal importin β binding (IBB) domain from rice importin α1a and human importin α3, which comprise the major subgroups of the importin α family (27, 28), we selected 218 and 253 sequences from 12X and XBBX IVV libraries,
which could encode monopartite NLSs, after five and six rounds of screening, respectively. Many of the selected sequences exhibited a significant level of nuclear import activity and could be classified into five classes of monopartite NLSs when aligned (Fig. 2). Unexpectedly, the sequences selected from the 12X and XBBX libraries contained many unconventional NLSs with only two or three consecutive basic amino acids, and these could be divided into three classes, designated classes 3–5 (Fig. 2). A native NLS observed class 3 NLS was observed at the C terminus of the mammalian nucleolar RNA helicase II/Gu involved in rRNA synthesis and was confirmed to be responsible for its nuclear import (Fig. 3). In contrast to classes 3 and 4, class 5 NLS was functional only in plant cells. An in vitro binding assay with recombinant importin α/H9251 variants from different species showed that a class 5 NLS strongly interacted only with a rice importin α, whereas the classical NLSs and class 3 and 4 NLSs bound to most importin α variants from yeast, rice, and humans, which indicates that class 5 NLS is plant-specific (Fig. S1).

To collect many more kinds of noncanonical NLSs, we extracted 29 unconventional NLSs from over 400 sequences previously reported in the literature. These unconventional NLSs had relatively short stretches of residues and appeared not to match the NLSs defined in this study. Unexpectedly, only three of these NLSs had a strong or medium level of activity, suggesting that the other 26 sequences may require additional sequences for complete activity (data not shown). Mutational analyses revealed that these three NLSs (BDV-P1-NLS, RHA-NLS, and HCV-NS5A-NLS) were unconventional class 2 NLSs.
that had a stretch of activating flanking residues (see Fig. 6C) (described below). This result indicates that most of the importin α-dependent monopartite NLS sequences are categorized into the five NLS classes defined in this study.

Class 3 and Class 4 NLSs Bind to the Minor Binding Groove of Importin α—Importin α contains two NLS binding grooves, the major and minor NLS binding sites, each located at the N-terminal Armadillo (Arm) repeat 2–4 and C-terminal Arm repeat 7–8 (Fig. 4A). We examined which of the two NLS binding grooves of importin α makes specific contact with class 3 and class 4 NLSs. We created a ΔIBB variant (Kapo) of Kap60p, an importin α of budding yeast, and its mutants (Kapo mut1 and Kapo mut2, each mutated at the major and minor NLS-binding sites) (29). An in vitro binding assay with these importin variants and GFP fusions with various classes of NLS showed that class 1 NLS (SV40 NLS) was specifically bound to the major binding site of importin α, whereas the bipartite NLS (NP-NLS) was bound to the both sites (Fig. 4B); this concurred with previous findings (13, 30). In contrast, class 3 (KRSWSMAF) and class 4 (KRKYF) NLSs exhibited specific binding to the minor binding site of importin α, although the binding of the class 4 sequence was considerably weak because of a lack of its N-terminal Arg or Pro (see Fig. 5). These results explain the reason underlying the considerable difference in sequence between class 1/2 and class 3/4 NLS subgroups, although they bind to the same importin α species.

Determination of Relative Levels of NLS Activity in Different Species—Although the importin α/β pathway is a well conserved system in eukaryotes, identification of the plant-specific class 5 NLS suggests that NLS specificities differ among species. Human and Arabidopsis possess six and eight importin α subtypes, respectively, which contain the conserved Arm repeat domain that is directly involved in NLS recognition. The Arm domains of human and Arabidopsis importin α subtypes exhibit 51–90% (average of 61%) and 45–87% (average of 66%) amino acid identities within species, respectively and 34–63% (average of 52%) identity between the species (Fig. S2). This suggests that the differences in NLS specificities among importin α subtypes of different species are likely to be larger than those of same-species importin α subtypes. Thus, we examined the specificities of various patterns and classes of NLS from different species.

To effectively assess the nuclear import activity of peptide sequences in different species, we developed a universal GFP expression system that functioned in budding yeast, higher plants, and mammals (Fig. 1C). This cotransfection system comprised a universal GFP reporter plasmid and a species-specific expression plasmid for the tTA transactivator, which is a

![FIGURE 3. Nuclear import of nucleolar RNA helicase II/Gu depending on class 3 NLS.](image)

![FIGURE 4. Class 3 and class 4 NLSs bind to the minor binding groove of importin α.](image)
Nonclassical NLSs Recognized by Importin \( \alpha \)

| Class 2 | Class 2 | Class 4 |
|---------|---------|---------|
| Y P M  | Y P M  | Y P M  |
| RAAKRPRRT | 10 10 9 | PPAAERPLO | 10 10 8 |
|         | 9 9 9  | A* E | 9 9 6  | A* E* | 9 9 5  | A* E* |
|         | 8 8 8  | A* E | 7 7 7  | A* E* | 8 8 8  | A* E* |
|         | 7 7 7  | A* E | 7 7 7  | A* E* | 7 7 7  | A* E* |
|         | 6 6 6  | A* E | 6 6 6  | A* E* | 6 6 6  | A* E* |
|         | 5 5 5  | A* E | 5 5 5  | A* E* | 5 5 5  | A* E* |
|         | 4 4 4  | A* E | 4 4 4  | A* E* | 4 4 4  | A* E* |
|         | 3 3 3  | A* E | 3 3 3  | A* E* | 3 3 3  | A* E* |
|         | 2 2 2  | A* E | 2 2 2  | A* E* | 2 2 2  | A* E* |
|         | 1 1 1  | A* E | 1 1 1  | A* E* | 1 1 1  | A* E* |
|         | 0 0 0  | A* E | 0 0 0  | A* E* | 0 0 0  | A* E* |

**TABLE 1**

Consensus sequences of six classes of importin \( \alpha \)-dependent NLS

| NLS class | Consensus sequence |
|-----------|-------------------|
| Class 1   | KR/K(R/R), K/K/RK |
| Class 2   | (P/R)XKR/K(R)/DE |
| Class 3   | KKR/W/F/YXXAF    |
| Class 4   | (R/P)XKR/K(R)/DE |
| Class 5   | LGKR(K/R)/W/F/Y  |
| Bipartite | KRX_{10–12}K(KR)   |
|           | KRX_{10–12}K(KR)(KR) |

*Sequence representation is as follows: (\( \alpha \)), any amino acid except Asp or Glu; \( X_{10–12} \), any 10–12 amino acids.

*For more optimal patterns, acidic residues should be rich in the central linker region and arginine in the terminal linker region, whereas basic and hydrophobic residues should be rare in the central region and proline-rich in the terminal region.

tetracycline repressor (TetR) fused with the herpes simplex virus VP16 activation domain. The universal reporter pTUE-GFP3 encodes an enhanced GFP fused with bacterial virus VP16 activation domain. The universal reporter pTUE-GFP3 and assay the nuclear import activities as one of the 10 levels (alternatively four levels). We cloned DNA fragments encoding peptides into pTUE-GFP3 and assayed the nuclear import activities of various NLSs, we ranked the relative levels of the nuclear import activities as one of the 10 levels (alternatively four levels) based on the localization phenotype of the GFP reporter, as explained in detail in Fig. S3. In this analysis, however, we cannot exclude the possibility that some of the observed nuclear and cytoplasmic phenotypes of the GFP fusions could be due to nuclear export signals or nuclear/cytoplasmic retention signals potentially contained in fused peptides (31, 32). Since these signal sequences are generally short, some screened and mutated NLS sequences could overlap with either of these signal sequences to affect their NLS activities.

**Optimal Consensus Patterns for Each NLS Class Determined by Amino Acid Replacement Analyses**—Using this GFP expression system, we conducted amino acid replacement analyses for each class of NLS and determined relative levels of the activities in the three organisms. The analyses revealed that flaming residues and basic patterns influence NLS activity, depending on the organisms assayed (Fig. 5 and Fig. S4). For class 1 NLSs, the nonarbitrary nature of the basic pattern is consistent with the biased basic patterns observed for the selected class 1 NLSs (Fig. 2 and Fig. S4). Similar amino acid biases were observed for 189 class 2 NLSs selected by screening of a BBXB library. Most of the selected class 2 NLSs contained KRX(K/R) as a basic core, and the activity levels varied depending on the flaming residues and the organisms assayed (Fig. S5). For other noncanonical classes, the basic and flaming patterns conserved in the selected NLSs were crucial for their NLS activities. These results confirm previous observations regarding the regulatory roles of flaming sequences of the classical NLSs and furthermore demonstrate that NLS activities can significantly vary, depending on the basic patterns, flaming residues, and organisms assayed. Although these results indicate that it is difficult to determine specific consensus sequences for each NLS class, optimal consensus patterns for each NLS class are suggested to be KR/K(R)/R or K/K/RK for class 1, (P/R)XKR/K(R)/DE for class 2, KRX(W/F/Y)XXAF for class 3, (R/P)XKR/K(R)/DE for class 4, and LGKR(K/R)/W/F/Y for class 5, where \( \alpha \) represents any amino acid except for Asp or Glu (Table 1).

**The Core Basic Residues of the Classical Class 2 NLSs Can Be Replaced by Hydrophobic Amino Acids, and KRX-containing NLSs Exhibit Redox Sensitivity**—In a class 2 NLS with a highly activating flaming sequence, the basic core residues at positions +2 and +4 could be replaced with several hydrophobic residues and asparagines (Fig. 6A), and this feature was most prominent in mammalian cells and was indeed found in several natural noncanonical NLSs from the Born disease virus P protein (BDV-PI) (33), the human RNA helicase A (RHA) (34), and the hepatitis C virus nonstructural protein (HCV-NS5A) (35) (Fig. 6C). In particular, cysteine and histidine at position +4 were comparable with arginine or lysine (Fig. 6A and B). This cysteine-substituted class 2 NLS was inactivated by cysteine-oxidizing reagents, diethyl maleate and tert-butylhydroquinone, and thereby, the first example of a redox-sensitive NLS was identified (Fig. 6D). Natural redox-sensitive NLSs were identified in human proteins including cell division cycle-associated protein 4 (Fig. 6E). These observations indicate that even classical NLSs are highly variable, and their fundamental core pattern is not restricted to the traditional consensus patterns.

**Consensus Patterns of Bipartite NLS**—A currently considered consensus sequence of bipartite NLS is also incomplete. To determine more reliable consensus sequences of bipartite NLS, we collected functional bipartite NLS sequences by screening three different IVV libraries. These peptide libraries (biX1, biX2, and biX3) contained different patterns of minimal bipartite basic residues separated by a linker sequence composed of 10–11 semirandom amino acids (Fig. 1B). The screening with...
Nonclassical NLSs Recognized by Importin α

The conserved basic pattern of the classical class 2 NLS is not essential in an NLS containing a highly activating flanking sequence. A, amino acid replacement analysis of the basic core residues of a class 2 monopartite NLS with a strong basal activity. The basic residues at positions +1, +2, and +4 of a class 2 NLS template sequence indicated at the top were replaced with other amino acids indicated in the left column. The mutated sequences were assayed for NLS activity in yeast, cultured tobacco, and NIH3T3 cells, and activity scores were represented as in Fig. S3. B, amino acid replacement analysis of class 2 NLS with a standard activity. In a class 2 NLS context with a standard basal activity, an arginine residue at a position +4 was exchangeable with cysteine and histidine in yeast (Y), tobacco (P), and NIH3T3 cells (M). It should be noted that the flanking sequences of the indicated NLSs have a substantially weaker activity than that indicated in A. C, noncanonical NLSs from the Borna disease virus P protein (BDV-P1), the human RNA helicase A (RHA), and the hepatitis C virus nonstructural protein (NS5A). These NLSs have been confirmed to be sufficient for localizing the GFP reporter to the nucleus. It should be noted that the hydrophobic residues highlighted at position +2 contribute to the activity in a sequence with a proline and/or arginine stretch in the N-terminal flanking region, as indicated in A. D, redox-sensitive NLS identified in noncanonical class 2 NLSs. Double-stranded oligonucleotides encoding the indicated NLSs, PAAKRPRLD (KRPR), PAAKRPCLD (KRPC), and PAAKRRLD (KRRC), and the plasmids were transfected to NIH3T3 cells. After 40 h of culture, the cells were treated with cysteine-oxidizing reagents (40), diethyl maleate (DEM), and tert-butylhydroquinone (tBHQ), each at 200 μM, and 0.1% DMSO as control, with serum-free Dulbecco’s modified Eagle’s medium for 1.5 h. Nuclei were indicated by arrowheads. E, nuclear localization of human proteins mediated by redox-sensitive NLSs. The cDNAs of cell division cycle-associated protein 4 (CDC4; AAH11736) and a hypothetical protein (Hyp1; CAH56177) were cloned into pCMV-dGFPN1 and assayed as in D. CDC4A K7Q and Hyp1 C28S amino acid substitution mutants for the predicted NLSs (CDC4A, 4-RGLKRKC; Hyp1, 23-PPKKRCL) were distributed throughout the cells, indicating that these redox-sensitive NLSs are the primary determinants for the nuclear import of these proteins. Nonclassical NLSs Recognized by Importin α

The sequences selected by different importins exhibited similar characteristic patterns in the N-terminal basic stretches, including KR, RKR, and RKH (Figs. S6 and S7), despite the fact that the RKR sequence arose from spontaneous frame shifts and/or nucleotide mutations during screening. To determine the N-terminal basic patterns required for NLS activity more definitively, the activity of bipartite NLSs was analyzed with variants containing different combinations of lysine and arginine in the N-terminal basic region (Fig. 7A and Fig. S9). Consistent with the screening results, the N-terminal basic patterns containing KR were required for a strong NLS activity in most cases and RKR was more effective than KR. A similar tendency was observed in 137 bipartite NLSs collected from the literature, of which 73 sequences (53%) contained KR in their N-terminal basic stretches (supplemental Table S2).

C-terminal Property of Bipartite NLS—Many biX3-derived NLSs were found to have a sequence KR or KK in their C-terminal basic regions (Fig. S6). To determine the sequence requirement of the C-terminal basic patterns more definitively, we generated bipartite NLS variants with different combinations of lysine and arginine within five residues of the C-terminal basic stretch (Fig. 7A and Fig. S10). The analysis for the NLS activity revealed that lysine at a position of +15, where the first lysine residue in the N-terminal basic stretch is numbered as +1, was critical for the activity (Fig. 7A and Fig. S10). An additional lysine and arginine downstream of position +15 enhanced the activity in a position- and residue-specific manner, and these C-terminal patterns had a more varied influence on the three species than the N-terminal patterns.

The compensatory interaction between the N-terminal and C-terminal region of the bipartite NLS was observed particularly in the screening result on the biX3 library (Fig. S8). Although the biX3 library has been designed to generate the class 2 monopartite NLS with a considerably weak activity at the C-terminal basic stretch, many of the selected NLS sequences generated the class 2 NLS with strong or moderate activities because of a frameshift and/or nucleotide mutations generated during screening. In agreement with these results, 44 sequences (32%) of the 137 bipartite NLSs from the literature contained the basic core of the class 2 monopartite NLS in their C-terminal regions (data not shown). In all of these NLSs, the position of X in the basic cores K(K/R)X(K/R) was not occupied by acidic residues, which severely impair the activity of the class 2 NLS at this position. These results suggest that the C-terminal basic stretch of bipartite three importin α variants (Osimp α1, Hsimp α3, and Kapa) as bait identified hundreds of functional NLSs that had various levels of activity (Figs. S6–S8).
The linker residues at the indicated positions were replaced with the indicated residues. Activity scores are represented as in Fig. S3.

Amino acid replacement analysis of bipartite NLSs. FIGURE 7. Amino acid replacement analysis of bipartite NLSs. A, amino acid exchange within the terminal basic regions. Two or three consecutive basic residues in the N-terminal and C-terminal basic regions of the indicated bipartite NLSs were replaced with Lys, Arg, or Thr to generate various basic patterns. The NLS mutants were assayed in yeast (Y), tobacco (P), and NIH3T3 (M). The asterisk represents unchanged residues. Activity scores are represented as in Fig. S3, A, amino acid exchange within the central linker region. The linker residues at the indicated positions were replaced with the indicated residues, and the resulting mutants were assayed as in A. NLS functions as the class 2 monopartite NLS with a weak or medium level of activity, thus producing the full NLS activity by incorporating activities from the N-terminal basic stretch and linker region.

Linker Region of Bipartite NLS—In the linker regions of the screened sequences with a strong NLS activity, proline and acidic amino acids were more abundant, and hydrophobic residues more rare than in the sequences with a weaker NLS activity. Bipartite NLSs collected from the literature also contained a greater number of proline and acidic residues and a smaller number of hydrophobic residues in their linker regions as compared with the naturally occurring amino acid frequency (supplemental Table S3). Outstandingly, many proline residues were located in the terminal regions, especially at positions +3, +11, +12, and +13, whereas acidic residues were located in the central region. Indeed, amino acid replacement analysis in the central linker region showed that acidic residues functioned as activator, whereas hydrophobic and basic residues functioned as repressor (Fig. 7B and Fig. S11A). These amino acid-specific effects were increased with increasing number of converted residues (Fig. 7B), indicating an additive contribution of these residues to the activity. These observations suggest that specific residues in the linker sequence make a substantial contribution to the NLS in a position-dependent manner. We then examined the effect on activity of the linker length using three template NLSs with different levels of activity (Fig. S11B). The linkers ranging from 10 to 16 residues provided a more effective spacing for the activity, whereas the linkers composed of 20 residues were functional only in the template with a high basal activity. This suggests that a long linker up to 20 residues is functional as long as a bipartite NLS is placed on a protein terminus, which could permit the NLS to form a flexible conformation.

From these observations, optimal activation patterns of the classical bipartite NLS were defined as KRX_{10–12}(K/R)(K/R) and KRX_{10–12}(K/R)(K/R) (Table 1). For more optimal patterns, it is preferred that acidic residues are rich in the central linker region and rare in the terminal linker region, whereas basic and hydrophobic residues are rare in the central region and proline is rich in the terminal region. However, there are many functional NLSs that only partially match these patterns in the selected and literature-derived NLS sets. It is likely that, in many bipartite NLSs, the activation patterns present in the N-terminal, C-terminal, or linker regions compensate for a weak activation pattern of the other region.

DISCUSSION

We identify three novel classes of importin α-dependent NLSs by high throughput screening of random peptide libraries. In addition to three classes of the known traditional NLSs, a total of at least six NLS classes mediate nuclear import of proteins through the importin α/β pathway. The presence of multiple NLS classes that can bind to a single importin α species is likely to cause the sequence diversity of NLS. Additionally, our mutational analyses demonstrate that many different residues throughout the entire region of each NLS class activate or repress NLS activity, depending on their position. Highly activating patterns in the flanking sequences of class 2 NLSs allow exchange between the core basic residues and several hydrophobic and asparagine residues.

In particular, the effect of cysteine at position +4 was comparable with that of basic residues, and the NLSs containing the K(K/R)X(C) pattern had reduced sensitivity. These observations suggest that the restricted but flexible exchangeability of the respective residues within an NLS is another cause of the NLS diversity. Moreover, although NLSs are thought to possess a common functionality in eukaryotes, we have frequently observed some degree of difference in the activities of many selected and mutated NLSs among yeast, plant, and mammalian cells. This is partly because of the fact that importin α subtypes within the same species possess shared but partly distinct NLS recognition specificities (29, 36–39). Together with the observed diversity of importin α subtypes between the species, our observation suggests that differences exist in the specificity of importin α subtypes between and within species, which is likely to be a third cause of the diversity of NLSs.

The novel types of monopartite NLS, class 3 and class 4, bind specifically to the minor binding site of importin α, correlating with the binding site of the N-terminal basic region of bipartite NLS. This indicates that a bipartite NLS comprises two imperfect monopartite NLSs, corresponding to the class 3/class 4 NLS at its N-terminal region and the class 1/class 2 NLS at its C-terminal region. In contrast, the linker region of the bipartite NLS has been reported to interact nonspecifically with importin α and to be tolerant to mutation (12, 14). Nevertheless, our mutational analyses demonstrate that the linker region substantially affects the NLS activity, depending on the residue, suggesting a functional contribution of the nonspecific interaction. A linker length of about 10–16 residues was effective but
also functional up to 20 residues in a sequence context bearing a high NLS activity, which is consistent with a previous observation (12). This high tolerance to a long linker length is, however, likely to be because of the structural flexibility of NLS peptides fused at the C terminus of the reporter protein in our analysis. Thus, for bipartite NLSs with a structurally restricted flexibility, an effective length of the linker would be 10–12 residues, which would be equal to the distance between the major and minor binding sites of importin α. Conversely, when the respective terminal regions of bipartite NLSs are placed to provide adequate space for binding to importin α on a surface of a protein or protein complex, it could become a functional NLS even if one terminal region is considerably distant from the other on the primary sequence. This may be the case for a dimerization- and conformation-dependent NLS seen in some STATs and zinc finger proteins.

This study defined optimal consensus sequences for each NLS class, but the level of NLS activity was found to vary depending on the flanking sequences and the organisms assayed. Thus, to obtain an accurate representation of the NLS consensus sequences, we would need to consider the contributions of all amino acid species at each position within an NLS.

REFERENCES
1. Görlich, D., and Kutay, U. (1999) Annu. Rev. Cell Dev. Biol. 15, 607–660
2. Mattaj, I. W., and Englmeier, L. (1998) Annu. Rev. Biochem. 67, 265–306
3. Chook, Y. M., and Blobel, G. (2001) Curr. Opin. Struct. Biol. 11, 703–715
4. Jans, D. A., Xiao, C. Y., and Lam, M. H. (2000) BioEssays 22, 532–544
5. Lange, A., Mills, R. E., Lange, C. J., Stewart, M., Devine, S. E., and Corbett, A. H. (2007) J. Biol. Chem. 282, 5101–5105
6. Pemberton, L. F., and Paschal, B. M. (2005) Traffic 6, 187–198
7. Weis, K. (2003) Cell 112, 441–451
8. Kalab, P., Weis, K., and Heald, R. (2002) Science 295, 2452–2456
9. Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984) Cell 39, 499–509
10. Dang, C. V., and Lee, W. M. (1988) Mol. Cell. Biol. 8, 4048–4054
11. Dingwall, C., and Laskey, R. A. (1991) Trends Biochem. Sci. 16, 478–481
12. Robbins, J., Dilworth, S. M., Laskey, R. A., and Dingwall, C. (1991) Cell 64, 615–623
13. Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuruvij, J. (1998) Cell 94, 193–204
14. Fontes, M. R., Teh, T., and Kobe, B. (2000) J. Mol. Biol. 297, 1183–1194
15. Fontes, M. R., Teh, T., Jans, D., Brinkworth, R. L., and Kobe, B. (2003) J. Biol. Chem. 278, 27981–27987
16. Conti, E., and Kuruvij, J. (2000) Structure 8, 329–338
17. Fontes, M. R., Teh, T., Toth, G., John, A., Pavo, I., Jans, D. A., and Kobe, B. (2003) Biochem. J. 375, 339–349
18. Hodel, M. R., Corbett, A. H., and Hodel, A. E. (2001) J. Biol. Chem. 276, 1317–1325
19. Makker, J. P., Dingwall, C., and Laskey, R. A. (1996) Curr. Biol. 6, 1025–1027
20. Xiao, C. Y., Hubner, S., and Jans, D. A. (1997) J. Biol. Chem. 272, 22191–22198
21. Horisawa, K., Tateyama, S., Ishizaka, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Doi, N., and Yanagawa, H. (2004) Nucleic Acids Res. 32, e169
22. Miyamoto-Sato, E., Ishizaka, M., Horisawa, K., Tateyama, S., Takashima, H., Fuse, S., Sue, K., Hirai, N., Masuoka, K., and Yanagawa, H. (2005) Genome Res. 15, 710–717
23. Kousig, S., and Ohashi, Y. (2002) Plant Physiol. 128, 833–843
24. Roberts, R. W., and Szostak, J. W. (1997) Proc. Natl. Acad. Sci. U.S.A. 94, 12297–1302
25. Nemoto, N., Miyamoto-Sato, E., Husumi, Y., and Yanagawa, H. (1997) FEBS Lett. 414, 405–408
26. Wilson, D. S., Keefe, A. D., and Szostak, J. W. (2001) Proc. Natl. Acad. Sci. U.S.A. 98, 3750–3755
27. Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T., and Adam, S. A. (2004) Trends Cell Biol. 14, 505–514
28. Shojo, K., Iwasaki, T., Matsuki, R., Miyao, M., and Yamamoto, N. (1998) Gene (Amst.) 212, 279–286
29. Mélen, K., Fagerlund, R., Franke, J., Köhler, M., Kinnunen, L., and Julkunen, I. (2003) J. Biol. Chem. 278, 28193–28200
30. Leung, S. W., Harreman, M. T., Hodel, M. R., Hodel, A. E., and Corbett, A. H. (2003) J. Biol. Chem. 278, 41947–41953
31. Bennett, R. P., Presnyak, V., Wedekind, J. E., and Smith, H. C. (2008) J. Mol. Biol. 38, 279–286
32. Shoya, Y., Kobayashi, T., Koda, T., Ikuta, K., Kakinuma, M., and Kishi, M. (1998) J. Virol. 72, 9755–9762
33. Aratani, S., Oishi, T., Fujita, H., Nakazawa, M., Fujii, R., Imamoto, N., Oneda, Y., Fukamizu, A., and Nakajima, T. (2006) Biochem. Biophys. Res. Commun. 340, 125–133
34. Ide, Y., Zhang, L., Chen, M., Inchauspe, G., Bahl, C., Sasaguri, Y., and Padmanabhan, R. (1996) Gene (Amst.) 182, 203–211
35. Yasuhara, N., Shibaizaki, N., Tanaka, S., Nagai, M., Kamiyama, Y., Os, E., Asally, M., Kamachi, Y., Kondoh, H., and Oneda, Y. (2007) Nat Cell Biol. 9, 72–79
36. Nadler, S. G., Triteschler, D., Haffar, O. K., Blake, J., Bruce, A. G., and Cleaveland, J. S. (1997) J. Biol. Chem. 272, 4310–4315
37. Miyamoto, Y., Imamoto, N., Sekimoto, T., Tachibana, T., Seki, T., Tada, S., Enomoto, T., and Oneda, Y. (1997) J. Biol. Chem. 272, 26375–26381
38. Köhler, M., Speck, C., Christiansen, M., Bischoff, F. R., Prehn, S., Haller, H., Görlich, D., and Hartmann, E. (1999) Mol. Cell. Biol. 19, 7782–7791
39. Velichkova, M., and Hasson, T. (2005) Mol. Cell. Biol. 25, 4501–4513