Dissection of a Nuclear Localization Signal*

Received for publication, September 18, 2000, and in revised form, October 16, 2000
Published, JBC Papers in Press, October 18, 2000, DOI 10.1074/jbc.M008522200

Mary R. Hodel, Anita H. Corbett, and Alec E. Hodel‡

From the Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30322

The regulated process of protein import into the nucleus of a eukaryotic cell is mediated by specific nuclear localization signals (NLSs) that are recognized by protein import receptors. This study seeks to decipher the energetic details of NLS recognition by the receptor importin-α through quantitative analysis of variant NLSs. The relative importance of each residue in two monopartite NLS sequences was determined using an alanine scanning approach. These measurements yield an energetic definition of a monopartite NLS sequence where a required lysine residue is followed by two other basic residues in the sequence K/K/R/X/K/R). In addition, the energetic contributions of the second basic cluster in a bipartite NLS (−3 kcal/mol) as well as the energy of inhibition of the importin-α importin-β binding domain (−3 kcal/mol) were also measured. These data allow the generation of an energetic scale of nuclear localization sequences based on a peptide’s affinity for the importin-α-importin-β complex. On this scale, a functional NLS has a binding constant of −10 nM, whereas a nonfunctional NLS has a 100-fold weaker affinity of 1 μM. Further correlation between the current in vitro data and in vivo function will provide the foundation for a comprehensive quantitative model of protein import.

The sequestering of genetic material in the nucleus by eukaryotic cells provides a powerful mechanism for the regulation of gene expression and other cellular processes through the selective translocation of proteins between the nucleus and the cytoplasm (1–3). Recently, the regulated transport of proteins across the nuclear envelope has been recognized as a crucial step in an increasing number of cellular processes (4–6). Understanding the mechanisms of regulated protein translocation through nuclear pores requires a detailed definition of the signals that mark a macromolecular complex for nuclear import or export.

The best characterized mechanism for translocation across the nuclear envelope is protein import which depends on the “classical” nuclear localization signal (NLS)1 (7). This NLS consists of a cluster of basic residues (monopartite) or two clusters of basic residues separated by 10–12 residues (bipartite) (8, 9). This signal is recognized by the heterodimeric import receptor complex comprising importin-α and importin-β (3). Importin-α is an adapter protein that consists of a small N-terminal importin-β-binding (IBB) domain and a larger C-terminal NLS-binding domain (10–14). Importin-β does not directly interact with the NLS cargo but acts to direct importin-α to the nuclear pore (15, 16). In the absence of importin-β, “NLS-like” sequences of the N-terminal IBB domain form an intramolecular bond with the NLS-binding site inhibiting the interaction between importin-α and the NLS cargo. Evidence for this auto-inhibition is found in the crystal structure of full-length importin-α as well as in vitro binding assays (16–19). Thus, the interaction between importin-α and the NLS cargo is regulated by importin-β. In an analogous manner, the interaction between importin-α and importin-β is regulated by the small GTPase Ran. Both structural and biochemical evidence indicate that in the GTP-bound state, Ran binds tightly to importin-β resulting in a conformational change that triggers the release of importin-α (20–23).

This cascade of regulated interactions suggests a molecular model for unidirectional import of proteins into the nucleus (2, 3, 24). The small GTPase Ran serves to distinguish the nucleus from the cytoplasm (23). In the nucleus, Ran is found primarily in a GTP-bound form, whereas, in the cytoplasm, the GDP-bound form is dominant (3). Thus, in the cytoplasm, importin-β is free to bind to importin-α removing the inhibition of NLS binding by the IBB domain. The importin-β-bound importin-α is then free to bind to the NLS cargo, and the ternary complex is translocated through the nuclear pore via interactions between importin-β and the nucleoporins (15, 16). In the nucleus, the ternary complex encounters Ran in the GTP-bound state. Ran-GTP binds to importin-β releasing the importin-α-NLS complex (19). The IBB domain of importin-α then competes with the NLS cargo for the NLS-binding site facilitating the release of the NLS cargo into the nucleus. This model for protein import requires precise tuning of the thermodynamic interactions between the various species for the reaction to proceed efficiently in a single direction. For example, the interaction between the NLS and importin-α must be tuned such that the affinity of the importin-α-importin-β complex for the NLS is tight enough to allow cytoplasmic capture and nuclear translocation of the cargo. However, the interaction between the NLS and importin-α alone (in competition with the IBB domain) must be weak enough to allow efficient release of the NLS cargo into the nucleus. These restrictions yield a simple thermodynamic definition of a classical nuclear localization signal.

Presently, the definition of a nuclear localization signal sequence is somewhat vague owing to the diversity of sequences that can apparently act as a functional NLS (7). The NLS of the SV40 large T antigen, with a sequence of PKKKKRKV, provides the prototypic monopartite NLS defined by a cluster of basic residues. Functional assays indicate that a lysine is essential in the third position of this sequence, but the importance of the other residues is somewhat ambiguous (8, 25). The NLS from the c-myc proto-oncogene, with a sequence of PAARRVKLĐ,
Dissection of a Nuclear Localization Signal

![Diagram of importin α and NLS sequences]

**FIG. 1.** The atomic structures of importin α bound to the Myc NLS (A) and the SV40 NLS (B). The surface of importin α is rendered as a molecular surface, whereas the bound NLS peptides are shown as rods. The peptides bind in an extended conformation with each side chain situated in a unique pocket of the protein. These pockets are labeled as per the number scheme outlined in Table 1 (note that pocket 3 is obscured by a loop of importin α in A). In the NLS peptides, carbon atoms are shown in white and oxygen and nitrogen atoms are drawn in red and blue, respectively. The surface of importin α is rendered according to the electrostatic potential, with red denoting a negative charge and blue denoting positive. The NLS binds in a predominantly acidic groove of importin α. The rendering was made using the program GRASP (36).

illustrates the diversity of peptide sequences that can act as functional localization signals (26). In analogy with the third lysine of the SV40 NLS, the fourth lysine was found to be critical in the Myc NLS. In addition, the initial proline and the final Leu-Asp dipeptide were also found to be important in the function of this NLS sequence (27). The crystal structures of importin α in complex with both the SV40 and the Myc NLS peptides have been reported recently (12, 28). These crystal structures showed that the similar but distinct NLS sequences bound to the same site on importin α in an extended conformation. As shown in Fig. 1, the SV40 sequence PKKKKRV bound in a nearly identical conformation to the Myc sequence of PAAKRVKL with equivalent residues from each binding in identical pockets on the protein (see Table 1).

In correlation with the functional data, the interactions in pocket 1 are extensive including three hydrogen bonds, a salt bridge, and hydrophobic interactions with the aliphatic segment of the lysine side chain. This is consistent with the hypothesis that the amino acid specificities of the NLS binding site of importin α define the sequence requirements for a functional nuclear localization signal.

A common variant of the classical NLS is a bipartite sequence with a small cluster of basic residues positioned 10–12 residues N-terminal to a monopartite-like sequence. The prototypal bipartite NLS is found in nucleoplasm in with the sequence KRPAATKKAGQAKKKKL (9). The additional binding energy contributed by the upstream cluster of basic residues relaxes the requirements for the downstream monopartite-like sequences. In fact, a nonfunctional SV40 variant where the critical third lysine residue is replaced with a threonine (PKTKRV) can be converted into a functional NLS through the addition of a second properly positioned basic cluster (KRTADSQHSTTPKTKRV) (27).

A complete understanding of nuclear import signals requires a quantitative model for the import reaction that correlates NLS amino acid sequence, *in vitro* interaction energies, and *in vivo* functionality. The first steps toward such a model were taken with the report of a quantitative assay for the affinity between importin α and an NLS sequence using an enzyme-linked immunosorbent assay-based method (29–32). We have recently expanded on these efforts by reporting a fluorescence-based assay for NLS-importin α interactions that is performed in solution at equilibrium (17). With this fluorescence assay, we have begun to reconstitute the molecular reactions of protein import *in vitro* to provide a detailed thermodynamic description of the translocation reaction. A detailed description of the energetic requirements for an NLS sequence will facilitate the recognition of these sequences in protein primary structure as well as suggest possible modes for the regulation of protein import.

Here we attempt to decipher the energetic details of NLS recognition by importin α through quantitative analysis of variant NLS affinities. The relative importance of each residue in two monopartite NLS sequences was determined using an alanine scanning approach. These measurements yield an energetic definition of a monopartite NLS sequence. In addition, the energetic contributions of the second basic cluster in a bipartite NLS as well as the energy of inhibition of the importin α IBB domain were also measured. These data allow the generation of an energetic scale of nuclear localization sequences and provide the foundation for a comprehensive quantitative model of protein import.

### MATERIALS AND METHODS

**Generation of NLS-GFP Variants and Protein Purification**—Various NLS sequences were cloned as in-frame N-terminal fusions to the green fluorescent protein (GFP) through polymerase chain reaction in a pet-28a (Novagen) expression vector. The amino acid sequences of the NLS variants are enumerated in Table II. The identity of each variant was confirmed by DNA sequencing. Each of these variants was expressed and purified as described elsewhere (17). Both full-length importin α and a fragment consisting of residues 89–530 (ΔIBB importin α) were expressed and purified as described (17).

**NLS Binding Assay**—The dissociation constant for the binding between NLS-GFP fusion proteins and importin α was measured through a fluorescence depolarization assay (17). The anisotropy of the GFP fluorescence was monitored using an ISS PCl fluorimeter with the sample maintained at 25 °C. The sample was excited at a wavelength of 492 nm, and the emitted fluorescence was measured after filtering through a 510-nm high pass filter. The changes in the anisotropy of NLS-GFP when titrated with various concentrations of importin α were then used to calculate the fraction of NLS-GFP bound yielding a binding isotherm for the reaction. The binding isotherm was then fit through nonlinear regression to a simple binding equation.

### TABLE I

| Pocket* | 12345 |
|---------|-------|
| SV40    | PKKKKRV |
| Myc     | PAAKRVKL |

* The position of each residue when bound to importin α as shown in Fig. 1.
Dissection of a Nuclear Localization Signal

Buried Surface Area—The surface area buried upon the binding of an NLS to importin α was calculated from the crystal structures for the SV40-NLS-importin α complex (Protein Data Bank code 1BK6) and the Myc-NLS-importin α complex (Protein Data Bank code 1EE4). Areas were calculated with the CNS program using the standard protein parameters therein (33). To determine the surface area, the NLS was first separated from importin α and the surface area of each molecule was obtained. Next, the surface area of the complex was calculated and subtracted from the sum of the individual, unbound surface areas. Both the protein and peptide were maintained in the same conformation in both the bound and unbound state in this calculation. The relative changes in buried surface area for a small peptide NLS, the inhibition constant of this peptide in the binding of SV40-GFP to IBB importin α was measured. The binding curve for SV40-GFP was measured in the presence of four different concentrations of the SV40-A5 peptide ranging between 5 and 100 μM. The resulting binding curves were then simultaneously fit to an equation for the fraction NLS bound, \( Y \), as a function of \( K_y \), for SV40-NLS, \( K_y \), for the peptide, the total NLS-GFP concentration, the total importin α concentration, and the total concentration of the peptide. There are three solutions to this equation, two of which are physically useful, that correspond to the situations where \( K_y > K_b \) and where \( K_y < K_b \). Using the latter case, the binding affinity between the SV40-A5 peptide and IBB importin α was calculated (data not shown).

Theoretical Energy Estimation—Using the published three-dimensional structures of the SV40 NLS and the Myc NLS in complex with IBB importin α (12, 28), a theoretical estimation of the relative free energy contribution of each residue of each NLS in the binding reaction with importin α was calculated. To perform a rigorous, atomic-level free energy simulation with this system would require an enormous effort, and such an effort would not yield much higher precision or accuracy than simpler, approximate methods. Thus, the relative free energy contributions reported here were generated using a number of reasonable methods based on approximations. In this calculation, we estimate the change in the binding energy, or \( \Delta G \), when a residue of the NLS is substituted with alanine. Thus, only terms that will differ between the side chain of the variant NLS and importin α undoubtedly change upon forming a complex, we assume that these changes will not have a significant effect when comparing the buried surface area from one NLS variant to the next. For each NLS alanine variant, the calculation was repeated with the appropriate NLS side chain atoms omitted.

Electrostatic Energy—The change in electrostatic interaction between the NLS and importin α with each alanine substitution was estimated first by calculating the electrostatic potential for the NLS-importin α complex with the appropriate side chain atoms removed. The electrostatic potential was calculated using the Poisson-Boltzmann method in the program GRASP with partial atomic charges taken from the CNS protein parameter file (35, 36). The electrostatic potential was then shifted to the position of each atom from the backbone chain. Then this potential was multiplied by the partial charges of each of the side chain’s atoms yielding the approximate energy of interaction between the side chain and the rest of the complex.

Helicity—The fractional helicity of each NLS variant sequence was calculated using the program AGADIR (37). The entire N-terminal sequence of each variant as shown in Table II was used as the input into this program and the helical content of each peptide sequence was obtained. We assume that the energy necessary to unravel the peptide into an extended conformation for binding to importin α is proportional to the helical content of the peptide.

Deconvolution of Energetic Terms—The approximations involved with calculating these three energetic terms preclude the possibility of obtaining a true theoretical free energy. These terms gives a relative scale for the contribution of each term to the \( \Delta G \) of binding of each variant NLS. The values obtained are proportional to \( \Delta G \) and are comparable within each term, but the three energy terms calculated above cannot be directly compared with each other. These calculations can be used to determine the relative importance of each term in the binding of each residue of the NLS to importin α. To determine the relative contributions of each term to NLS binding, a simple linear deconvolution was performed. For each of the two NLS sequences, the SV40 and the Myc NLSs, \( \Delta G \) values were calculated for each alanine variant in comparison to the wild type sequence. Assuming that the three energy terms calculated above, buried surface, electrostatic interaction, and helicity, are proportional to \( \Delta G \) and are internally consistent, then \( \Delta G \) can be expressed as a linear function of these three terms. Suppose the vector \( \Delta G = (\Delta G_1, \ldots, \Delta G_N) \), where \( \Delta G_i \) is the experimentally measured \( \Delta G \) value for the NLS alanine mutant \( i \). Then, given vectors for the calculated terms \( \Delta \text{surf} \) (buried surface), \( \Delta \text{elec} \) (electrostatics), and \( \Delta \text{hel} \) (helicity), \( \Delta G \) can be fit to Equation 2.

\[
\Delta G = C_{\text{surf}} \Delta \text{surf} + C_{\text{elec}} \Delta \text{elec} + C_{\text{hel}} \Delta \text{hel} \quad (\text{Eq. 2})
\]

where \( C_{\text{surf}}, C_{\text{elec}}, \) and \( C_{\text{hel}} \) are scalar constants. If a reasonable fit is achieved, then the resulting terms \( C_{\text{surf}} \Delta \text{surf} \) (buried surface), \( C_{\text{elec}} \Delta \text{elec} \) (electrostatics), and \( C_{\text{hel}} \Delta \text{hel} \) (helicity) give the relative contributions of each energy term to the binding of each residue of the NLS to importin α.

RESULTS AND DISCUSSION

Energetic Contributions of Each Residue in a Monopartite NLS—Nuclear localization signals are an apparently diverse set of sequences with a generally polybasic character (7, 8). To obtain a more detailed description of the conformational features of a monopartite NLS, the energetic contributions of each residue in two classical NLS sequences (the SV40 NLS and the Myc NLS) to the binding of the NLS to yeast importin α were measured through an alanine scan. To simplify the interpretation and collection of the data, the affinity of each NLS variant was measured for binding to an importin α fragment (ΔIBB

\[
Y(K_y, L_v, R_v) = K_y + L_v + R_v - \left(K_y + L_v + R_v\right)^2 - 4 L_v \frac{R_v}{2L_v R_v} \quad (\text{Eq. 1})
\]
importin α) that lacks the auto-inhibitory N-terminal importin β binding domain. This fragment is identical to the one that was crystallized in complex with both the SV40 and the Myc NLS peptides (12), facilitating direct comparison of the thermodynamic data to the atomic structure.

Each nonalanine residue of both the Myc and SV40 NLS sequences was mutated one at a time to alanine, and the affinity of each of these mutant NLS sequences for ΔIBB importin α was measured. Each NLS sequence was fused to the N terminus of GFP (see Table II). The binding of this NLS-GFP fusion to ΔIBB importin α was measured by monitoring the fluorescence depolarization of GFP while titrating the fused NLS with ΔIBB importin α. The resulting binding curves were then fit to obtain a dissociation constant for the NLS-GFP/ΔIBB importin α interaction (Fig. 2). The change in the affinity of each NLS variant is plotted as a ΔΔG value in Fig. 3, illustrating the energy profile for each NLS sequence.

As expected from previous functional and structural studies, a single lysine dominates the energetic profile of the monopartite NLS (8, 12, 25). This lysine corresponds to Lys-3 in the SV40 NLS and Lys-4 in the Myc NLS, where both residues bind to pocket 1 of the protein (numbered from Table I, see Fig. 1). Mutation of this residue in the SV40 NLS destroys its function in vivo (8). This mutant NLS bound with affinity nearly identical to that of the alanine mutation. A methionine substitution (SV40M3) was generated to test whether the hydrophobic surface of a long side chain could contribute a significant amount of binding energy. Surprisingly, this variant bound more weakly than the alanine and threonine variants. Finally, an arginine was substituted at this position (SV40R3) to test whether the site was specific for lysine or could accommodate either basic residue. Although this variant bound more tightly than the alanine variant, its affinity for ΔIBB importin α was significantly weaker than that of the wild type SV40 sequence. Thus, this pocket appears to be fairly specific for a lysine residue.

Although the Myc and SV40 NLS sequences differ in the residues of pockets 2, 3, 4, and 5, the energy profile for the five corresponding residues of both Myc and SV40 highlighted in Table I are quite similar (Fig. 3). Pocket 1 dominates the energy profile, pockets 2 and 4 have intermediate contributions, and pockets 3 and 5 have fairly weak individual contribu-

**Table II**

NLS-GFP sequences and affinities for ΔIBB importin α

| NLS variant   | N-terminal sequence of GFP fusiona | Kα, ΔIBB importin α |
|---------------|-----------------------------------|-------------------|
| GFP           | [S1] - QMGR GSEFE LARQ - [S2-GFP] | TWTMb             |
| Myc           | [S1] - QMGR GSEFE PAAKRVYKDE - [S2-GFP] | 6 ± 3             |
| MycA1         | [S1] - QMGR GSEFE AAGKRVYKDE - [S2-GFP] | 120 ± 14          |
| MycA4         | [S1] - QMGR GSEFE PAAKRVYKDE - [S2-GFP] | 15,000 ± 7000     |
| MycA5         | [S1] - QMGR GSEFE PAAKAVYKDE - [S2-GFP] | 1400 ± 850        |
| MycA6         | [S1] - QMGR GSEFE PAAKAVKDE - [S2-GFP] | 25 ± 11           |
| MycA7         | [S1] - QMGR GSEFE PAAKVADE - [S2-GFP] | 650 ± 70          |
| MycA8         | [S1] - QMGR GSEFE PAAKVKAD - [S2-GFP] | 85 ± 7            |
| MycA9         | [S1] - QMGR GSEFE PAAKVKLAD - [S2-GFP] | 13.5 ± 8          |
| SV40          | [S1] - QMGR GSEFE SPEKKRVE - [S2-GFP] | 9 ± 4             |
| SV40A1        | [S1] - QMGR GSEFE SKKKRVE - [S2-GFP] | 36 ± 2            |
| SV40A2        | [S1] - QMGR GSEFE PAKKRVEK - [S2-GFP] | 16.5 ± 1          |
| SV40A3        | [S1] - QMGR GSEFE SPEKKRVEK - [S2-GFP] | 3000 ± 1400       |
| SV40A4        | [S1] - QMGR GSEFE SPEKKRVEK - [S2-GFP] | 335 ± 7           |
| SV40A5        | [S1] - QMGR GSEFE SPEKKRVEK - [S2-GFP] | 38 ± 2            |
| SV40A6        | [S1] - QMGR GSEFE SPKKRVEK - [S2-GFP] | 310 ± 100         |
| SV40A7        | [S1] - QMGR GSEFE SPKKRVEK - [S2-GFP] | 53 ± 4            |
| SV40T3        | [S1] - QMGR GSEFE SPEKKVKE - [S2-GFP] | 3000 ± 1414       |
| SV40R3        | [S1] - QMGR GSEFE SPEKKVKE - [S2-GFP] | 850 ± 200         |
| SV40M3        | [S1] - QMGR GSEFE SPEKKQKE - [S2-GFP] | 6500 ± 700        |
| BP-GFP        | [S1] - KRTAD GSEFE LARQ - [S2-GFP] | 2000 ± 1000       |
| BPSV40        | [S1] - KRTAD GSEFE SPEKKRVE - [S2-GFP] | TTTTM             |
| BPSV40A4      | [S1] - KRTAD GSEFE SPEKKRVE - [S2-GFP] | 1 ± 1             |
| BPSV40A5      | [S1] - KRTAD GSEFE SPEKKRVE - [S2-GFP] | TTTTM             |
| BPSV40A6      | [S1] - KRTAD GSEFE SPEKKRVE - [S2-GFP] | 0.9 ± 1           |
| BPSV40T3      | [S1] - KRTAD GSEFE SPEKKRVE - [S2-GFP] | 13.5 ± 6          |
| SV40A5        | SPEKKRVEAC | 10,000 ± 1000 |

a [S1] represents the sequence: MGSSHIIHHIHSSGLVPRGSHMASMTGG--; [S2-GFP] represents the sequence: -ASGLVPRGSH- (GFP).

b TWTM, too weak to measure; TTTM, too tight to measure.

**FIG. 2.** Typical curves for the binding of an NLS-GFP fusion protein to ΔIBB importin α measured through fluorescence anisotropy. The fluorescence anisotropy of the NLS-GFP fusion protein is plotted as a function of the concentration of ΔIBB importin α on a log scale. The data for two NLS variants were plotted. Triangles, SV40; diamonds, SV40A7 variant. For each variant, the data were fit to an ideal binding curve to yield a measurement of $K_a$. The fit curves are displayed with the data as solid lines.
Dissection of a Nuclear Localization Signal

![Graph](image)

**Fig. 3.** The energy profile for the Myc (A) and the SV40 (B) NLS sequences. The amino acid sequence of each NLS is shown along the X axis. Above each residue letter is plotted the change in the free energy of binding between the NLS and importin α when that residue is replaced with an alanine. Each value is calculated as \( \Delta \Delta G = kT \ln(K_a/\text{alanine variant})/K_a/\text{wild type NLS} \) where \( kT = 0.59 \text{ kcal/mol} \). The error bars shown are calculated by calculating \( \Delta \Delta G \) at the 95% confidence limits of \( K_a \), and reporting the error in \( \Delta \Delta G \) as the larger of the values from \( \Delta \Delta G(K_a > 95\% \text{ confidence interval}) \). 

B contributions to the NLS binding energy. It is interesting to note that pocket 3 has the weakest contribution of the five positions, while it also has the least conserved amino acids between the two NLSs: an Arg in SV40 and a Val in Myc. Also notable is that this pocket is the site of a fairly large conformational change in importin α when comparing the SV40 complex with the Myc complex. A reasonable conclusion is that this flexible site can accommodate a large variety of amino acids, Arg, Val, or Ala, with little change in the energy of binding.

**Establishment of Boundaries—**A consideration that was not mentioned in the above discussion is that, with a repetitive sequence like that of the SV40 NLS, there is a possibility for variants to bind in a different register than that observed for the wild type sequence. For example, in the SV40A3 mutant, given the specificity of pocket 1 for lysine, the variant could bind to importin α with either Lys-2 or Lys-4 in pocket 1 switching the register of binding by one residue (see Table III). This appears not to be the case. One would not expect such large variations in the binding energy between adjacent alanine variants if the interaction was plastic enough to accommodate a register shift. A detailed comparison of adjacent alanine variants while considering the possibility of register shifts can lead to further information about the specificity of positions within the binding site and those positions immediately flanking. As a specific example, a comparison of the SV40A2, -A3, and -A4 mutations defines the specificity of the terminal edges of the NLS binding site. Thus, the standard mode for binding observed in the crystal structure can be compared with other possible modes where the register has shifted by 1 residue (Table III).

We measured a 3 μM dissociation constant for the SV40A3 variant. Thus regardless of what conformation or register this variant adopts when binding to importin α, the tightest binding mode possible between the NLS and the receptor yields a binding constant of 3 μM. As shown in Table III, if the SV40 variant was to shift the position of its binding by one residue in the N-terminal direction (register −1), the specific positions of lysines and alanines with regard to the five binding pockets of importin α would be somewhat similar to the binding of the SV40A2 variant in the standard register. However, SV40A2 has the much tighter \( K_a \) of 17 nM, 2 orders of magnitude tighter than the binding of the SV40A3 variant. The large difference in the binding affinity of SV40A3 in the −1 register compared with the SV40A2 variant suggests that at least one of the substitutions shown in the SV40A3, register −1 mode of binding costs at least 3 kcal/mol in binding energy. These substitutions include a Lys → Arg at position 2, a Lys → Val at position 4, and a Val → Glu at position 5. If these substitutions are less deleterious than the alanine substitution at position 1 (in the standard register), then the binding mode of the SV40A3 variant observed would be in the −1 register rather than the mode observed in the crystal structure. From the sequence of the Myc NLS, one can conclude that the Lys → Arg at position 2 can be accommodated. Thus the combination of Val at position 4 and Glu at position 5 must cost at least 3 kcal/mol. Similarly, one can compare SV40A3 in the +1 register to SV40A4 and conclude that the combination of the Lys → Pro at the −1 position, the Arg → Lys at position 3, the Lys → Arg at position 4 and the Val → Lys at position 5 must cost at least 1.2 kcal/mol in binding energy. Thus, the register of binding of the polybasic SV40 NLS is most probably set by the specificity of the −1 position, position 4, and/or position 5. Thus, the large variation in binding energy observed for sequential alanine substitutions in the NLS suggests that all the alanine variants bind in an identical conformation to that observed for the wild type sequences in the crystal structures.

An important conclusion drawn from these data is that the requirements for a monopartite NLS sequence are more specific than a simple cluster of basic residues. Our results are consistent with previous functional and structural studies, suggesting a basic core of the monopartite NLS with a sequence K(K/R)(X/K/K/R)8 (6, 12, 25). The analysis of possible register shifts above suggests further requirements in the residues preceding the N-terminal anchoring lysine and/or at the C terminus of the basic core; however, the elucidation of the specific character of these terminal positions will require further study.

**Modeling Energetic Terms—**With the availability of the atomic structures of the NLS-importin α complexes, it seemed a reasonable goal to attempt to correlate the experimental data with the crystal structure in a quantitative manner. With such an analysis, the energetic contributions of each residue might be further dissected into terms such as hydrophobic entropy and electrostatic interactions. As a rigorous free energy simulation with a system as large as the NLS-importin α complex would be technically difficult, a more approximate approach was taken. The goal of this analysis was to calculate individual

| Variant and register | Pocket position of side chain when bound to ΔIBB importin α | \( K \) |
|----------------------|-------------------------------------------------------------|------|
| SV40                | SPYKKKKEVVEA nM | 9 ± 4 |
| SV40A3, register 0  | SPYKKKKEVVEA | 3000 ± 1400 |
| SV40A3, register −1 | SPYKKKKEVVEA | 3000 ± 1400 |
| SV40A2              | SPYKKKKEVVEA | 16.5 ± 1 |
| SV40A3, register +1 | SPYKKKKEVVEA → | 3000 ± 1400 |
| SV40A4              | SPYKKKKEVVEA | 335 ± 7 |
terms for the $\Delta \Delta G$ values obtained in the alanine scanning data and then attempt to correlate these calculated energies with the experimentally measured quantities. Three energetic terms were considered. First, the changes in the surface area that was buried in the complex between the NLS and importin $\alpha$ was calculated as a relative measure of van der Waals interactions along with entropic interactions with the solvent (including the hydrophobic effect). Second, the interaction of each side chain in the NLS with the surrounding electrostatic potential was calculated as a relative measure of the electrostatic and hydrogen bonding interactions between the peptide side chains and the surrounding protein. Finally, the helical contents of the various NLS sequences were calculated as a relative measure of the differences in the conformational energy of the unbound NLS sequence that must be overcome to bind to importin $\alpha$.

These three terms were then considered to be internally consistent and proportional to the free energy such that a linear combination of the three terms should be equivalent to the total $\Delta \Delta G$ values obtained by experimentation. As described under “Experimental Procedures,” three scalar coefficients for the three terms were calculated by fitting the linear combination of these terms to the experimental data. The results of these calculations is illustrated in Fig. 4. Although the experimental data from the Myc NLS correlate well with the theoretical data, the SV40 calculations did not produce as good of a fit. When one set of coefficients was fit to all the data (both SV40 and Myc alanine scans), the calculated $\Delta \Delta G$ correlated fairly well with the experimental data ($r = 0.83$ overall, $r = 0.97$ Myc data alone); however, $\Delta \Delta G$ for most of the variants was underestimated (Fig. 4). This is primarily due to the overestimation of the theoretical terms describing the SV40A5 and SV40A2 variants. As shown in Fig. 4B, when the three terms are separated, it becomes apparent that the buried surface area due to the residues in the SV40A5 (Arg) and SV40A2 (Lys) variants is much more substantial in the crystal structure than expected from the experimental data, where both alanine substitutions had a very modest effect.

When the coefficients for the three energy terms were fit for just the Myc data alone, the correlation coefficient for these data increased to $r = 0.996$. When these coefficients are applied to all the calculations, the resulting energies are shown in Fig. 4C. The Myc data is fit nearly perfectly, but again, the calculated energy for SV40A2 and SV40A5 were significantly overestimated. Although this overestimation in these residues cannot be presently explained, the fit obtained for the rest of the data ($r = 0.94$ excluding SV40A2 and SV40A5) suggests that these calculations provide useful information.

When the data are fit using just the Myc alanine scan, the buried surface area yields the dominant energy term for the free energy profiles of the NLSs (Fig. 4D). For both the Myc and the SV40 NLS, the electrostatic term is comparable to the surface area term for the residues bound in pocket 1 (SV40A3 and MycA4). When each of these terms are individually fit to the experimental data, the surface area term yields a correlation coefficient of $r = 0.7$, whereas the electrostatic term yields a correlation of $r = 0.73$. When the fit is performed using a linear combination of the surface area and the electrostatic terms (omitting the helicity term), the correlation yields $r = 0.82$ compared with $r = 0.83$ when all three terms are used. The helicity term appears to be insignificant in comparison to the other two terms. When considering the Myc NLS data alone, the buried surface area is even more dominant. The fit of the surface area term alone to the experimental data yields an $r = 0.95$, the electrostatic term alone yields $r = 0.85$, and the linear combination of surface area with electrostatics yields an $r = 0.985$ (compared with $r = 0.996$ with all three terms). Thus, in a majority of our variants, the experimentally determined binding affinities correlate well with the amount of surface that is observed to be buried in the crystal structure of the NLS-importin $\alpha$ complex.

**The Additional Energy of a Bipartite Sequence**—A majority of the putative NLS sequences recognized to date appear to be bipartite in structure (7). These sequences have a loose consensus of two basic clusters separated by 10–12 residues. Crystallographic analysis suggests that importin $\alpha$ binds these two basic clusters in two unique binding sites (28). The C-terminal cluster of the NLS binds to the monopartite-NLS binding site at the N terminus of the importin $\alpha$ Armadillo domain. The N-terminal cluster of the bipartite NLS binds to a smaller site near the C terminus of importin $\alpha$. It has been shown that adding a cluster of two basic amino acids 10–12 residues upstream of a nonfunctional or very weak NLS (for example the SV40T3 variant) converts the sequence into a functional signal (27). We have analyzed this effect quantitatively using our
SV40T3 variant. As shown in Table II, when two basic residues are added at the appropriate distance N-terminal to the SV40T3 sequence, the binding affinity of this sequence for ΔIBB importin α increases from 3 μM for the monopartite sequence to 13.5 nM for the bipartite variant (BPSV40T3). This further illustrates that there is a correlation between the ability of an NLS to function in vivo and the ability for the sequence to bind to importin α. This example also yields a measurement of the additional binding energy conferred by the addition of a second basic cluster to a monopartite NLS. The addition of two basic residues upstream of the monopartite sequences of the SV40T3, SV40A4, and SV40A6 variants contributed an average of ~3.1 kcal/mol to the binding affinity for ΔIBB importin α. The addition of this second cluster to the SV40 and SV40A5 sequences produced NLS-GFP fusions whose affinity for ΔIBB importin α was too tight to measure accurately using our methods (K_d < 0.5 nM). Thus, the inclusion of the upstream basic cluster dramatically increases the variety of sequences that can serve as a functional NLS.

Interestingly, the addition of the two basic residues to the GFP control protein yielded a fusion (BP-GFP) with measurable affinity for ΔIBB importin α (see Table II). This fusion binds to ΔIBB importin α with a K_d of 2 μM that is nearly identical to the binding constant for the SV40A3 NLS variant. There are two probable binding modes for the complex of the BP-GFP variant with ΔIBB importin α. The first mode would have the newly added basic cluster (with the sequence KR) binding to the monopartite binding site of ΔIBB importin α with a lysine in the pocket 1 position. A second possible mode would make use of two arginines from the original GFP vector positioned 10 residues down from the newly added KR cluster. This second arrangement would suggest that the original GFP protein from our vector, as a model of a random peptide, binds to ΔIBB importin α —3.1 kcal/mol more weakly than the BP-GFP variant. One can then conclude that a random peptide without a sequence nonspecific interactions between the residues N-terminal to the monopartite sequence and importin α as observed in the atomic structure of a bipartite NLS bound to ΔIBB importin α (28).

**Auto-inhibition of NLS Binding by the IBB Domain**—The crystal structure of full-length murine importin α revealed that the flexible, largely unstructured N-terminal importin β-binding domain contains NLS-like sequences that, in the absence of other proteins, will bind in the monopartite NLS binding site (18). We and others have previously shown that the inclusion of this domain inhibits the binding of monopartite NLS sequences to full-length importin α and that this inhibition is relieved in the presence of importin β (16, 17, 19). A model for these observations is that the IBB domain of importin α competes with monopartite NLS sequences for the NLS binding site. However, importin β binds tightly to this N-terminal IBB domain and sequesters it from the NLS binding site. The removal of this intramolecular competitive ligand increases the effective affinity of importin α for the NLS in trans.

Although we observed no binding between an SV40-GFP fusion and full-length importin α, the competitive inhibition of the IBB domain could, in principle, be overcome by increasing the affinity of the NLS sequence for importin α. To test this conjecture, the affinity of full-length importin α for the high-affinity BPSV40-GFP fusions was determined. The binding of BPSV40, BPSV40A4, -A5, and -A6 to full-length importin α was measurable through the fluorescence depolarization assay (Table IV).

To compare the affinities of NLS-GFP fusion proteins for both ΔIBB importin α and full-length importin α, the functional stoichiometry of the protein preparations needed to be confirmed. To this end, an end point titration was performed for both ΔIBB importin α and full-length importin α using BPSV40-GFP at high concentration as a standard. From this assay, 1 mol of full-length importin α was functionally equivalent to 0.96 mol of ΔIBB importin α. This assay confirms that both ΔIBB importin α and full-length importin α are equally functional and folded properly. Thus, the affinities of these two proteins may be directly compared.

The binding affinity of ΔIBB importin α for four SV40 NLS variants (SV40, SV40A4, -A5, and -A6) is compared with the affinity of full-length importin α for the bipartite versions of the same SV40 variants in Fig. 5. The relative change in binding energy with each alanine mutation is nearly identical for the SV40 and their bipartite variants. This suggests that the energy gained by the addition of the second basic cluster is nearly equivalent to the energy cost of the competition with the IBB domain. The profile in Fig. 5 additionally suggests that these energy modifications are independent of the specific sequence of the monopartite NLS. This conclusion is not entirely unexpected. The IBB domain should have the same intramolecular binding energy regardless of the NLS against which it competes. In addition, the distance between the second basic cluster and the monopartite sequence is large enough to expect these segments to interact with importin α independently from each other. With these assumptions, the intramolecular binding energy of the IBB domain in competition with an NLS in trans can be estimated from the differences in the affinities in Fig. 5. The binding energy of a bipartite variant interacting with full-length importin α is ~0.4 ± 0.3 kcal/mol stronger than the energy of the same monopartite variant interacting with ΔIBB importin α. Given the estimate for the interaction energy added with the second basic cluster of 3.2 kcal/mol, the intramolecular competition of the IBB domain reduces the binding affinity for an NLS by ~2.8 kcal/mol.

| NLS variant | K_d full-length importin α (nM) |
|-------------|---------------------------------|
| SV40        | TWTM*                           |
| BPSV40      | 5 ± 3                           |
| BPSV40A4    | 110 ± 25                        |
| BPSV40A5    | 35 ± 4                          |
| BPSV40A6    | 120 ± 32                        |

* TWTM, too weak to measure.
With this simplistic view of the auto-inhibitory behavior of importin α, the function of this intramolecular inhibition is not immediately obvious. A common hypothesis is that this intramolecular competition for the NLS binding site is responsible for the delivery of the cargo in the nucleus. Upon reaching the nuclear basket of the pore, the trimeric complex of importin α, importin β, and the NLS cargo encounter the small GTPase Ran in its GTP-bound state. Ran-GTP binds to importin β, which, through a conformation change, releases importin α from importin β (3). The release of the IBB domain from importin β effectively reduces the affinity of importin α for the NLS cargo through competitive inhibition and the cargo is released into the nucleoplasm.

This intuitive model breaks down when one considers the binding energies measured in this study. What happens when a cargo contains an NLS with an extremely high affinity for importin α such as the BPSV40 NLS? The affinity of the BPSV40 NLS for the inhibited full-length importin α is as high as the affinity of a functional SV40 NLS for the uninhibited ΔIBB importin α. If ΔIBB importin α is considered a model for the importin α-importin β complex, then the BPSV40 NLS should bind to importin α in the nucleus with the same strength as the functional SV40 NLS binds to the importin α-importin β complex in the cytoplasm and nuclear pore. This suggests that the release of these high affinity cargoes, and thus the transport of these cargoes, would be fairly inefficient if the mechanism for nuclear release was dependent solely on the competitive inhibition of the importin α IBB domain. Such high affinity NLS sequences may not occur in vivo; indeed, the sequence of the BPSV40 NLS is completely artificial, but this model for release based on the auto-inhibition of importin α suggests that abnormally high affinity NLS sequences would have to be avoided in a natural setting. A plausible alternative would be that the auto-inhibition observed in vitro is just one element of a more active release mechanism that can accommodate high affinity NLS sequences. So far, evidence for such an active mechanism has not been observed.

**In Vitro Energy Scale for Nuclear Localization**—One goal of this quantitative analysis is to provide a thermodynamic foundation for a numerical model of the process of nuclear transport. One would expect that the *in vivo* process of nuclear import would correlate in some manner to the energetics of the individual protein-protein interactions that drive the process. It has been suggested that the initial rate of protein import is linearly correlated with the equilibrium constant for the interaction between the NLS cargo and importin α (30, 34, 38). This relationship would hold true in a simple model where the rate of protein import would depend on the equilibrium concentration of the importin α-importin β-cargo NLS ternary complex. One correlation that is clear from numerous studies is that there is some sort of functional threshold of affinity that an NLS must possess for importin α for the cargo to be imported into the nucleus. When the SV40 NLS is mutated to the SV40T3 sequence, its affinity for importin α decreases by ~3 kcal/mol and it also loses its ability to function as a nuclear localization signal *in vivo* (8). Given the arguments proposed above regarding the function of importin α auto-inhibition in cargo release, it will be interesting to see whether there are both lower and upper thresholds for the binding energy of a functional NLS.

The quantitative data presented here yield a numerical skeleton on which to build a comprehensive model for this complicated process. With the assumptions made above, a given NLS can be situated on a linear scale that describes its affinity for the importin α-importin β complex (using ΔIBB importin α as a model) as well as its affinity for importin α alone. For an NLS to function in nuclear import, one might hypothesize that it must have an affinity for the importin α-importin β complex that is tight enough to stimulate the uptake of the NLS cargo into the nuclear pore, but it must also have an affinity for the importin α that is weak enough to allow efficient release of the cargo into the nucleus. A scale of the values measured and calculated here is illustrated in Fig. 6. We are currently undertaking experiments to correlate the thermodynamic values for these interactions with the kinetics of nuclear import *in vivo*.

**Acknowledgments**—We thank Dr. Elena Conti for making the coordinates of the importin α structures available before publication. We also thank Patrizia Fanara for critical reading of the manuscript.

**REFERENCES**

1. Weis, K. (1998) *Trends Biochem. Sci.* **23**, 185–189
2. Mattaj, I. W., and Englmeier, L. (1998) *Annu. Rev. Biochem.* **67**, 265–306
3. Gorlich, D., and Kutay, U. (1999) *Annu. Rev. Cell Dev. Biol.* **15**, 607–660
4. Hood, J. K., and Silver, P. A. (1999) *Curr. Opin. Cell Biol.* **11**, 241–247
5. Hood, J. K., and Silver, P. A. (2000) *Biochim. Biophys. Acta* **1471**, M31–M41
6. Jans, D. A., Xiao, C. Y., and Lam, M. H. (2000) *BioEssays* **22**, 582–584
7. Dingwall, C., and Laskey, R. A. (1991) *Trends Biochem. Sci.* **16**, 478–481
8. Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984) *Cell* **39**, 499–509

**Fig. 5.** A comparison of the binding energy increase in a bipartite NLS to the energy decrease resulting from auto-inhibition. The binding affinities of monopartite NLS variants for ΔIBB importin α are compared with the binding affinities of full-length importin α for the bipartite versions of the same NLS variants. Energies are calculated as $-\Delta T \ln K_d$.

**Fig. 6.** An *in vitro* energy scale for nuclear localization. The relative dissociation constants, both measured and calculated, for various peptide sequences in complex with importin α are shown on two logarithmic scales. On the left is a scale representing the affinity for the importin α-importin β complex (as approximated using ΔIBB importin α). The scale to the right represents the affinity for full-length importin α alone (with inhibition from the IBB domain). Using the current model for protein import, the *left* scale represents the ability for each peptide to direct the cytoplasmic capture and translocation of the cargo to which the peptide is fused. The scale on the *right* would then represent the efficiency with which the cargo would be released in the nucleus. Functional studies suggest a lower threshold for a functional NLS on these scales, which exists between the affinities of the SV40 NLS and the SV40T3 variant. This model predicts the possible existence of an upper threshold for the affinity of an NLS sequence above which nuclear release by importin α would be inefficient.
9. Robbins, J., Dilworth, S. M., Laskey, R. A., and Dingwall, C. (1991) Cell 64, 615–623
10. Weis, K., Ryder, U., and Lamond, A. I. (1996) EMBO J. 15, 1818–1825
11. Gorlich, D., Henklein, P., Laskey, R. A., and Hartmann, E. (1996) EMBO J. 15, 1810–1817
12. Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. (1998) Cell 94, 183–204
13. Moroianu, J., Blobel, G., and Radu, A. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 6572–6576
14. Moroianu, J., Blobel, G., and Radu, A. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 2008–2011
15. Radu, A., Moore, M. S., and Blobel, G. (1995) Cell 81, 215–222
16. Moroianu, J., Hjikata, M., Blobel, G., and Radu, A. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 6532–6536
17. Fanara, P., Hodel, M. R., Corbett, A. H., and Hodel, A. E. (2000) J. Biol. Chem. 275, 21218–21223
18. Kobe, B. (1999) Nat. Struct. Biol. 6, 388–397
19. Rexach, M., and Blobel, G. (1995) Cell 83, 683–692
20. Bischoff, F. R., and Ponstingl, H. (1995) Methods Enzymol. 257, 135–144
21. Cingolani, G., Peitsa, C., Weis, K., and Muller, C. W. (1999) Nature 399, 221–229
22. Chook, Y. M., and Blobel, G. (1999) Nature 399, 230–237
23. Izaurralde, E., Kutay, U., von Kobbe, C., Mattaj, I. W., and Gorlich, D. (1997) EMBO J. 16, 6535–6547
24. Moore, M. S. (1998) J. Biol. Chem. 273, 22857–22860
25. Colledge, W. H., Richardson, W. D., Edge, M. D., and Smith, A. E. (1986) Mol. Cell. Biol. 6, 4136–4139
26. Dang, C. V., and Lee, W. M. (1988) Mol. Cell. Biol. 8, 4048–4054
27. Makkerh, J. P. S., Dingwall, C., and Laskey, R. A. (1996) Curr. Biol. 6, 1025–1027
28. Conti, E., and Kuriyan, J. (2000) Struct. Fold. Des. 8, 329–338
29. Johnson-Saliba, M., Siddon, N. A., Clarkson, M. J., Tremethick, D. J., and Jans, D. A. (2000) FEBS Lett. 467, 169–174
30. Hubner, S., Xiao, C. Y., and Jans, D. A. (1997) J. Biol. Chem. 272, 17191–17195
31. Xiao, C. Y., Hubner, S., and Jans, D. A. (1997) J. Biol. Chem. 272, 22191–22198
32. Hu, W., and Jans, D. A. (1999) J. Biol. Chem. 274, 15820–15827
33. Pres, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1992) in Numerical Recipes: The Art of Scientific Computing, pp. 689–699, Cambridge University Press, Cambridge, United Kingdom
34. Xiao, C. Y., Jans, P., and Jans, D. A. (1998) FEBS Lett. 440, 297–301
35. Brugger, A. T., Adams, P. D., Clore, G. M., DeLone, W. L., Gres, P., Grose-Kunslette, R. W., Jiang, J. S., Kuszeawski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonsen, T., and Warren, G. L. (1998) Acta Crystallogr. D Biol. Crystallogr. 54, 905–921
36. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Proteins 11, 281–296
37. Munoz, V., and Serrano, L. (1997) Biopolymers 41, 495–509
38. Efthymiadis, A., Shao, H., Hubner, S., and Jans, D. A. (1997) J. Biol. Chem. 272, 22134–22139