Cell, Vol. 94, 193–204, July 24, 1998, Copyright ©1998 by Cell Press

Crystallographic Analysis of the Recognition of a Nuclear Localization Signal by the Nuclear Import Factor Karyopherin α

Elena Conti,* Marc Uy,*‡ Lore Leighton,*‡ Günter Blobel,†‡ and John Kuriyan*‡§

*Laboratories of Molecular Biophysics

†Laboratory of Cell Biology

‡Howard Hughes Medical Institute

The Rockefeller University

New York, New York 10021

Summary

Selective nuclear import is mediated by nuclear localization signals (NLSs) and cognate transport factors known as karyopherins or importins. Karyopherin α recognizes "classical" monopartite and bipartite NLSs. We report the crystal structure of a 50 kDa fragment of the 60 kDa yeast karyopherin α, in the absence and presence of a monopartite NLS peptide at 2.2 Å and 2.8 Å resolution, respectively. The structure shows a tandem array of ten armadillo repeats, organized in a right-handed superhelix of helices. Binding of the NLS peptide occurs at two sites within a helical surface groove that is lined by conserved residues. The structure reveals the determinants of NLS specificity and suggests a model for the recognition of bipartite NLSs.

Introduction

Topogenic sequences are protein sequence determinants that specify the intracellular location of proteins (Blobel, 1980). Distinct topogenic sequences are recognized by specific receptors that function as the primary effectors for the distribution of proteins to specific cellular locations. Topogenic sequences for protein import into the nucleus (nuclear localization sequences [NLSs]) and for protein export from the nucleus (nuclear export sequences [NESs]) and their cognate receptors, termed karyopherins (Kaps), have been identified. In fact, the existence of several distinct NLSs and NESs and their cognate Kaps has been demonstrated, indicating that protein import as well as protein export proceeds via distinct pathways (for reviews, Görlich and Mattaj, 1996; Nigg, 1997; Pemberton et al., 1998; Wozniak et al., 1998). The various karyopherin substrate complexes bind to a subset of nucleoporins (a collective term for nuclear pore complex [NPC] proteins). Transport across the central gated channel of the NPC is assisted by additional factors, such as the small GTPase Ran and various Ran interacting proteins. The detailed mechanisms for transport across the NPC are still unknown (Ohno et al., 1998; Yang et al., 1998).

The first (now often termed "classical") nuclear import pathway that was discovered operates with the karyopherin α/β1 heterodimer, also known as importin α/β (Adam and Adam, 1994; Enenkel et al., 1995; Görlich et al., 1995; Imamoto et al., 1995; Moroianu et al., 1995a; Radu et al., 1995). Kapα functions in binding the so-called classical NLSs (Adam and Adam, 1994; Görlich et al., 1994; Moroianu et al., 1995a; Weis et al., 1995), whereas Kapβ1 docks the heterotrimeric complex to a subset of nucleoporins (Moroianu et al., 1995b; Radu et al., 1995). In contrast to this classical nuclear import pathway, other recently discovered pathways for nuclear import and export use different Kapβ1 homologs that bind directly to their cognate NLSs and NESs (i.e., without a Kapα adaptor [see reviews]).

Kapα is the product of an essential gene originally identified as a suppressor of temperature-sensitive RNA polymerase I mutations in yeast and named Srp1 (Yano et al., 1992). Subsequently, *Saccharomyces cerevisiae* Kapα was isolated from yeast cytosol as a heterodimeric complex with another protein that was identified as yeast Kapβ1 (Enenkel et al., 1995). Kapα is a 60 kDa molecular weight protein, with hydrophilic amino- and carboxy-terminal regions and a large central domain that consists of tandemly repeated modules known as armadillo (arm) motifs (Görlich et al., 1994; Weis et al., 1995). The arm-repeat domain is linked by a variable spacer (≈30 amino acid residues) to a small N-terminal region that is rich in conserved basic residues and is required for interaction with karyopherin β (Görlich et al., 1996; Moroianu et al., 1996; Weis et al., 1996). In vitro binding of Kapα to NLS was found to be considerably enhanced in the presence of Kapβ (Moroianu et al., 1995b; Rexach and Blobel, 1995). *S. cerevisiae* has a single gene for Kapα that shares a remarkable degree of sequence conservation with α homologs encoded by multigene families in higher eukaryotes (amino acid identity >40%).

Arm motifs have been found in several functionally unrelated proteins such as the *Drosophila melanogaster* segment polarity protein and β-catenin (Peifer et al., 1994). The crystal structure of β-catenin was determined recently, revealing that the repetitive arrangement of the motifs forms an elongated, α-helical molecule (Huber et al., 1997). The structure of β-catenin suggests that arm repeats provide a structural framework that is able to mediate diverse protein–protein interactions by virtue of the extensive surface groove formed by the arrays of α helices. The arm-repeat domain of Kapα has been proposed to harbor the binding site for NLS motifs (Cortes et al., 1994; Sekimoto et al., 1997).

Classical NLS motifs consist of short sequences (5–20 residues) containing several lysine and arginine residues (Dingwall and Laskey, 1991). When fused to a heterologous protein, these small NLS motifs are sufficient to direct the chimeric polypeptide to the nucleus (Kalderon et al., 1984). The simplest NLSs contain a single cluster of positively charged residues and are generally referred to as monopartite signals. A second class consists of bipartite NLSs where two stretches of basic residues are close together in the amino acid sequence but not actually adjacent (Table 1). Both variants of the classical NLS are recognized by Kapα homologs.

A prototypical nuclear localization signal is that of the simian virus 40 (SV40) large T antigen, which contains

§To whom correspondence should be addressed.

Table 1. Monopartite and Bipartite Nuclear Localization Sequences

NLS Source	NLS Sequence (Seq Numbering)	Nuclear Localization
Monopartite NLS		
SV40	P**KKKRK**V(132)	+ (Kalderon et al., 1984)
SV40 mutant	P**KT**A**RK**V(132)	− (Kalderon et al., 1984)
H2B	G**KKR**S**K**V(35)	+ (Moreland et al., 1987)
v-Jun	KS**RKRK**L(253)	+ in G2 (Chida and Vogt, 1992)
c-*myc*	PAA**KR**V**K**LD(328)	+ (Dang and Lee, 1988)
c-*myc* mutant	PAA**KT**A**KR**LD(328)	− (Makkerh et al., 1996)
Consensus	**KK**x**K**	
Bipartite NLS		
Nucleoplasmin	**KR**PAATKKAGQA**KKKK**L(171)	+ (Robbins et al., 1991)
SV40 mutant	**KR**TADSQHSTPP**KT**A**RK**V(132)	+ (Makkerh et al., 1996)
Consensus	**2K/R** 10–12 aa **3K/R**	

The important features of monopartite and bipartite NLSs are highlighted; bold letters indicate essential basic residues, while important neutral and acidic amino acids are underlined. The sequence numbers of the carboxy-terminal residues are indicated in parentheses. Their nuclear or cytosolic localization in the cell cycle is denoted with + or −.

a single cluster of five contiguous positively charged residues in the sequence ^{126}Pro-Lys-Lys-Lys-Arg-Lys-Val^{132} (Kalderon et al., 1984; Lanford and Butel, 1984). Mutation of individual basic residues in the SV40 T antigen NLS results in either nuclear accumulation or cytoplasmic retention, depending on which particular residue is replaced (Table 1). While Lys-128 is crucial for nuclear import, mutation of one of the adjacent basic residues reduces import but does not abolish it (Colledge et al., 1986). Interestingly, certain relatively rare NLS motifs have been identified that are able to localize a protein to the nucleus with as few as three positively charged amino acids, provided they are flanked by specific nonbasic residues (see the c-*myc* sequence in Table 1, Dang and Lee, 1988; Makkerh et al., 1996).

The bipartite NLS motif contains two interdependent positively charged clusters separated by a mutation-tolerant linker region of 10–12 amino acids, as found in the nucleoplasmin nuclear localization signal (Robbins et al., 1991). In bipartite NLSs, the upstream cluster consists of two lysines or arginines, while the downstream cluster bears strong resemblance to the SV40 T antigen NLS motif, although it is not sufficient by itself to localize a protein to the nucleus (Dingwall et al., 1988). The presence of an upstream cluster appears to relax the stringent sequence requirements of monopartite signals. Indeed, the Thr-128 mutant of the SV40 T antigen that is defective in nuclear import can be rescued by placing two basic amino acid residues ten residues upstream of the defective NLS (Table 1), presumably converting an ineffective single basic cluster into an active bipartite signal (Makkerh et al., 1996).

Little is known at present regarding the molecular mechanism underlying the recognition of nuclear localization signals by karyopherin α. In the absence of three-dimensional structural information, it has been difficult to explain the diversity of NLS motifs, as well as the constraints on their sequences. We now report the crystal structures of the entire arm-repeat region of *S. cerevisiae* karyopherin α in the unliganded form and in a complex with a peptide corresponding to the NLS of the SV40 large T antigen. The structure shows that the repeating architectural pattern of the arm motifs is utilized by the protein to generate an array of binding sites that are specific for lysine and arginine. These sites are formed within a long surface groove that is lined by conserved solvent-exposed residues. The number and placement of the binding sites dictate the sequence requirements for monopartite and bipartite nuclear localization signals. The protein forms a homodimer in the crystals, and this self-association has implications for a possible autoinhibitory mechanism.

Results and Discussion

Domain Identification and Structure Determination

Limited proteolysis was used to delineate the structural domains of yeast Kapα (Figure 1A). Proteolytic digestion using a very high concentration of subtilisin (approximate ratio of protein:protease of 1:1) resulted in a stable fragment encompassing residues 46–530, as verified by mass spectrometry and N-terminal sequencing (data not shown). Digestion with trypsin yielded a similar fragmentation pattern, suggesting that the whole arm-repeat region and part of the N-terminal domain together form a stable structure. However, this protease-resistant fragment failed to give crystals of good quality. Instead, crystals diffracting to high resolution were obtained with a shorter fragment, comprising residues 88–530, which consists of the entire conserved arm-repeat domain (Figure 1A) and spans the binding site for nuclear localization sequences (Cortes et al., 1994; Moroianu et al., 1996; Sekimoto et al., 1997). The boundaries of this 50 kDa fragment, which we shall refer to as Kapα50, were determined based on sequence homology and the results of the limited proteolytic digestion.

The structure of the unliganded Kapα50 was determined by conventional multiple isomorphous replacement methods (Table 2) and refined at 2.2 Å resolution to a crystallographic R value of 23.7% and a free R value of 27.5% with good stereochemistry. The diffraction data exhibit significant anisotropy, resulting in somewhat higher R values than expected (see Experimental Procedures). Crystals of Kapα50 contain two molecules per asymmetric unit. The refined model comprises residues 89–509 of karyopherin α, 140 water molecules, and a cobalt ion, while the 20 carboxy-terminal residues are disordered and not modeled.

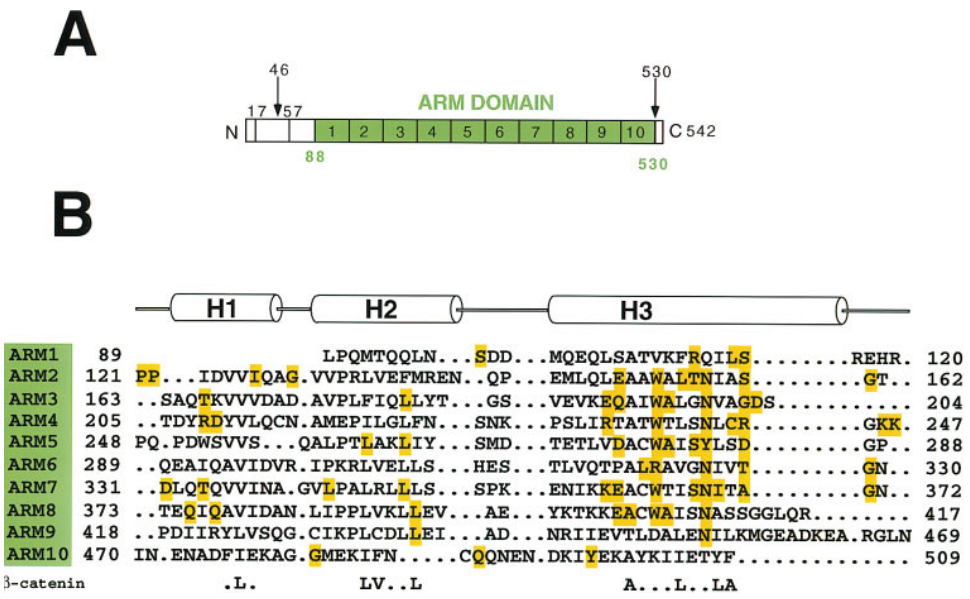

Figure 1. Primary and Secondary Structure of Yeast Karyopherin α

(A) Schematic of the domain structure, with sequence numbering of the conserved N-terminal domain (residues 17–57) and the large arm-repeat domain (residues 88–530). The locations of the cleavage sites that produce a stable product by subtilisin treatment are shown by numbered arrows. The green portion represents the molecular limits of the construct used for structure determination. The arm repeats are numbered 1–10.

(B) Structure-based sequence alignment of the ten arm motifs in yeast Kapα. The residue numbers at the beginning and the end of each repeat are indicated. The top line illustrates the structural elements of the repeat: helices H1, H2, and H3. The consensus shown at the bottom is from the analysis of arm repeats in β-catenin (Huber et al., 1997) and specifies the position of the hydrophobic core residues that are characteristic of arm motifs. The amino acids that are strictly conserved in 13 karyopherin α sequences are shaded in orange.

We have been unable to cocrystallize Kapα50 with NLS containing–peptides. In addition, attempts to soak crystals of the unliganded protein with various NLS peptides led almost invariably to severe degradation of the diffraction pattern. In the case of the SV40 T antigen NLS, we increased the concentration and soak time of the peptide (Ser-Pro-Lys-Lys-Lys-Arg-Lys-Val-Glu) to the point where the crystals just began to crack visibly. We were able to obtain a data set to 2.8 Å resolution and have determined the structure of the peptide complex of Kapα50 from these data. The liganded structure is currently refined to 2.8 Å resolution, with working and free R values of 25.6% and 30.7%, respectively. The model includes residues 89–509 of Kapα, two SV40 T antigen peptides, and four water molecules. Six and two side chains of the peptide are well localized in the

Table 2. Data Collection and MIR Statistics

Data Set	Native	PbAc	UAc	$AuCN_2$	NLS
Synchroton beamline[a]	X25	A1	A1	A1	A1
Maximum resolution (Å)	2.2	2.9	3.0	2.9	2.8
Independent reflections	48,255	22,898	20,738	22,448	22,717
No. of measurements	525,332	427,630	334,913	331,756	270,860
Completeness (%)[b]	95.0 (95.0)	98.1 (93.0)	98.8 (98.6)	90.7 (77.7)	86.9 (82.0)
R_{sym} (%)[b]	6.5 (29.9)	7.7 (14.8)	9.4 (29.6)	8.0 (12.7)	6.4 (18.0)
I/σ[b]	13.6 (2.8)	23.0 (10.3)	19.4 (6.7)	8.6 (5.9)	17.4 (9.9)
Heavy atom concentration (mM)		1	3	15	
No. of sites		2	2	2	
R_{iso} (%)[c]		34.1 (41.0)	24.6 (31.1)	10.8 (16.5)	
Phasing power[c]		0.73 (0.68)	0.99 (0.93)	0.92 (0.80)	
R_{cullis} (%)[c]		0.92 (0.92)	0.81 (0.92)	0.83 (0.92)	

$R_{sym} = \Sigma |I - \langle I \rangle| / \Sigma I$, where I is the integrated intensity of a given reflection.
$R_{iso} = \Sigma |F_{PH} - F_P| / \Sigma F_P$, where F_{PH} and F_P are the derivative and native structure factor amplitudes.
Phasing power = <rms heavy atom structure factor>/<rms lack of closure> for acentric reflections.
R_{cullis} = <rms lack of closure>/<rms isomorphous difference> for centric reflections.
NLS, nuclear localization signal peptide from the SV40 antigen (residues 125–133).
[a] X25, BNL; A1, CHESS.
[b] Values for the outermost resolution shell are given in parentheses.
[c] Statistics calculated to 3.0 Å resolution, with the outermost resolution shell given in parentheses.

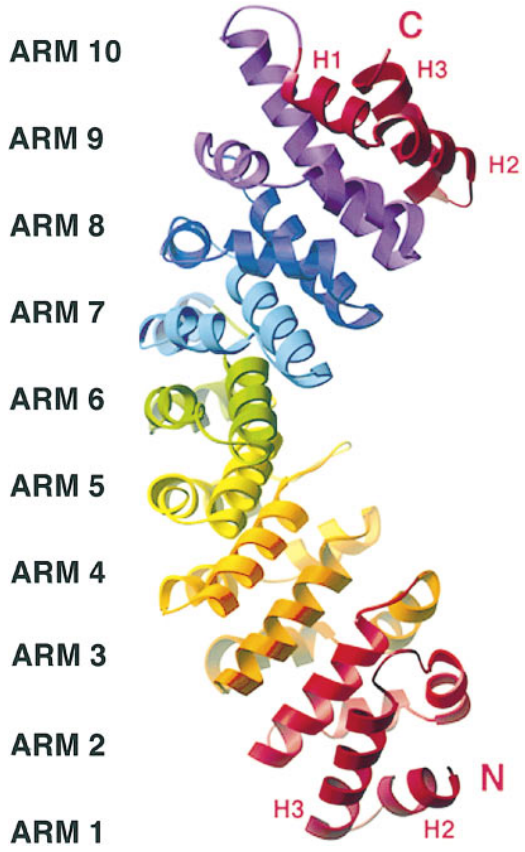

Figure 2. Three-Dimensional Structure of Kapα50

The molecule contains ten tandem arm repeats, which are shown in different colors. With the exception of the first motif, each arm repeat includes three α helices (H1, H2, and H3). The superhelical axis of the molecule is vertical. All ribbon diagrams were generated using RIBBONS (Carson, 1991).

electron density maps at the two binding sites, respectively. The average temperature factor for the backbone and side chain atoms for the two protein molecules in the asymmetric unit are 41.4 and 41.0 Å^2, respectively. The temperature factors for the peptide are very similar (44.7 Å^2 and 44.8 Å^2, respectively). This suggests that the two peptide binding sites are fully occupied, which is also consistent with the electron density maps (see Experimental Procedures).

Overall Architecture

Kapα50 folds into one contiguous structural domain (Figure 2). The overall shape of the domain is cylindrical, with a height of 100 Å and a diameter of 30 Å. The molecule is built by the tandem stacking of ten arm repeats in much the same way as reported for the 12 arm–containing protein β-catenin (Huber et al., 1997). The structural units are arranged roughly parallel, with neighboring repeats separated by an average translation of 9 Å. A rotation of approximately 30° between adjacent structural units causes a right-handed superhelical twist. The superhelical axis roughly follows a straight course for the whole length of the molecule; the 50° bend observed at the ninth repeat of β-catenin is absent in Kapα50 (Huber et al., 1997). Superposition of Kapα with β-cateinin results in 395 topologically equivalent α carbons with an rmsd of 3.8 Å and an overall sequence identity of 17%.

The arm motif consists of about 40 amino acid residues that form three helices (Figure 1B). As in β-catenin, the choice of which three helices make up a particular arm repeat is arbitrary, since the tight and repetitive packing of the helices leads to a contiguous hydrophobic core that extends throughout the structure. An alignment of the sequences of the ten arm repeats of Kapα is shown in Figure 1B, along with a consensus sequence motif that was derived from the β-catenin structure. Despite the high variability in their amino acid sequences, arm motifs share very similar structures due to the conservation of this hydrophobic core. The rms difference between Cα atoms of pairs of Kapα50 arm repeats is 0.9 Å on average. As in the case of β-catenin, the first arm motif lacks the H1 helix; however, in both cases the preceding residues are not present in the protein fragments that were crystallized. The first and the last arm repeat in the Kapα50 structure display a weaker level of sequence similarity to the arm consensus sequence, and their existence had been overlooked in previous sequence analyses.

A Surface Groove Presents Highly Conserved Solvent-Exposed Side Chains

The roughly triangular cross section of each arm repeat results in the formation of a shallow groove, which spirals about the superhelical axis of the molecule. The concave surface of this groove is shaped by a ladder of parallel H3 helices. A similar array of parallel H2 helices, flanked by the short H1 helices, forms the convex surface of the molecule (Figure 2). A similar arrangement is present in the structure of β-catenin, although the residues lining the surface groove are different in the two proteins (Huber et al., 1997). A remarkable feature of the Kapα50 structure is the presence within the groove of a large number of solvent-exposed and highly conserved side chains (Figure 3). Indeed, more than 70% of the invariant residues in the Kapα family are present on the surface of the molecule and are contributed predominantly by the H3 helices that are between the second and the eighth arm repeats. This is a striking observation, since in general highly conserved residues are found most often in the hydrophobic core of proteins and not on their surfaces.

Strictly conserved tryptophans are present at analogous positions in the third turn of the H3 helices of six arm motifs (Figure 1B). The side chains of these tryptophan residues are well-defined in the electron density maps, despite being completely exposed to solvent. In addition, a set of invariant asparagine residues is found four residues downstream from most of the tryptophans. The tryptophan-asparagine pairs are located one turn of a helix apart, and they generate a ridge that spans the concave surface of the molecule (Figure 4). The regularity of the array is interrupted at the fifth and sixth arm repeats, where an invariant tyrosine and an invariant arginine substitute for the asparagine and the tryptophan residues, respectively. The

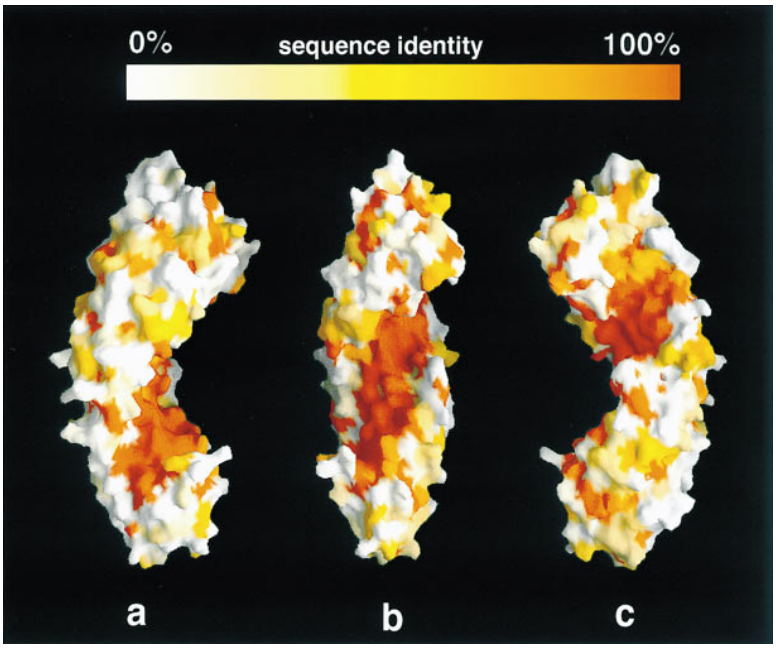

Figure 3. Conservation of the Molecular Surface of Kapα50

The molecular surface is shown in three views (a, b, and c), related by successive 90° rotations about the superhelical axis. In view a, the molecule has a similar orientation as in Figure 2, with the N terminus at the bottom and the C terminus at the top. The coloring scheme represents the level of amino acid similarity observed on aligning 13 sequences of karyopherin α. The colors range from 0% (white) to 100% (orange) sequence identity. Amino acid residues were each assigned a parameter as a measure of the degree of conservation (P. Brick, unpublished); the parameter values were mapped onto a color scale and displayed on the surface of the molecule using GRASP (Nicholls et al., 1991).

ridge of absolutely conserved residues shapes the base of the surface groove and is surrounded by conserved polar and charged amino acids from the H1 and H3 helices.

Karyopherin α Contains Two NLS Binding Sites

The SV40 T antigen NLS binds in an extended conformation at two sites on the concave surface of the arm domain of karyopherin α (Figure 4). One site is larger and binds the Lys-Lys-Lys-Arg-Lys-Val portion of the nonameric peptide used in the structure determination. These six amino acids correspond to residues 127–132 (hereafter denoted P_1 to P_6) of the SV40 T antigen sequence. The larger binding site is in the N-terminal region of Kapα50, between the second and the fourth arm motifs (residues 121–247). The smaller binding site is located at the seventh and eighth arm repeats (residues 331–417), and it recognizes only two residues specifically. The first and the ninth arm repeats are not directly involved in recognition of the NLS but contribute essential interactions to the maintenance of the binding site structure.

Similar features govern the binding of the SV40 T antigen NLS peptide at both sites. The peptide chains run antiparallel to the overall direction of the Kapα polypeptide. Their main chain roughly follows the superhelical axis of the molecule, running along the ridge formed by the absolutely conserved tryptophan and asparagine pairs of the parallel H3 helices that form the long surface groove (Figure 5A). The interactions between the NLS and Kapα50 can be divided into three components. First, the direction of the NLS main chain is determined by backbone interactions with the conserved asparagine side chains of Kapα50. Second, the aliphatic portions of the lysine side chains of the NLS are bound in shallow grooves on the surface of the protein, some of them shaped by the conserved tryptophan side chains of Kapα50. Finally, the extended lysine side chains form salt bridges with conserved negatively charged residues that surround the hydrophobic floor of the surface groove.

The structure of the NLS-Kapα complex is consistent with the results of Sekimoto et al (1997), who mapped the binding site for the SV40 T antigen NLS to the region covering the first and the ninth arm repeats of human karyopherin α1 (NPI-1). In this context, the binding of the SV40 T antigen NLS to the very C-terminal region of the protein observed by Moroianu et al. (1996) is likely to be due to nonspecific binding in the presence of large excess of peptide.

Interactions with the Asparagine Array

At the larger binding site, the peptide backbone of the NLS is held in an extended conformation by evenly spaced hydrogen bonds with Kapα (Figure 5B). Asn-199 and Asn-157 make bidentate hydrogen bonding interactions with the main chain amide groups of the lysines at positions P_3 and P_5 of the NLS peptide. The asparagine rotamers are fixed by donor–acceptor interactions of their amino group with the main chain carbonyl oxygens of Leu-115 and Thr-156, respectively. The NLS backbone deviates slightly from this pattern at position P_1, where the carbonyl oxygen and amide nitrogen of the NLS backbone are about 3.5 Å from the corresponding side chain nitrogen and oxygen of Asn-241 (Figure 5B). This region is close to a crystallographic contact, and the intermolecular contacts may be perturbed.

A very similar pattern of asparagine-based interactions has been found to anchor peptides in class II MHC proteins (Stern et al., 1994) and might therefore represent a general mechanism used to extend and orient peptides along protein binding sites. Intriguingly, Asn residues are found at corresponding locations in several of the H3 helices in β-catenin. The utilization of asparagine residues rather than glutamine to bind and orient the peptide main chain is most likely a consequence of

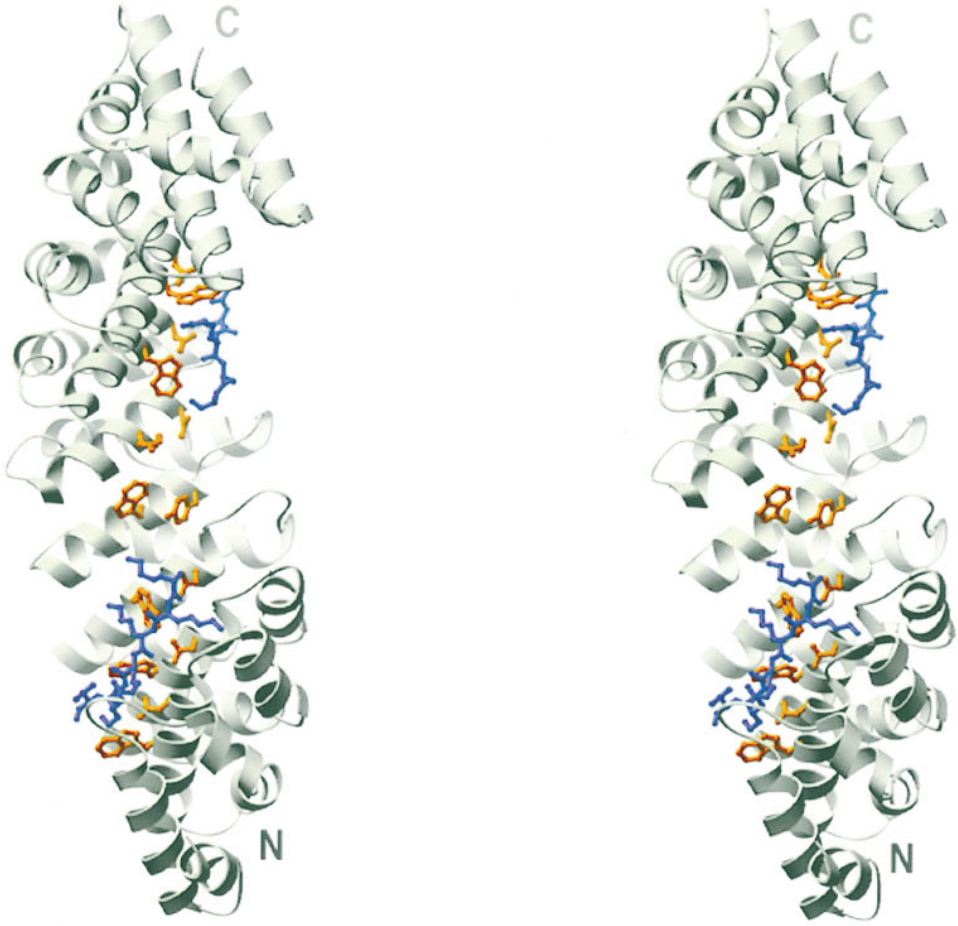

Figure 4. Stereo View of Kapα50 with the SV40 T Antigen NLS Peptide Bound at Two Sites

Stereo drawing of a ribbon representation of Kapα50 shown in a similar orientation to view b in Figure 3. Highlighted in orange is the ridge of conserved residues presented by the H3 helices, which spirals about the superhelical axis, forming the floor of the surface groove. The SV40 T antigen peptides (blue) bind at a larger (below) and a smaller (above) site in the conserved groove.

the less flexible nature of the Asn side chains, which minimizes entropic loss upon ligation.

Interactions with the Tryptophan Array

The side chains of the NLS peptide form extensive hydrophobic and van der Waals interactions with Kapα50. As a result of the extended β-strand conformation of the NLS peptide, the side chains of the residues at P_2, P_4, and P_6 protrude on one side into mostly apolar pockets, while the lysines at P_1, P_3, and P_5 extend on the opposite side and pack against the aromatic ring systems of three tryptophan residues (Figure 5B).

The almost parallel arrangement of the side chains of Trp-237, Trp-195, and Trp-157 generates three equidistant and apolar shallow pockets that are able to accommodate the aliphatic portion of the lysine side chains of the NLS (Figure 5C). To complete the network of hydrophobic interactions, Trp-153 is sandwiched between the peptide lysine residue at P_5 and the side chain of Arg-112, which is invariant in all Kapα proteins. This arginine is in turn positioned by a salt bridge interaction with Glu-150, such that its N_ϵ atom is 3.0 Å away from the center of the six-membered ring of Trp-153. The guanidinium moiety of Arg-112 is thus at an optimal distance for participating in an amino-aromatic interaction with the π electrons of the Trp-153 side chain (Burley and Petsko, 1988). Another key structural interaction that orients the Trp-153 side chain is the hydrogen bonding of the indole nitrogen with the side chain of Ser-116. A Ser-116→Phe mutant of yeast Kapα is defective in nuclear import and causes the arrest of cell cycle progression (Loeb et al., 1995); this point mutation is likely to disrupt the structural integrity of the NLS binding site.

Electrostatic Interactions

Negatively charged residues are arrayed along the outer edge of the NLS binding site, removed from the vicinity of all but the terminal atoms of the peptide side chains. For example, the lysine at P_2 contacts the main chain carbonyl of Gly-161, the hydroxyl group of Thr-166, and the carboxylate of Asp-203 (Figure 5B). These three oxygen atoms of Kapα are well positioned to coordinate all three hydrogen atoms of the sp^3 hybridized nitrogen of the NLS side chain. Asp-203 is partially buried and makes hydrogen bonding interactions with two serine side chains (residues 163 and 204). On the opposite

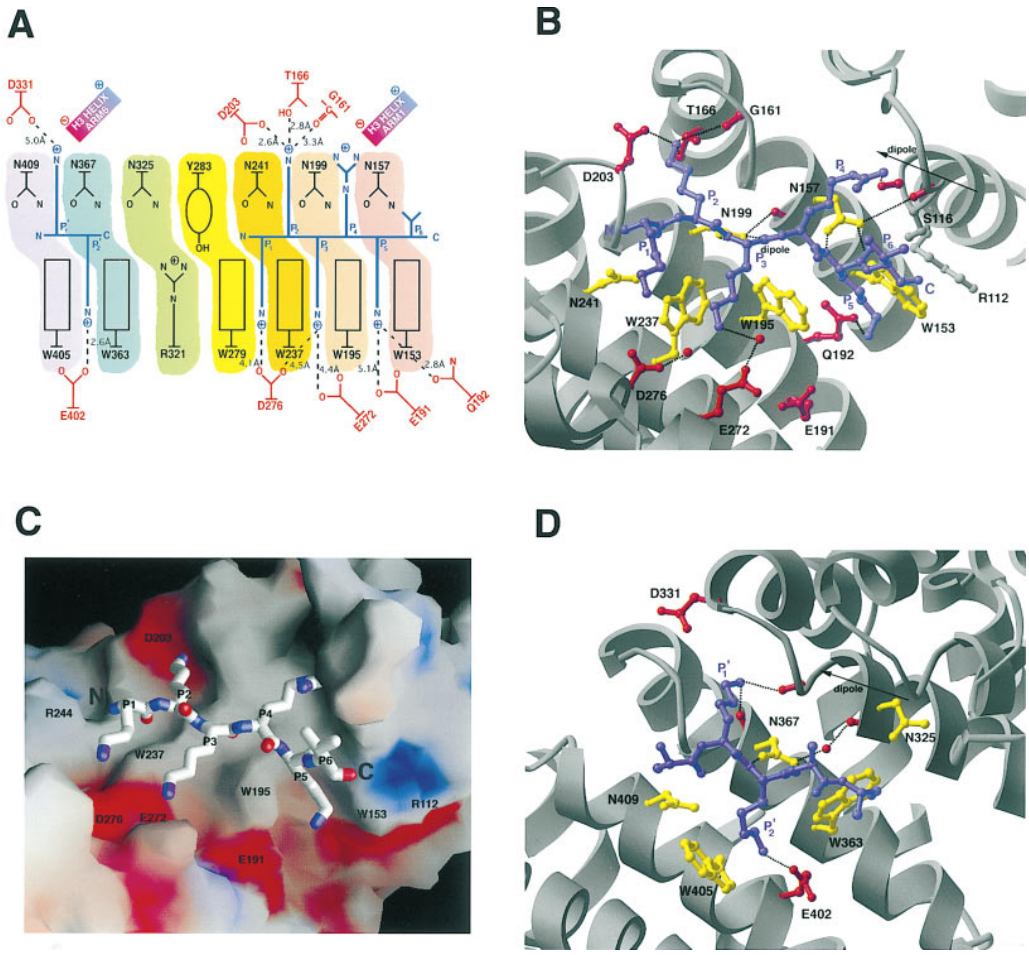

Figure 5. Specificity of NLS Binding to Karyopherin α

(A) Schematic representation showing the position of the two SV40 T antigen peptides (in blue) within the ridge of conserved Trp and Asn residues. In red are negatively charged and polar residues, while helix dipoles are shown as rectangles. Six side chains of the NLS are visible at the larger binding site (denoted P_1 to P_6), while only two are recognized at the minor site (P_1' and P_2'). The peptide binds at the Trp-Asn pairs of arm2, arm3, and arm4 (153–157, 195–199, and 237–241) at the larger binding site and of arm7 and arm8 (363–367 and 405–409) at the smaller site. The regular array of Trp-Asn pairs is interrupted at the fifth and sixth arm motifs by Tyr-283 and Arg-321. The conserved pairs are colored per arm repeat (as in Figure 2).

(B) Interactions of the SV40 T antigen NLS peptide at the larger N-terminal site of Kapα50. The peptide is shown in blue, with residues 127–132 denoted P_1 to P_6. The asparagine and tryptophan residues of the protein are in yellow, while polar and acidic residues are in red. Water molecules are represented as red spheres. The backbone of Kapα is shown in gray, together with two protein side chains that, though not engaged in direct recognition of the peptide, are involved in proper structuring of the binding site. Polar and electrostatic interactions are indicated with dotted lines.

(C) Molecular surface of the Kapa50 larger binding site colored according to electrostatic potential, with the bound SV40 T antigen peptide. The molecule is shown in a view similar to Figure 5B. The surface was generated with GRASP (Nicholls et al., 1991) using dielectric constants of 2.0 for the protein interior and 80.0 for solvent at an ionic strength equivalent to 0.1 M monovalent salt. Regions of positive charge on the surface of Kapα50 appear blue, electronegative regions are red, with neutral regions colored white. The location of several residues labeled in Figure 5B are indicated.

(D) Interactions at the smaller binding site. The peptide is in blue, with the two bound side chains labeled P_1' and P_2'. The ribbon representation of the molecule is in gray, and residues at the binding site are colored as in Figure 5B.

side, the lysine at P_3 contacts Glu-272 and Asp-276 through water molecules, while the P_5 lysine amino group hydrogen-bonds with the side chain of Gln-192 (Figure 5B). The residue at P_1 is involved in long range electrostatic interactions, with its amino group positioned 4.1 Å and 4.5 Å from the carboxylate oxygens of Asp-276. The arginine side chain at position P_4 is not complemented electrostatically by interaction with negatively charged residues. However, a hydrogen bond to the main chain carbonyl of residue 117 anchors the arginine guanidinium group at the carboxy terminus of the arm1 H3 helix, thus stabilizing its positive charge by interaction with the helix dipole.

The Smaller Binding Site

A smaller NLS binding site is located in the C-terminal region of the arm domain of karyopherin α. The main chain of five peptide residues together with the side chains of two lysines (denoted P_1' and P_2') are well localized in the electron density. The peptide backbone and

the two lysine side chains are bound by a pattern of interactions similar to those described at the larger binding site (Figure 5D). There are no obvious recognition sites for the basic residues of the NLS peptide adjacent to the P_1' and P_2' lysines. The relatively limited nature of the interactions between the NLS peptide and this smaller binding site suggests that monopartite NLS sequences might not bind here to a significant extent. Rather, the smaller binding site is likely to be important in the recognition of bipartite NLSs (see below).

The Presence of Two Binding Sites Suggests a Mechanism for the Recognition of a Bipartite NLS

The larger binding site is structured optimally for the recognition of five lysine or arginine residues, three of which can pack against three tryptophans, while the smaller binding site allows specific recognition for two basic residues, one of which can be sandwiched between two tryptophans. The independent recognition of positively charged clusters at two distinct sites on the same Kapα molecule suggests a possible mechanism for the binding of bipartite signals, with simultaneous interactions at both sites. At the larger binding site, the five contiguous positively charged residues span about 14 Å (Cα P_1–Cα P_5). The 28 Å distance, which separates the Cα atoms of the downstream P_1 residue from the upstream P_2' residue, would allow a 10-residue spacer to link the two peptide segments, while a shorter linker would impair the simultaneous binding of the two clusters. This is in agreement with mutational analysis, showing that the two basic stretches can still induce nuclear localization when their separation is increased but not when it is decreased (Robbins et al., 1991).

An additional constraint on bipartite NLS sequences is that the smaller basic cluster is required to be upstream of the larger cluster (Robbins et al., 1991). The backbone orientation of the NLS peptide is determined by the bidentate hydrogen bonding with the asparagine side chains and is consistent with the sequence direction of bipartite NLSs, which always bear their monopartite-like cluster at the C terminus. The maximum number of positively charged residues that can be accommodated by the upstream and the downstream clusters is dictated by the regular and asymmetric distribution of Trp-Asn pairs on the surface groove of karyopherin α (Figure 5A). The regularity of the Trp-Asn array is interrupted at the surface groove between the two binding sites, precisely where the nonbasic linker in a bipartite NLS would be located.

Implications for the Specificity of NLS–Karyopherin Interactions

The binding sites on the karyopherin α surface combine hydrophobic and electrostatic elements in a manner that is uniquely suited for the recognition of lysine and arginine. The tryptophan cradles that interact with the methylene groups of the lysine side chains of the NLS are likely to provide a significant fraction of the total binding energy via the hydrophobic effect. The interaction between the NLS peptide and Kapα50 buries 76% of the peptide surface (385 $Å^2$), thus attesting to the extensive surface complementarity between the ligand and its binding pocket. The crucial aspect of the NLS binding site that then prevents indiscriminate ligation of hydrophobic peptides is the strategic placement of aspartates and glutamates around the edges of the binding site. The presence of these negatively charged residues ensures the avoidance of all but lysine and arginine in the ligand, since a side chain of the ligand that disturbed the solvation of the acidic residues without compensating the charge would pay a severe energetic penalty.

The interaction distances between the lysine amino groups and the acidic side chains are relatively long at most of the sites. The distances between the terminal nitrogen atom of the lysines and the closest carboxylate oxygen of the associated aspartates or glutamates are typically between 4–5 Å (Figure 5A), compared to distances of around 2.5 Å that are expected for charged hydrogen bonds. Despite the lack of tight hydrogen bonding, the long range nature of the electrostatic energy would result in charge complementarity still being a crucial factor over the distance ranges that are seen here. The formation of charged hydrogen bonds on the surfaces of proteins is often destabilizing due to energetic penalties associated with desolvating the charged groups (Hendsch and Tidor, 1994). It may be that by interacting at a somewhat greater distance the lysines and the negatively charged side chains allow water molecules to participate in the interaction to a greater extent, thereby offsetting some of the desolvation energy associated with localizing the peptide on the protein surface. Furthermore, the presence of relatively loose specificity pockets might provide the flexibility required for binding both lysine and arginine residues.

The structure of the NLS-Kapα50 complex explains not only the preference for the positively charged residues in general, but also the relative importance of each of the positions in the NLS motif. The more extensive interactions observed at the P_2, P_3, and P_5 positions of the larger site are consistent with the consensus sequence KKxK seen in monopartite NLS peptides, with the second lysine of the motif (P_2) being strictly required. At the P_2 position, the near optimal nature of the hydrogen bonding geometry around the lysine amino group is likely to be energetically favorable. Another important factor is the coordination of the amino group of the lysine at P_2 by a partially buried negatively charged residue, that of Asp-203. The desolvation term for this interaction is likely to be reduced, since the protein carboxylate group has already been partially sequestered from the solvent.

The arginine at the P_4 site of the SV40 T antigen NLS is stabilized by a helix dipole and does not interact with a negatively charged residue. Consequently, a positive charge at this position of the ligand is not required (see, for example, Histone H2B in Table 1). Similarly, there are monopartite NLSs without a lysine or arginine at P_1 (see, for example, v-Jun in Table 1), consistent with the absence of tight hydogen-bonding interactions at this position (Figure 5A). A negative charge at P_1 would however be disfavored energetically, and this might be responsible for the cell cycle–dependent partitioning of the oncogenic viral Jun protein out of the nucleus upon phosphorylation of the corresponding serine at this position (Chida and Vogt, 1992). In comparison to other

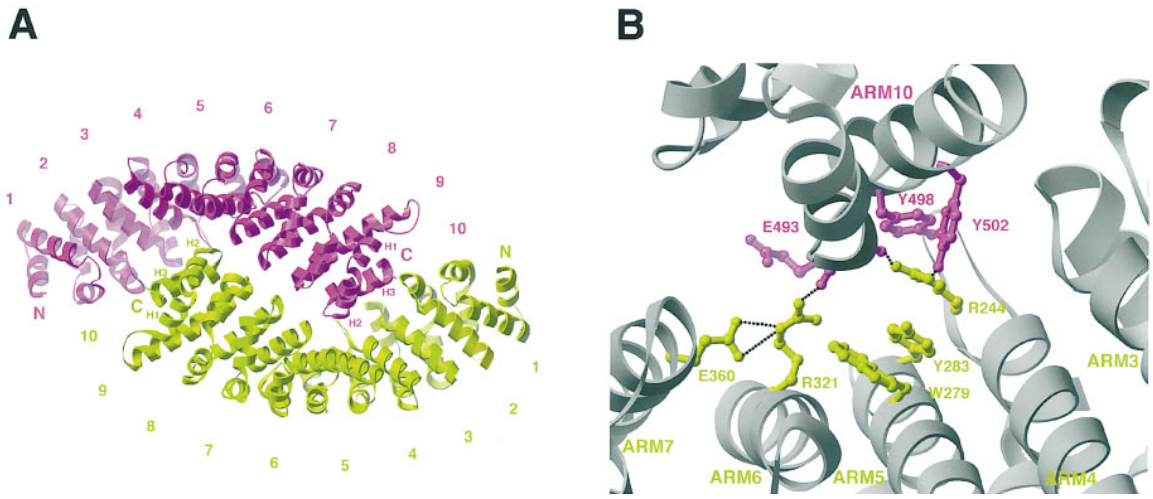

Figure 6. The Kapα50 Dimer in the Crystals

(A) Dimer of Kapα50 viewed approximately down the local molecular dyad axis, with the two monomers colored in green and magenta. The arm repeats are numbered sequentially from the N to the C terminus. The helices of the tenth arm motifs are labeled. Note that the first arm repeat lacks the H1 helix.

(B) The dimer interactions between the tenth arm repeat of one monomer (in gray with relevant residues depicted in magenta) contacting the H3 helices of the central repeats of the second monomer (in gray with important residues highlighted in green). All the labeled side chains are highly conserved. Hydrogen bonding contacts are shown in dotted lines.

monopartite and bipartite NLSs, the SV40 T antigen is efficient in nuclear targeting (Efthymiadis et al., 1997), and the structural analysis shows that it utilizes electrostatic and/or helix dipole interactions at all the specificity pockets, including P_1 and P_4.

Karyopherin α Forms a Dimer in the Crystal with a Conserved Interface

The arm-repeat domain of karyopherin α forms a homodimer in the crystal, in which the two monomers pack side by side via their carboxy-terminal regions (Figure 6A). Kapα50 appears to be homodimeric in aqueous solutions at a concentration of 1 mg/ml, as seen by dynamic light scattering. Since no functional role has previously been assigned to dimer formation in Kapα50, it is important to consider whether this observation is significant.

The dimer interface of Kapα50 is quite extensive and buries approximately 13% (2300 Å^2) of the surface area of each molecule. This is unusually large for normal crystal interfaces (Naismith et al., 1996). There are three main regions of contacts between monomers. A central area is located at the local dyad axis and involves the H2-H3 loops of the eighth and ninth arm motifs (Figure 6A). Two adjacent symmetrical areas of interaction are contributed by the tenth arm repeat of one monomer and the fourth, fifth, and sixth motifs of the other one. Hydrophobic patches and intersubunit hydrogen bonds are distributed across the dimer interface and involve a set of conserved residues, in particular Arg-244, Tyr-283, Arg-321, and Tyr-498 (Figure 6B).

The conserved dimer interface of Kapα50 is between the two NLS binding sites of each monomer. This occludes the intervening surface at the fifth and sixth arm motifs, where Tyr-283 and Arg-321 are found to interrupt the Asn-Trp pairs that mediate NLS binding (Figures 4 and 5A). The nature of the homodimer interface is incompatible with the binding of a bipartite NLS, because the space between the larger and smaller binding sites of the NLS is blocked in the crystal. Although the spacer region linking the two clusters of a bipartite NLS does not have specific sequence requirements, it is intriguing that the interface between the two sites is nonetheless conserved (Figure 3). The conservation of Tyr-283 and Arg-321 and their involvement in monomer–monomer interactions point to a possible functional role for these residues in dimer formation.

The binding of an NLS-containing peptide to Kapα is enhanced by the presence of the Kapβ1 protein. One proposed mechanism for this cooperative binding involves the ligation of the N-terminal region of Kapα by Kapα itself, with activation resulting from the displacement of this interaction by Kapβ1 (Moroianu et al., 1996; but see Görlich et al., 1996). The amino terminus of the α subunit (17–57, yeast numbering) is rich in positively charged residues and contains an internal NLS sequence that might bind to Kapα. Indeed, the limited proteolysis experiments reported here indicate that part of the highly basic N-terminal region of Kapα forms a stable compact fragment with the arm-repeat domain, suggesting that the region spanning residues 47–56 functions as an internal NLS. It is possible that homodimerization provides an additional level of autoinhibition, whereby binding of the β subunit would not only displace the amino-terminal autoinhibitory tail, but also disrupt the α homodimer, resulting in complete access to the binding site.

Concluding Remarks

The clustering of positively charged residues within classical nuclear localization signals suggests that NLS recognition might proceed most simply via electrostatic

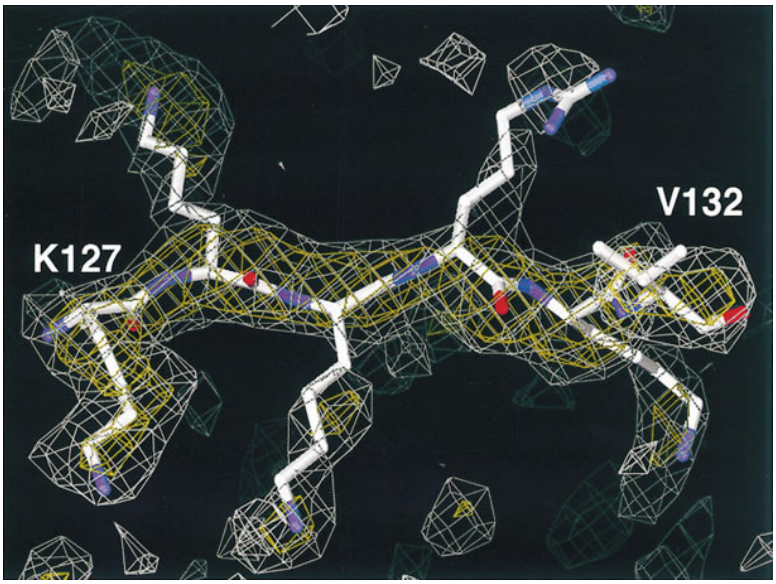

Figure 7. Refined Model of the SV40 T Antigen Signal Peptide at the Larger Binding Site and Electron Density from a Simulated Annealing Omit Map

The peptide was deleted from the model, and ten independent simulated annealing refinements were used to remove bias from the calculated phases, which were obtained by averaging the structure factors calculated from the ten resulting structures. The map was calculated using all data between 30 and 2.8 Å resolution and contoured at 2.0 σ (white) and 3.0 σ (yellow).

pairing between arrays of negatively charged residues of Kapα and the lysines and arginines of the NLS. However, the necessity to first desolvate the charged residues before formation of the intermolecular interface makes it difficult to achieve high affinity by relying exclusively on electrostatic interactions. The structure of the Kapα-NLS complex shows that the repeated arm motifs of Kapα provide a much more complex set of interactions with the backbone and the side chains of the NLS. Hydrogen bonding and hydrophobic interactions between Kapα and the NLS backbone and side chains are likely to be the primary determinants of affinity, with specificity arising from electrostatic complementarity. In addition, the spacing between a defined number of binding sites acts as a molecular ruler that sets further constraints on target specificity.

The superhelical assembly of α helices in the repeated architecture of Kapα provides a larger accessible surface area for interactions than in more globular proteins of comparable molecular weight, and it is therefore ideally suited for protein–protein association. This principle has already been observed in leucine-rich and ankyrin-repeat proteins (Kobe and Deisenhofer, 1994) and is also likely to govern the binding of Kapα to the repetitive HEAT motifs of Kapβ (Andrade and Bork, 1995). It has been proposed that the arm repeats of Kapα and the HEAT repeats of Kapβ are related and that the transport factors Kapα and Kapβ are in fact members of the same protein superfamily (Malik et al., 1997). Evolutionary specialization of the nuclear targeting machinery presumably resulted in Kapα becoming an adaptor protein in nuclear import. Kapα has a range of substrate specificity, in that it recognizes a diversity of topogenic sequences that can all direct a protein to the nucleus. The structure described here explains the general requirements for classical monopartite and bipartite nuclear localization signals, revealing the determinants for simultaneous sequence conservation and variation.

Experimental Procedures

Protein Preparation

Full-length yeast karyopherin α (residues 1–542) was expressed and purified as previously described (Enenkel et al., 1995). The fragment of yeast karyopherin α spanning residues 88–530 used for the structure determination was subcloned into a pProEX-HTb expression vector (Life Sciences), preceded by six histidine residues and a TEV protease cleavage site. N-terminally His_6-tagged Kapα50 was overexpressed at 30°C in *Escherichia coli* strain BL21(DE3) following induction at an optical density of 0.6 (600 nm) with 0.5 mM IPTG for 5 hr in Luria broth. Cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris-HCl [pH 8.0], 100 mM NaCl, 10% glycerol) with 1 mM PMSF and 1 mg/ml aprotinin. Following cell lysis with a French pressure cell and separation of cell debris by high speed centrifugation, the soluble supernatant was loaded onto a Ni-NTA resin (Qiagen), washed with buffer B (20 mM Tris-HCl [pH 8.0], 500 mM NaCl, 10% glycerol), and eluted with an imidazole gradient. The peak fractions were further purified by anion exchange at pH 6.5, followed by dialysis in buffer A and cleavage of the hexahistidine tag at 4°C over 24 hr with TEV protease. Kapα50 was depleted of the protease and of the uncleaved His_6-Kapα50 by affinity and anion exchange chromatography at pH 8.0. Six liters of culture generally yielded 100 mg of purified Kapα50, which was concentrated to 20 mg/ml in buffer A and stored at −20°C. The protein is dimeric and monodisperse in solution at 1 mg/ml, as determined by dynamic light scattering performed with a Dyna-Pro-801 molecular size detector (Protein Solutions, Charlottesville, VA).

Crystallization and Data Collection

Crystals of Kapα50 were grown at 4°C in hanging drops containing equal volumes of the protein solution and of a reservoir solution that consisted of 18%–20% (w/v) polyethylene glycol (PEG) 4000, 200 mM sodium acetate, 100 mM HEPES (pH 7.0), and 20 mM cobaltous chloride ($CoCl_2$). Despite the presence of heavy precipitate, crystals typically grew to 0.3 × 0.3 × 0.4 mm^3 in 3 or 4 weeks. The space group is $P2_12_12_1$, with cell dimensions a = 158.0 Å, b = 74.1 Å, and c = 84.1 Å. The crystals contain two molecules per asymmetric unit, consistent with a solvent content of 50%.

Crystals were stabilized in 25% PEG 4000, 100 mM HEPES (pH 7.0), 200 mM sodium acetate, 15 mM $CoCl_2$, and 5% glycerol and flash-frozen in liquid propane after a quick transfer into a similar solution containing 20% glycerol. Crystallographic data were measured at the Cornell High Energy Synchrotron Source (CHESS) A1 beamline and at beamline X25 of the National Synchrotron Light

Source at Brookhaven National Laboratory. X-ray intensities were processed using the Denzo/HKL package (Otwinowski and Minor, 1997), and the CCP4 suite was used for subsequent calculations (CCP4, 1994). A summary of the data collection statistics is given in Table 2.

The diffraction pattern generated by the Kapα50 crystals exhibits significant anisotropy, with strong diffraction observed to Bragg spacing of 2.0 Å resolution or better along a* and c*. Along b*, however, the diffraction pattern becomes relatively weak beyond 2.4 Å.

Structure Determination

Phases were determined by the method of multiple isomorphous replacement using the program MLPHARE (CCP4, 1994). The final MIR phases had an overall figure of merit of 0.50 for data between 20 and 3.0 Å resolution and were improved by solvent flattening and 2-fold averaging (Table 2). A polyalanine model for 400 amino acid residues was traced with the graphics program O (Jones et al., 1991) using the backbone of the β-catenin molecule (Huber et al., 1997) as a guide. After correction for the anisotropy of the data, the experimental electron density map improved significantly and allowed unambiguous assignment of the amino acid sequence.

The model was refined using the maximum-likelihood target with CNS (Brünger et al., 1998). A random sample containing 5% of the total data between 2.2 and 30 Å resolution (with no σ cutoff) was excluded from the refinement to monitor the course of the refinement procedure. A bulk solvent correction and an anisotropic temperature factor correction (final tensor elements of B_{11} = 13.2 Å^2, B_{22} = −27.0 Å^2, B_{33} = 13.6 Å^2) were applied throughout the refinement. Strict noncrystallographic 2-fold symmetry was enforced and released only when the free R value had dropped to 29%. The refined model has an R_{free} = 27.5% and $R_{working}$ = 23.7%, with rmsd from ideality of 0.007 Å in bond lengths and 1.37° in bond angles. Stereochemical analysis using PROCHECK (CCP4, 1994) showed 96.0% of the residues lying in the most favored regions of the Ramachandran plot and none in the disallowed areas. The average temperature factor for all protein atoms is 45.0 Å^2.

The final R values of the model are somewhat higher than would be expected based on the quality of the final model. Although the $R_{working}$ increases sharply at about 2.4 Å, inclusion of all data to 2.2 Å resolution was beneficial in the interpretation of the electron density, and it was therefore decided not to truncate the data set to lower resolution.

The NLS-Kapα50 complex was formed by soaking the unliganded crystals for 48 hr in a stabilizing solution containing 0.8 mg/ml of a Ser-Pro-Lys-Lys-Lys-Arg-Lys-Val-Glu peptide. Data to 2.8 Å resolution were measured at CHESS, and the structure was refined in CNS using the structure of the unliganded protein as the initial model. The peptide could be fitted unambiguously in the electron density (Figure 7) also due to the nature of the donor–acceptor interactions between its main chain and the protein. The crystallographic $R_{working}$ is 25.6%, and the R_{free} is 30.7%. Each of the peptide binding sites is free of direct intermolecular packing interactions in the crystal lattice. However, there are crystal packing contacts between loops of Kapα that are near the peptide binding sites, and small changes in their conformation upon NLS binding might underlie the sensitivity of the Kapα crystals to soaking in solutions containing peptide.

Acknowledgments

We thank M. Rexach for the clone of yeast karyopherin α. Protein sequence determination and peptide synthesis was performed by the protein/DNA Technology Center of the Rockefeller University. We are grateful to F. Gharahdaghi and B. Chait for assistance with mass spectrometry. We thank L. Berman, R. Sweet, the staff of National Synchrotron Light Source, and the staff of Cornell High Energy Synchrotron Source for assistance with synchrotron data collection. We thank S.K. Burley, J. Bonanno, A. Boriack-Sjodin, X. Chen, M. Huse, D. Jeruzalmi, S. Soisson, H. Yamaguchi, Y. Zhang, J. Marcotrigiano, J. Gulbis, and J. Cabrao for advice and help. E. C. was supported by an EMBO fellowship and a Human Frontier Science Program fellowship.

Received May 20, 1998; revised June 25, 1998.

References

Adam, E.J., and Adam, S.A. (1994). Identification of cytosolic factors required for nuclear-location sequence-mediated binding to the nuclear envelope. J. Cell Biol. *125*, 547–555.

Andrade, M.A., and Bork, P. (1995). HEAT repeats in the Huntington's disease protein. Nature Genet. *11*, 115–116.

Blobel, G. (1980). Intracellular protein topogenesis. Proc. Natl. Acad. Sci. USA *77*, 1496–1500.

Brünger, A.T., Adams, P.D., Clore, G.M., Gros, P., Grosse-Kuntsleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J. et al. (1998). Crystallographic and NMR system: a new software system for macromolecular structure determination. Acta Crystallogr. D, in press.

Burley, S.K., and Petsko, G.A. (1988). Weakly polar interactions in proteins. Adv. Protein Chem. *39*, 125–189.

Carson, M. (1991). Ribbons 2.0. Acta Crystallogr. *24*, 958–961.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. *D50*, 760–763.

Chida, K., and Vogt, P.K. (1992). Nuclear translocation of viral Jun but not cellular Jun is cell cycle dependent. Proc. Natl. Acad. Sci. USA *89*, 4290–4294.

Colledge, W.H., Richardson, W.D., Edge, M.D., and Smith, A.E. (1986). Extensive mutagenesis of the nuclear location signal of simian virus 40 large-T antigen. Mol. Cell. Biol. *6*, 4136–4139.

Cortes, P., Ye, Z.-S., and Baltimore, D. (1994). RAG-1 interacts with the repeated amino acid motif of the human homologue of the yeast protein SRP1. Proc. Natl. Acad. Sci. USA *91*, 7633–7637.

Dang, C.V., and Lee, W.M.F. (1988). Identification of the human c-*myc* protein nuclear translocation signal. Mol. Cell. Biol. *8*, 4048–4054.

Dingwall, C., and Laskey, R.A. (1991). Nuclear targeting sequences: a consensus? Trends Biol. Sci. *16*, 178–181.

Dingwall, C., Robbins, J., Dilworth, S.M., Roberts, B., and Richardson, W.D. (1988). The nucleoplasmin nuclear location sequence is larger and more complex than that of SV40 large T antigen. J. Cell Biol. *107*, 841–849.

Efthymiadis, A., Shao, H., Hubner, S., and Jans, D.A. (1997). Kinetic characterization of the human retinoblastoma protein bipartite nuclear localization sequence (NLS) in vivo and in vitro: a comparison with the SV40 large T antigen NLS. J. Biol. Chem. *272*, 22134–22139.

Enenkel, C., Blobel, G., and Rexach, M. (1995). Identification of a yeast karyopherin heterodimer that targets import substrate to mammalian nuclear pore complexes. J. Biol. Chem. *270*, 16499–16502.

Görlich, D., and Mattaj, I.W. (1996). Nucleocytoplasmic transport. Science *271*, 1513–1518.

Görlich, D., Prehn, S., Laskey, R.A., and Hartmann, E. (1994). Isolation of a protein that is essential for the first step of nuclear protein import. Cell *79*, 767–778.

Görlich, D., Kostka, S., Krarft, R., Dingwall, C., Laskey, R.A., Hartmann, E., and Prehn, S. (1995). Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. Curr. Biol. *5*, 383–392.

Görlich, D., Henklein, P., Laskey, R.A., and Hartmann, E. (1996). A 41 amino acid motif in importin-alpha confers binding to importin-beta and hence transit into the nucleus. EMBO J. *15*, 1810–1817.

Hendsch, Z.S., and Tidor, B. (1994). Do salt bridges stabilize proteins?: a continuum electrostatic analysis. Protein Sci. *3*, 211–226.

Huber, A.H., Nelson, W.J., and Weis, W.I. (1997). Three-dimensional structure of the armadillo repeat region of β-catenin. Cell *90*, 871–882.

Imamoto, N., Tachibana, Y., Matsubae, M., and Yoneda, Y. (1995). A karyophilic protein forms a stable complex with cytoplasmic components prior to nuclear pore binding. J. Biol. Chem. *270*, 8559–8565.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110–119.

Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino acid sequence able to specify nuclear location. Cell *39*, 499–509.

Kobe, B., and Deisenhofer, J. (1994). The leucine-rich repeat: a versatile binding motif. Trends Biochem. Sci. *19*, 415–421.

Lanford, R.E., and Butel, J.S. (1984). Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. Cell *37*, 801–813.

Loeb, J.D., Schlenstedt, G., Pellman, D., Kornitzer, D., Silver, P.A., and Fink, G.R. (1995). The yeast nuclear import receptor is required for mitosis. Proc. Natl. Acad. Sci. USA *92*, 7647–7651.

Makkerh, J.P.S., Dingwall, C., and Laskey, R.A. (1996). Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. Curr. Biol. *6*, 1025–1027.

Malik, H.S., Eickbush, T.H., and Goldfarb, D.S. (1997). Evolutionary specialization of the nuclear targeting apparatus. Proc. Natl. Acad. Sci. USA *94*, 13738–13742.

Moreland, R.B., Langevin, G.L., Singer, R.H., Garcea, R.L., and Hereford, L.M. (1987). Amino acid sequences that determine the nuclear localization of yeast histone H2B. Mol. Cell. Biol. *7*, 4048–4057.

Moroianu, J., Blobel, G., and Radu, A. (1995a). Previously identified protein of uncertain function is karyopherin α and together with karyopherin β docks import substrate at nuclear pore complexes. Proc. Natl. Acad. Sci. USA *92*, 2008–2011.

Moroianu, J., Hijikata, M., Blobel, G., and Radu, A. (1995b). Mammalian karyopherin $\alpha 1\beta$ and $\alpha 2\beta$ heterodimers: $\alpha 1$ or $\alpha 2$ subunit binds nuclear localization signal and β subunit interacts with peptide repeat-containing nucleoporins. Proc. Natl. Acad. Sci. USA *92*, 6532–6536.

Moroianu, J., Blobel, G., and Radu, A. (1996). The binding site of karyopherin alpha for karyopherin beta overlaps with a nuclear localization sequence. Proc. Natl. Acad. Sci. USA *93*, 6572–6576.

Naismith, J.H., Brandhuber, B.J., Devine, T.Q., and Sprang, S.R. (1996). Seeing double: crystal structures of the type I TNF receptor. J. Mol. Recognit. *9*, 113–117.

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

Nigg, E.A. (1997). Nucleocytoplasmic transport: signals, mechanisms and regulation. Nature *386*, 779–787.

Ohno, M., Fornerod, M., and Mattaj, I.W. (1998). Nucleocytoplasmic transport: the last 200 nanometers. Cell *92*, 327–336.

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Peifer, M., Berg, S., and Reynolds, A.B. (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. Cell *76*, 789–791.

Pemberton, L.F., Blobel, G., and Rosenblum, J.S. (1998). Transport routes through the nuclear pore complex. Curr. Opin. Cell Biol. *10*, 392–399.

Radu, A., Moore, M.S., and Blobel, G. (1995). The peptide repeat domain of nucleoporin 98 functions as a docking site in transport across the nuclear pore complex. Cell *81*, 215–222.

Rexach, M., and Blobel, G. (1995). Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. Cell *83*, 683–692.

Robbins, J., Dilworth, S.M., Laskey, R.A., and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell *64*, 615–623.

Sekimoto, T., Imamoto, N., Nakajima, K., Hirano, T., and Yoneda, Y. (1997). Extracellular signal-dependent nuclear import of Stat1 is mediated by nuclear pore-targeting complex formation with NPI-1, but not Rch1. EMBO J. *16*, 7067–7077.

Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L., and Wiley, D.C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature *368*, 215–221.

Weis, K., Mattaj, I.W., and Lamond, A.I. (1995). Identification of hSRP1α as a functional receptor for nuclear localization sequences. Science *268*, 1049–1053.

Weis, K., Ryder, U., and Lamond, A.I. (1996). The conserved amino-terminal domain of hSRP1 alpha is essential for nuclear protein import. EMBO J. *15*, 7120–7128.

Wozniak, R.W., Rout, M.P., and Aitchison, J.D. (1998). Karyopherin and kissing cousins. Trends Cell Biol., in press.

Yang, Q., Rout, M.P., and Akey, C.W. (1998). Three dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. Mol. Cell *1*, 223–234.

Yano, R.M., Oakes, M., Yamaghishi, M.M., Dodd, J.A., and Nomura, M. (1992). Cloning and characterization of SRP1, a suppressor of temperature-sensitive RNA polymerase I mutations in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *12*, 5640–5651.

Brookhaven Protein Data Bank Accession Numbers

The accession numbers for the structures described in this manuscript are 1bk5 and 1bk6. Coordinates are also available on the web site www.rockefeller.edu/kuriyan.