Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin α

Elena Conti[1,2*] and John Kuriyan[1]

Background: Karyopherin α (importin α) is an adaptor molecule that recognizes proteins containing nuclear localization signals (NLSs). The prototypical NLS that is able to bind to karyopherin α is that of the SV40 T antigen, and consists of a short positively charged sequence motif. Distinct classes of NLSs (monopartite and bipartite) have been identified that are only partly conserved with respect to one another but are nevertheless recognized by the same receptor.

Results: We report the crystal structures of two peptide complexes of yeast karyopherin α (Kapα): one with a human c-*myc* NLS peptide, determined at 2.1 Å resolution, and one with a *Xenopus* nucleoplasmin NLS peptide, determined at 2.4 Å resolution. Analysis of these structures reveals the determinants of specificity for the binding of a relatively hydrophobic monopartite NLS and of a bipartite NLS peptide. The peptides bind Kapα in its extended surface groove, which presents a modular array of tandem binding pockets for amino acid residues.

Conclusions: Monopartite and bipartite NLSs bind to a different number of amino acid binding pockets and make different interactions within them. The relatively hydrophobic monopartite c-*myc* NLS binds extensively at a few binding pockets in a similar manner to that of the SV40 T antigen NLS. In contrast, the bipartite nucleoplasmin NLS engages the whole array of pockets with individually more limited but overall more abundant interactions, which

Addresses: [1]Laboratory of Molecular Biophysics, Howard Hughes Medical Institute, The Rockefeller University, New York 10021, USA and [2]European Molecular Biology Laboratory, Heidelberg D-69117, Germany.

*Corresponding author.
E-mail: conti@embl-heidelberg.de

Key words: bipartite, crystal structure, karyopherin α, nuclear import, nuclear localization signal

Received: **10 January 2000**
Revisions requested: **26 January 2000**
Revisions received: **2 February 2000**
Accepted: **4 February**

Published: **29 February 2000**

Structure 2000, **8**:329–338

0969-2126/00/$ – see front matter

Introduction

Cellular activities in the nucleus and the cytosol of eukaryotic cells are coordinated by a continuous exchange of macromolecules across the nuclear envelope. Nuclear proteins and RNAs are shuttled from their site of synthesis to their final destination by an active and selective mechanism that is mediated by signal-dependent binding to specific and saturable receptors [1–3].

The first nuclear import pathway to be discovered, referred to as the 'classical' pathway, relies on the recognition by transport proteins of a nuclear localization signal (NLS) presented by the nuclear protein to be translocated. The NLS receptor is a cytosolic αβ heterodimer known as karyopherin or importin [4–8]. Karyopherin α (Kapα) is the saturable component of the receptor and binds the NLS-containing cargo in an energy-independent manner [9–11]. Kapα is the carrier component of the receptor, which docks the complex to the nuclear pore and translocates it to the nucleus [7,8]. Upon reaching the nucleoplasmic side, the two karyopherin subunits are dissociated via the direct interaction of the β subunit with the small GTPase Ran [12,13]. As a result, the α subunit and the NLS-containing nuclear protein are released to the nucleoplasm, whereas the β subunit accumulates at the nuclear envelope. After releasing its NLS cargo, the α subunit is recycled back to the cytoplasm via an active and specific nuclear export pathway that involves a β homologue receptor [14].

Many nuclear proteins are imported via the classical import pathway. A number of similar but independent pathways, however, have been characterized for the import of ribosomal and mRNA-binding proteins and for the export of Kapα and RNAs (see reviews [1–3]). These transport pathways mostly rely on the direct interaction of the target macromolecule with a β homologue receptor, which mediates both saturable cargo binding and translocation through the nuclear pore complex. In contrast, the classical import pathway requires Kapα as an adaptor protein to bind NLSs.

NLSs are sequence motifs that are sufficient for targeting their respective proteins to the nucleus. Proteins that are normally restricted to the cytoplasm can be imported to the nucleus if fused to an NLS motif [15]. NLSs are not cleaved after entry into the nucleus and can be found anywhere within a protein sequence. NLS motifs lack a

stringent consensus sequence, although they are generally short stretches of amino acids that contain a high proportion of positively charged residues [16].

The importance of basic residues in nuclear import was first established with the extensive characterization of the NLS in the SV40 T antigen polypeptide [17,18]. The SV40 T antigen NLS consists of seven residues, five of which are basic (^{126}PKKKRKV^{132}, in single-letter amino acid code). Despite the apparently non-descript repetition of positively charged amino acids in the signal, point mutations identified one particular lysine residue (Lys128) as essential, and the adjacent residues were found to be either less critical or in one case even dispensable [19]. These differences in the relative importance of individual basic residues in the motif suggested that the receptor imposes particular structural requirements on the NLS that cannot be satisfied by the simple juxtaposition of positive charges.

As more NLSs were identified, it became apparent that those that are monopartite and defined by the presence of a short single stretch of contiguous positively charged residues are in fact an exception. More than 50% of nuclear proteins contain a sequence that fits a longer bipartite NLS motif [16]. Bipartite NLSs are approximately 18 amino acids in length with two clusters of basic residues separated by a linker, as italicised in the nucleoplasmin sequence ^{155}*KR*PAATKKAGQA*KKKK*LD^{172} [20,21]. The larger downstream basic cluster resembles a monopartite NLS, but is not sufficient by itself to target a protein to the nucleus. It requires a smaller upstream cluster separated by a linker of 10–12 residues with variable sequence. Strikingly in this context, the c-*myc* proto-oncogene is directed to the nucleus by a short peptide (^{320}PAAKRVKLD^{328}) that contains even fewer positively charged residues than the typical downstream cluster of a bipartite NLS [22]. In contrast to the SV40 NLS, the c-*myc* NLS requires the presence of hydrophobic residues at certain positions [23].

The determinants for the specific recognition of a canonical monopartite NLS were elucidated when the crystal structure of the NLS-binding domain of *Saccharomyces cerevisiae* karyopherin α (Kapα50) was determined in a complex with the nuclear targeting signal of the SV40 T antigen [24]. Kapα50 consists of the 50 kDa C-terminal portion of the full-length molecule. It lacks the importin β-binding (IBB) N-terminal domain [25], which contains an autoinhibitory segment [26]. Kapα50 is an elongated molecule consisting of ten helical repeats known as armadillo (ARM) motifs. Each ARM motif has an approximate triangular cross-section and is characterized by three α helices. Two α helices (H1 and H2) form the convex outer surface of the molecule, and the third (H3) forms its concave portion. The concave surface of Kapα is shaped into a shallow and extended groove, which is lined by a ladder of conserved tryptophan and asparagine residues and surrounded by acidic amino acids. This arrangement results in an array of binding sites that combine hydrophobic and electrostatic elements to recognize the positively charged residues of the SV40 NLS specifically.

Kapα is able to recognize not only the monopartite SV40 T antigen NLS, but also the more hydrophobic c-*myc* NLS as well as the longer bipartite nucleoplasmin NLS. Despite the diversity in its targets, however, Kapα is not indiscriminate because many sequences that are only slightly different from active NLS motifs are not recognized. To address the structural basis for the versatile specificity of Kapα we have determined the crystal structures of the yeast Kapα50 in complex with two distinct NLS-containing peptides: a human c-*myc* NLS peptide in which hydrophobic residues play an important role, and a *Xenopus* nucleoplasmin peptide that contains a bipartite NLS.

Results and discussion

Structure determination

Initial attempts at the co-crystallization of Kapα50 with peptides corresponding to the sequence of the c-*myc* NLS or to that of the nucleoplasmin NLS were unsuccessful. Furthermore, in contrast to the short SV40 T antigen NLS, neither of the peptides used in this study could be soaked productively into the native Kapα50 crystals [24]. Inspection of the Kapα50 structure suggested that the inability to soak most peptides into this crystal form was perhaps because of the substrate-binding site of Kapα50 being partially occluded by the formation of a dimeric structure. To prevent the dimerization of Kapα50 we engineered a point mutation (Tyr→Asp at position 397, Y397D). In the native crystal form, a central element of the intermolecular interface is a hydrophobic interaction between the sidechains of the conserved Tyr397 residue and its symmetry mate at the local twofold of the Kapα50 homodimer. Mutation of residue 397 from tyrosine to aspartic acid is expected to destabilize the homodimer. Although residue 397 is buried in the crystallographic dimer, it would be exposed to solvent in the monomeric form and the mutation is therefore unlikely to alter the mainchain conformation to a significant extent.

The Y397D Kapα50 mutant was expressed better in bacterial cells than the wild-type protein was and showed improved solubility properties in crystallization experiments. Co-crystals with a c-*myc* NLS and a nucleoplasmin NLS peptide were obtained using the vapor diffusion method with polyethylene glycol as the precipitating agent, with different additives and pH conditions. The c-*myc* NLS–Kapα50 crystals are in space group $P2_1$ and diffract X-rays to 2.1 Å resolution, and the nucleoplasmin NLS–Kapα50 crystals are in space group C2 and diffract to 2.4 Å resolution. Both structures were determined using molecular replacement, using the native Kapα50 model. The models are currently refined to free and

working R values of 27.7% and 24.6% respectively for the c-*myc* complex, and 26.5% and 24.1% respectively for the nucleoplasmin complex.

Overall architecture and conformational flexibility

The overall structure of the Y397D mutant of yeast Kapα50 is quite similar to that of the wild-type protein and of the mouse homologue reported previously [24,26]. The molecule presents an extended surface groove lined by a set of conserved residues, notably a set of solvent-exposed tryptophan and asparagine residues. Each Trp–Asn pair is contributed by one ARM motif. The tryptophan residues occupy a conserved position in the H3 helices of each ARM motif and the asparagine residues are located one helical turn downstream from the tryptophan residues. This arrangement creates an array of Trp–Asn pairs positioned at regular intervals along the surface groove. The regularity of the Trp–Asn ladder is interrupted in the fifth and sixth ARM motifs by the presence of an arginine and a tyrosine residue at the corresponding positions.

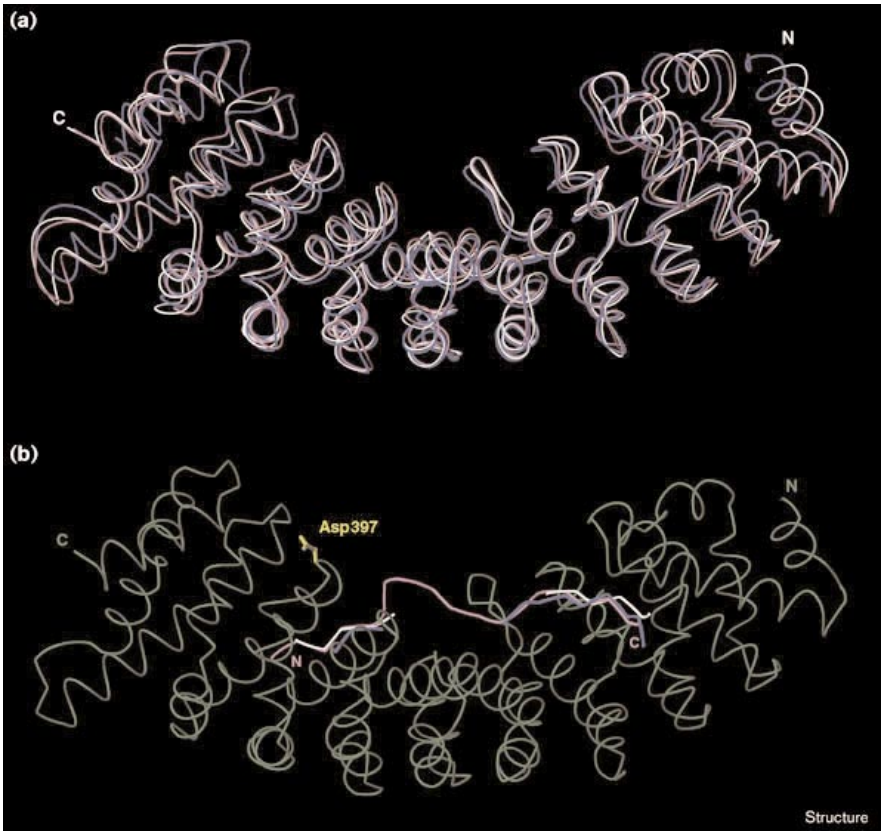
There is a small but significant change in the overall shape of the Kapα50 molecule in complex with the c-*myc* or nucleoplasmin NLS peptides when compared with that of the SV40 T antigen NLS complex (Figure 1a). The overall root mean square deviations (rmsd) in Cα positions are 1.6 Å and 1.1 Å, respectively. In these global comparisons, the central repeats (ARM 4–8) superpose closely with an rmsd of about 0.7 Å. In contrast, the three N-terminal and the two C-terminal ARM repeats deviate with an average rotation of 6° and 5°, respectively. A larger global deviation of 3.1 Å is obtained from the comparison of the c-*myc* and nucleoplasmin structures; a rotation of 6.4° in this case relates the N-terminal ARM repeats, and ARM 9 and 10 deviate by 9.8°.

The overall difference in the structure of Kapα50 between the c-*myc* and nucleoplasmin complexes is larger than that observed when comparing the yeast and mouse Kapα proteins in the SV40 complex and in the autoinhibited form, respectively [26]. The conformational differences appear to be caused by peptide binding more than by crystal packing. Such flexibility is the product of small additive changes that propagate along the polypeptide chain, and results in a more or less open structure.

General features of peptide binding

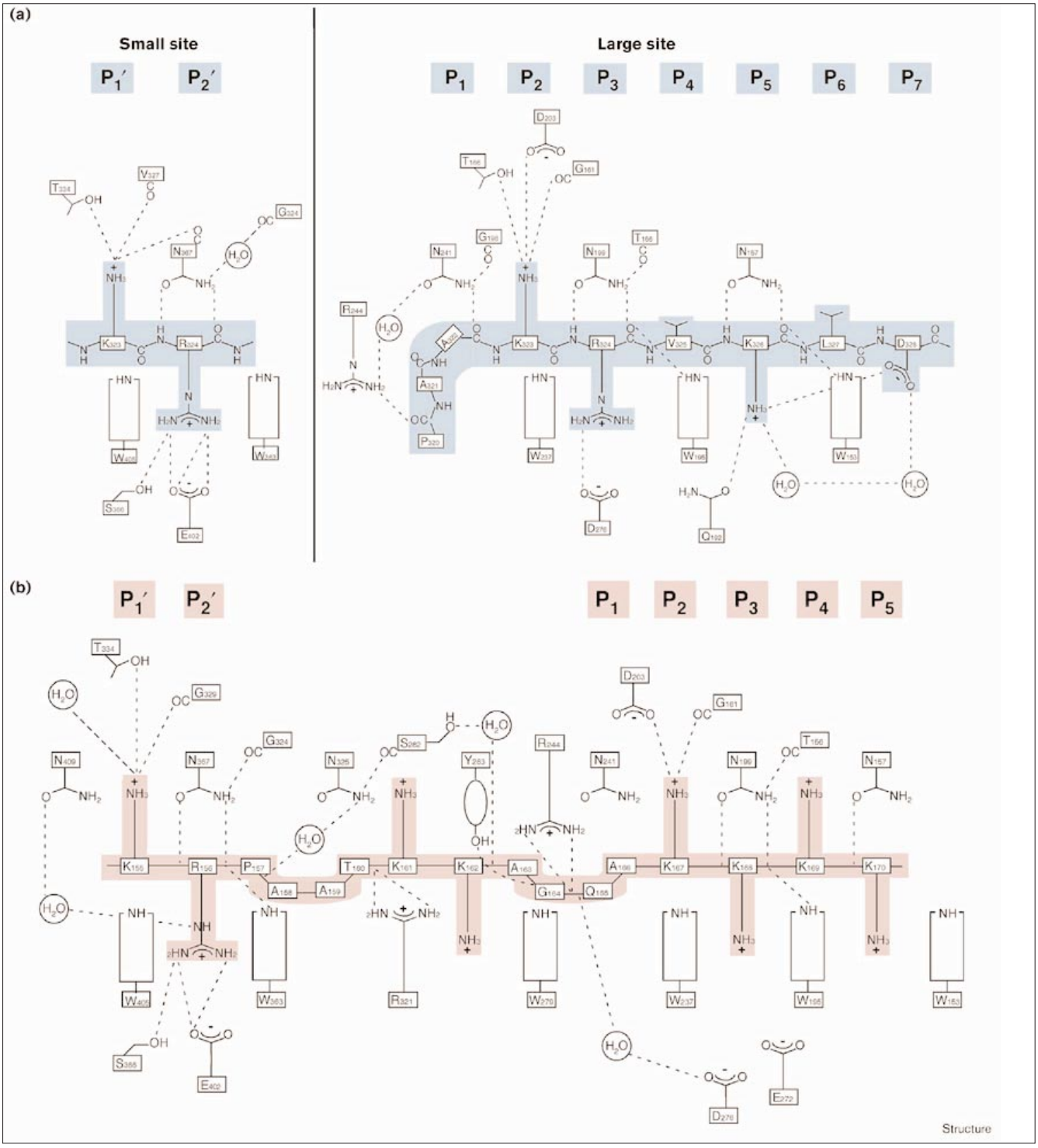
The c-*myc* and nucleoplasmin NLS peptides bind along the spine of the surface groove in an almost fully extended conformation (Figure 1b). Their mainchains run in the opposite direction to that of the protein, with the NLS N termini bound to the C-terminal region of Kapα50 and their C termini to the N-terminal region.

Figure 1



Structure of yeast Kapα50. **(a)** Comparison of the overall structure of yeast Kapα50 in the NLS complexes. The Cα backbones of Kapα in the c-*myc* complex (shown in blue) and in the nucleoplasmin NLS complex (shown in purple) have been globally superposed to that of the SV40 T antigen NLS complex (shown in white). Although the overall structure of Kapα in the complexes is the same, some flexibility of the ARM repeats at the N- and C-terminal portion of the molecule is observed upon binding of different NLSs. The figure was created using the program RIBBONS [46]. **(b)** Comparison of the NLS peptide-binding sites. The c-*myc* NLS peptide (blue) and the nucleoplasmin NLS peptide (purple) are shown after a global protein superposition to Kapα (green) in a complex with the SV40 T antigen NLS (white). The peptide backbones closely superpose at the large and small binding sites at the N and C termini of the protein, respectively. The mutated residue (Y397D) used to obtain the c-*myc* and nucleoplasmin co-crystals of Kapα is shown in yellow.

Figure 2



Interactions of distinct NLSs with Kapα. The NLS peptides (shown in colour) nestle between the conserved Trp–Asn pairs of residues that line the surface of Kapα. Each pair is provided by a single ARM repeat, shaping the regularly spaced P and P′ specificity pockets at the large and small binding sites, respectively. The polar and electrostatic interactions between conserved protein and NLS residues are indicated with dotted lines. **(a)** Schematic diagram of the interactions between the human c-*myc* NLS peptide (highlighted in blue) and yeast Kapα. The monopartite NLS binds at the large (functional) and small sites with mainchain and sidechain interactions. **(b)** Schematic diagram of the interactions between the *Xenopus* nucleoplasmin NLS peptide (highlighted in pink) and protein residues. The bipartite NLS binds along the whole surface groove with mainchain and sidechain interactions at the small site, and with mainly backbone interactions of the linker and the downstream cluster.

The point mutation engineered in Kapα50 appears to not influence peptide binding. The aspartate sidechain at position 397 (which replaces the tyrosine found in the wild-type protein) points into solvent. It is 16 Å from the nearest atom of the NLS peptide and is unlikely to alter the affinity of NLS binding significantly.

Two copies of the monopartite NLS of human c-*myc* (residues 320–328) bind yeast Kapα50 at two sites, as was the case for the single SV40 T antigen targeting signal [24]. A large site is present in the N-terminal half of the molecule and is characterized by three Trp–Asn pairs (at ARM 2–4), whereas two such pairs line a small site in the C-terminal half (at ARM 7 and 8). The bipartite NLS (residues 155–170) of *Xenopus* nucleoplasmin spans the whole surface groove between the small and the large sites defined for monopartite NLSs as indeed predicted [24,27], and it additionally interacts with the central fifth and sixth ARM repeats.

The 'backbones' of the monopartite and bipartite peptides are closely superimposable at the large and small sites. The deviation in the Cα positions of the nucleoplasmin residues 165–170 and the c-*myc* residues 320–326 is 1.1 Å. At the small site, the superposition of three α carbon atoms of the bipartite NLS (155–157) with the corresponding ones of c-*myc* (323–325) results in an rmsd of 0.7 Å. The detailed set of polar and hydrophobic interactions exploited by the peptides at the binding pockets is, however, not identical (Figure 2), and is tailored to the requirements of their distinctive sequence. For consistency with the notation used previously for the SV40 T antigen NLS recognition, the binding pockets will be referred to as P and P′ at the large and small site, respectively.

c-*myc* NLS recognition

Well-defined high-resolution electron density (Figure 3a) marks the recognition of the human c-*myc* NLS (^{320}PAAKRVKLD^{328}) at the seven binding pockets that shape the large site of Kapα50 (Figure 3b). The central portion of the c-*myc* NLS (^{323}KRVKL^{327}) binds in an extended conformation from the P_2 to the P_6 pocket (Figure 2a). The tryptophan and asparagine sidechains of the third and second ARM motifs (W195–N199, W153–N157 pairs) bind the peptide mainchain by fully satisfying its hydrogen-bonding potential at the P_3 and P_5 positions. The third Trp–Asn pair at the large site (W237–N241) interacts only partially with the NLS mainchain, which deviates from a β-strand conformation with a sharp bend at residue Ala322. The PAA tripeptide (residues 320–322) is anchored by a hydrogen-bonding interaction of its N-terminal proline residue to a conserved arginine sidechain.

The sidechains of the three positively charged residues of the NLS protrude into the P_2, P_3 and P_5 pockets and are involved in interactions that are similar to those observed in the binding of the SV40 T antigen NLS [24]. In the case of the c-*myc* NLS, the basic residue at P_5 forms an intramolecular salt bridge with the Asp328 residue of the peptide at P_7 (Figure 2a). The hydrophobic residues at P_4 and P_6 fit into apolar areas on the surface of the protein, which are not immediately surrounded by acidic residues. At P_1, the sidechain of Ala322 points in the same direction as Lys323, as is the case for the autoinhibitory segment of mouse Kapα but not for the SV40 NLS complex, in which a lysine residue points in the opposite direction and occupies the P_1 pocket. It is unclear whether the differences observed at the P_1 site in the SV40 structure are genuine or altered by the presence of a crystallographic dimer interface near this region.

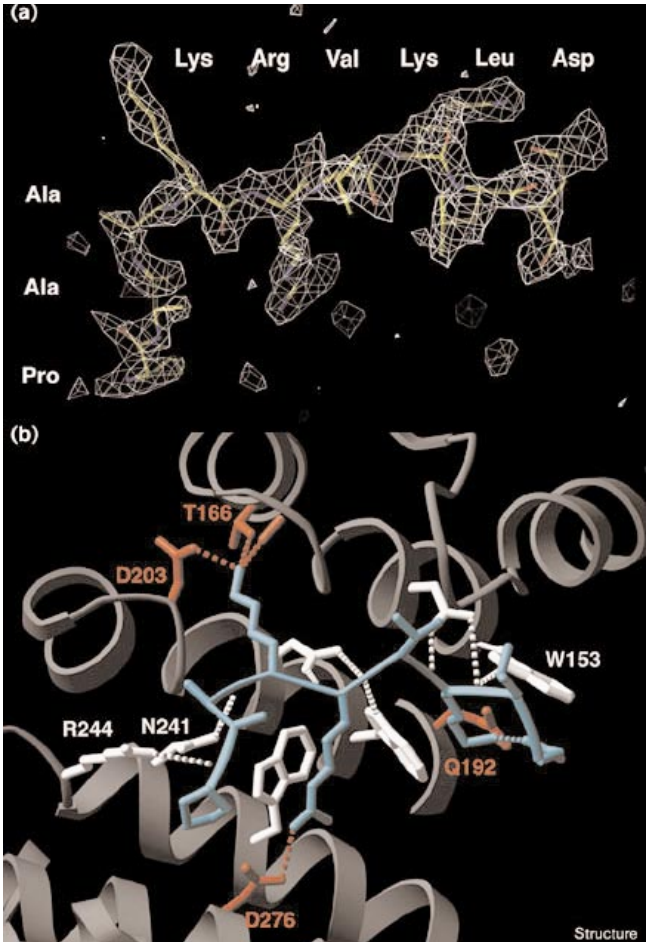
A smaller additional binding site for the NLS is present in the C-terminal half of Kapα at the Trp–Asn pairs of ARM 7 and ARM 8 (W363–N367, W405–N409 pairs) where two contiguous positively charged residues occupy the P_1' and P_2' pockets (Figure 2a). Although the binding of the c-*myc* peptide to this small site might not be physiologically relevant, the site reveals a set of polar and hydrophobic interactions that are similar to those observed earlier in the case of the SV40 T antigen NLS, which had led us to speculate the importance of the dual binding sites in bipartite NLS binding [24,27].

Nucleoplasmin NLS recognition

The bipartite NLS (^{154}VKRPAATKKAGQAKKKKLD^{172}) of *Xenopus* nucleoplasmin binds along the surface groove of Kapα50 with the upstream basic cluster (residues 155–156) at the small site and the downstream basic cluster (residues 167–170) at the large site. The electron density and the monotonic increase in B factors from the N terminus to the C terminus of the NLS suggest more prominent binding of the upstream cluster and of the linker rather than of the larger downstream cluster (Figure 4a). The peptide is recognized in a mostly extended conformation with two short turns, one of them (^{157}PAAT^{160} residues) following the smaller cluster and the other (^{163}AGQA^{166} residues) preceding the larger one (Figure 4b). Despite the fact that the peptide used was longer than the characterized NLS motif, only the latter is ordered in the structure.

The recognition of the KKKK downstream cluster of nucleoplasmin at the large site is mediated by fewer interactions than those established in the same area by the c-*myc* or SV40 T antigen monopartite NLSs (Figure 2b). The sidechains at P_2 and P_3 are well defined in the electron density, and exploit some but not all of the binding properties at the corresponding pockets. Conversely, poor electron density and no sidechain interactions are detected at P_4 and P_5. The C-terminal ^{171}LD^{172} segment is disordered, in contrast to the structure of the c-*myc* complex in which the same sequence at the same position is well ordered and participates in contacts at the P_6 and P_7 pockets. Similarly, the detailed interactions engaging the KR upstream cluster of nucleoplasmin at the small site are

Figure 3



c-*myc* NLS recognition. **(a)** The refined model of the c-*myc* NLS peptide bound at the large site on the surface of the protein is shown superposed to 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. The F_o–F_c electron-density map (white) is at 2.1 Å resolution and contoured at 3σ. **(b)** Ribbon representation of the nine-residue long c-*myc* NLS peptide (shown in blue) binding at the large site of Kapα (shown in grey). Key protein residues are shown in white and hydrogen-bonding interactions engaging the peptide backbone are indicated with white dotted lines. Intermolecular polar and electrostatic interactions between the NLS sidechains and protein residues (in red) are shown with red dotted lines, and the intramolecular salt bridge is indicated by a blue dotted line. The orientation of the molecule corresponds to the schematic representation of Figure 2a.

not identical to those involving the same amino acids in the equivalent pockets in the c-*myc* NLS complex.

The ten linker amino acids separating the two basic clusters of the nucleoplasmin NLS make extensive mainchain interactions with the sidechains of conserved residues which line a conspicuous part of the Kapα surface groove. The short turn following the upstream cluster and that preceding the downstream cluster are engaged in hydrogen-bonding interactions with Arg321 and Tyr283 in particular. These two conserved residues interrupt the regularity of the Trp–Asn array at the fifth and sixth ARM repeats (W279–Y283, R321–N325 pairs) and their function in nuclear import had, so far, been unclear. The nucleoplasmin GQA sequence occupies the P_1 pocket in a conformation similar to that adopted by the c-*myc* PAA tripeptide. In contrast to the extensive mainchain interactions, the sidechain interactions of the linker amino acids are scarce and limited to the Lys161 residue, possibly in a species-dependent manner.

NLS structural determinants

The crystal structure of the c-*myc* complex shows many of the central features that had been described previously for the highly basic SV40 T antigen monopartite NLS [24]. Most probably the physiological recognition occurs at the large site, where the c-*myc* peptide binds with all of its nine amino acids, whereas the simultaneous binding of a partial peptide sequence at the small site is probably because of the high concentration of NLS peptide present in the co-crystallization experiments.

One lysine residue (323) appears to play a crucial role in the nuclear localization signal of the c-*myc* NLS; it is engaged in extensive interactions at the P2 site. The essential PAA segment occupies the P_1 pocket, taking full advantage of its hydrophobic character and curved shape. The LD dipeptide that had been shown to be important in mutagenesis studies [23] stabilizes the P_5 residue by neutralizing its charge intramolecularly.

The crystal structure of the nucleoplasmin complex bears out all the characteristics of bipartite NLSs predicted by Dingwall and Laskey several years ago [16]. A large surface area with the benefit of multiple interactions mediates nucleoplasmin NLS recognition, providing a rationale for the absence of an absolute requirement for any single amino acid in the sequence in contrast to the monopartite NLSs of the SV40 T antigen and c-*myc*. Point-mutation experiments identified the lysine residues 155–167 as the most important, consistent with the more extensive set of interactions observed at the P_1', P_2' and P_2 pockets. Combination of any two such mutations abolished nuclear targeting [21]: the two clusters function in an interdependent manner by binding simultaneously at two sites on the same molecule. The correct orientation of the clusters is ensured by hydrogen bonds involving the peptide backbone at the small and large sites, and their correct positioning is made possible when a linker of at least ten residues is present to cover the physical separation between the two sites.

The paucity of sidechain interactions between the linker and the protein is in agreement with the lack of precise amino acid requirements in the intervening sequence, which can be replaced by a polyalanine segment without affecting nuclear targeting [21]. The recurrence of proline and glycine residues, however, appears to facilitate the nestling of the

peptide along the curvature of the Kapα surface and around the sidechains of Arg244, Arg321 and Tyr283, thus maintaining the wealth of mainchain interactions with the linker. In linkers of 11 or 12 amino acids [28], the preference for residues that are stereochemically compatible with sharp turns might be relaxed, but more stringent requirements on the correct positioning of the P_2 residue might be necessary.

Detailed modelling of longer bipartite NLSs such as the signal of the SWI5 transcription factor [29] is not straightforward, but general features can nevertheless be inferred. For example, electrostatic effects and steric hindrance (between the Trp279 residue of Kapα and the phosphoserine 646 in the bipartite NLS linker) are likely to disrupt the interaction with Kapα when the SWI5 transcription factor is phosphorylated, resulting in its cytoplasmic retention [29]. As the linker length increases, the cooperative binding of the clusters at the two sites would be more hampered than facilitated; linkers longer than 12 amino acids are in fact not common [16].

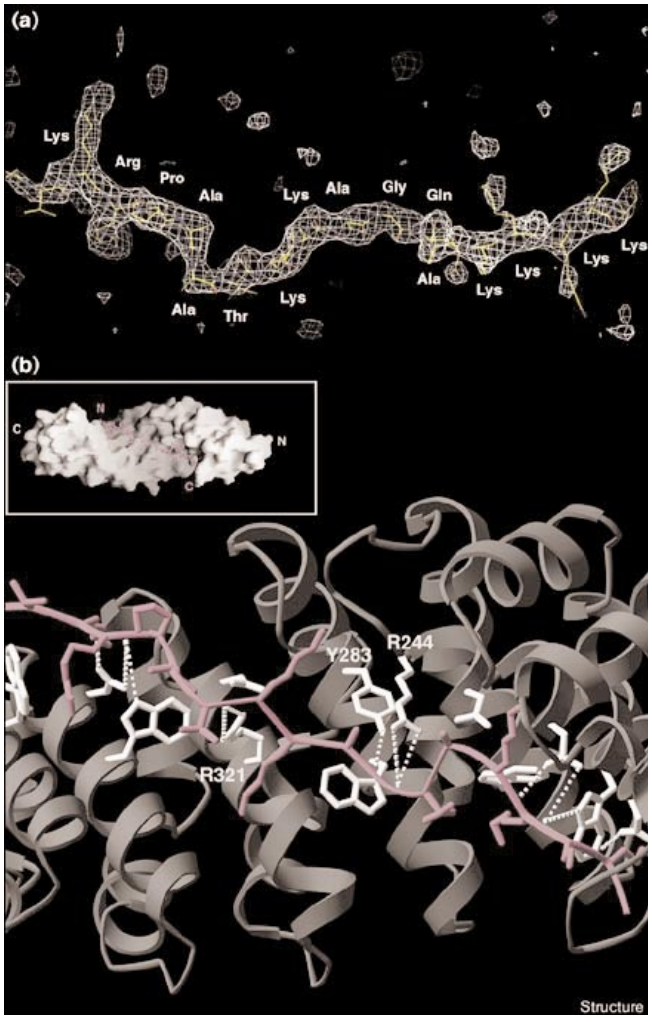
NLS sequence requirements

In the case of classical monopartite NLSs, a K–K/R–X–K/R (where X is any residue) consensus sequence appears to be a necessary minimal element [30]. The combination of hydrophobic and electrostatic properties at the P_2, P_3 and P_5 pockets of the large site restrains productive binding to residues with long aliphatic sidechains and a positive charge at the tip. The particular shape and chemistry of the P_2 pocket constrains the binding to lysine. Recognition of these three key positively charged residues is the major specificity determinant not only for the very basic SV40 NLS, but also for the more hydrophobic c-*myc* NLS.

The consensus sequence is necessary but not sufficient. Whereas in the case of the SV40 T antigen NLS the consensus is flanked by positively charged residues making contacts at the adjacent pockets, in the case of the c-*myc* NLS it is flanked by hydrophobic and negatively charged residues. The adjacent pockets pose a less-strict requirement on the incoming NLS residue, although a small hydrophobic residue at P_6 and a proline nestling at the P_1 pocket appear to be favoured. These additional interactions at the large site, however, are not able to rescue a mutation of the critical lysine residue at the P_2 pocket [17,23].

In the case of bipartite NLSs, no individual critical requirement in the sequence of the positively charged clusters appears to be necessary, and in fact the absence of a lysine residue at the P_2 pocket is tolerated [23]. The conserved positively charged residues of nucleoplasmin are engaged in fewer interactions than the analogous ones in the monopartite NLSs that have been studied. Surprisingly, the non-conserved intervening sequence makes multiple interactions via its mainchain with conserved Kapα residues and contributes to the binding of this bipartite NLS.

Figure 4



Nucleoplasmin NLS recognition. **(a)** A simulated-annealing electron-density map is shown in white with the final model of the bipartite nucleoplasmin NLS. The 2.4 Å resolution F_o–F_c map is contoured at 2.5σ. **(b)** Ribbon representation of the nucleoplasmin NLS peptide (in purple) binding along the extended surface groove that shapes the surface of Kapα. The molecule is shown rotated about the horizontal axis by approximately 90° in respect to the orientation in Figure 1. The viewer looks down the extended surface groove as in Figure 2b. Polar inter-actions between the bipartite NLS backbone and conserved protein residues (white) along the spine of the surface groove are indicated by dotted lines. The figure was generated using RIBBONS [46] and GRASP [47].

Canonical NLS-containing proteins are all imported by yeast α and by the several known human α isoforms, albeit with different efficiencies [31]. The weaker binding of the yeast receptor for certain NLSs, including nucleoplasmin, observed in comparison to other species is at present unclear. These differences might be because of additional interactions of the full-length protein rather than to differential binding of their NLSs, given the level of conservation in the recognition events of the corresponding NLS motifs observed in the structures. The NLS motifs of the

SV40 T antigen, human c-*myc* and *Xenopus* nucleoplasmin bind *S. cerevisiae* Kapα50 at surface residues most of which are absolutely conserved across species.

The structures fully support the prediction that the classical monopartite NLSs are a very efficient variant of the downstream cluster of a bipartite NLS [21], in which strengthening of the downstream cluster can be achieved using a variable set of interactions. It is expected that other single NLSs might interact at the smaller site [32] by strengthening the interactions at the upstream cluster to optimally exploit the binding pockets in the C-terminal part of Kapα. The Stat protein is a probable candidate, as are the nuclear targeting signals of the NPI and LEF-1 proteins [32–34]. In this case the consensus sequence would be a K/R–K/R motif, which is too short to be unequivocally identified by simply looking at the sequence of interest. The surface groove C-terminal to the P_1' and P_2' pockets is not conserved across species, and the potential binding pockets at ARM 9 and ARM 10 will have characteristics that depend on which α homologue is involved in substrate recognition. In this context, a reasonable expectation from the structures is that mutation of two residues on the surface of the protein (Asp203 at the P_2 pocket and Glu402 at the P_2' pocket) might be sufficient to abolish binding of most if not all NLSs recognized by Kapα.

Although the sequence of the NLS peptide is somewhat variable, its conformation is strictly bestowed by hydrogen-bond interactions that straighten the NLS backbone into an extended conformation. This is presumably an enforced conformation, in which the peptide is molded on the protein template as in other 'peptide-surface' recognition events [35]. The monopartite NLS of the NFκB transcription factor for example, is flexible and unstructured in the nuclear form of the protein when bound to DNA [36], and is in a helical conformation in the cytoplasmic form when bound to its inhibitor IκBα [37]. Similarly, the N-terminal autoinhibitory segment of Kapα is in a helical conformation when bound to Kapβ [38], and in a β strand conformation when bound to Kapα [26]. This is even more striking, given the similarity of the structural frameworks of Kapα and Kapβ. Bipartite NLSs can analogously be partially disordered [39], or entirely helical [40]. It is unclear how the structure of the karyophilic protein is modified during recognition and transport into the nucleus, particularly in the case of bipartite NLS-containing proteins. The NLS would have to be presented on the surface of the protein in a conformation that is sufficiently unconstrained to allow its recognition and binding as an extended β strand.

Biological implications

Transport of cytoplasmically synthetized proteins into the nucleus is a highly regulated process in eukaryotes. It relies on the presence of a targeting signal (NLS or nuclear localization signal) encrypted in the sequence of the macromolecule to be imported and on its recognition by a nuclear transport receptor. In the case of the classical nuclear import pathway, the recognition of a variety of NLS cargoes is mediated by an adaptable adaptor molecule, karyopherin α (Kapα or importin α).

Two idiosyncratic features of Kapα seem to allow adaptability in substrate recognition without limiting specificity. First, the repeated architecture of the protein is responsible for the creation of an array of binding pockets, which are regularly spaced along the concave surface groove of the adaptor. Second, the relative flexibility of the protein scaffold allows fine-tuning of the modular arrangement of the binding pockets, which can be exploited to different extents according to the number of sidechains and sequence properties of each individual NLS. The segmental nature of the molecular construction of Kapα facilitates such small conformational rearrangements.

Classical monopartite signals utilize the large site in the N-terminal part of the surface groove with a multitude of interactions. A minimal set of positively charged residues plays a pivotal role in binding, supplemented by additional interactions that can be either electrostatic (SV40) or hydrophobic (c-*myc*). Bipartite NLSs like nucleoplasmin distribute their binding potential along the entire conserved surface groove. The abundance of the contacts made allows in this case suboptimal interactions to occur at the single binding pockets. In fact, optimal interactions along the whole of the surface groove might not be attainable and might also be not as advantageous given that the NLS-containing protein has to be released once in the nucleus. The Kapα adaptor protein appears to have evolved to maximize specificity and versatility in NLS recognition, but not necessarily affinity.

Materials and methods

Protein and peptide purification

A Tyr→Asp mutation was introduced at residue 397 in the *S. cerevisiae* Kapα construct spanning residues 88–530, which had been previously subcloned in a pProEx-HTb expression vector [24]. This single amino acid substitution was performed with the Quikchange site-directed mutagenesis kit (Stratagene), following the instructions of the manufacturer. The Y397D mutant of yeast Kapα 88–530 was isolated from an overproducing strain of *Escherichia coli* (BL21 DE3) and purified with similar protocols used for the wild-type protein [24]. The Y397D Kapα mutant exhibited improved expression and solubility properties as compared with the wild-type construct, with 30 mg of pure protein being obtained from each litre of bacterial culture. The pure protein was concentrated to 40 mg/ml in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10% glycerol and stored at −20°C.

An undecapeptide corresponding to the c-*myc* NLS (^{319}GPAAKRVKLDS^{329}) was expressed as a fusion with a portion of the Trp ΔLE sequence and purified from inclusion bodies in bacteria [41]. In this plasmid, the linker between the Trp-leader and the peptide sequence codes for a polyhistidine tract and for an adjacent methionine, to allow for affinity purification and for cyanogen bromide (CNBr) cleav-

age, respectively. The fusion polypeptide was expressed in *E. coli* (BL21 DE3) by induction at an OD_{600} of 0.3 with 0.5 mM IPTG. Cells from 3 l of culture were harvested by centrifugation and resuspended in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM PMSF. Following cell lysis with a French pressure cell and high-speed centrifugation, the insoluble pellets were collected and dissolved in 20 mM Tris-HCl pH 8.0, 6 M guanidinium-HCl (GuCl) by gentle rocking overnight. Cell debris were removed by high-speed centrifugation and the supernatant loaded onto a Ni-NTA resin (Qiagen), washed with 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 6 M urea, 5 mM imidazole, and eluted with an imidazole gradient. The peak fractions were extensively dialyzed against water, and the fusion polypeptide (about 100 mg as estimated by absorbance at 280 nM) was lyophilized. Cleavage of the downstream 116 amino acid sequence from the c-*myc* sequence was achieved by incubating the lyophilized material with 15 ml of 70% formic acid, 40 mg/ml CNBr for 90 min in the dark. The cleavage reaction was stopped by adding ten times the reaction volume of water. After lyophilization, the sample was solubilized in 20 mM Tris-HCl pH 8.0, 6 M GuCl and purified by reverse-phase high performance liquid chromatography (HPLC) using a C4 VIDAC column equilibrated with 0.1% trifluoroacetic acid (TFA). Elution with a 0.09% TFA, 70% acetonitrile (CH_3CN) gradient was monitored at a wavelength of 230 nM. The fraction containing the c-*myc* peptide eluted at a CH_3CN concentration of 29.4%, as confirmed using electrospray mass spectrometry (measured mass 1141.4 Da; predicted mass 1141.3 Da). A similar method was used for producing a nucleoplasmin NLS peptide, which eluted from the reverse-phase column at a CH_3CN concentration of 19.0%. The results reported in this paper, however, were obtained with a nucleoplasmin NLS-like peptide CATAQGGGA^{154}VKRPAATKKAGQAKKKKLD^{172}, a generous gift of Daniela Rhodes (MRC, Cambridge UK).

Crystallization and data collection

Co-crystallization experiments of wild-type Kapα50 with NLS peptides failed to yield crystals, the protein precipitating upon NLS addition. Co-crystals of the Y397D Kapα50 mutant with either the c-*myc* or nucleoplasmin peptides were readily obtained with standard crystallization and additive screenings (Hampton Research), using a protein concentration of 20 mg/ml and a threefold excess of peptide. X-ray data were measured at the Cornell High Energy Synchrotron Source (CHESS) A1 beamline and processed using the Denzo/HKL package [42]. A summary of the data collection statistics for both the c-*myc* and nucleoplasmin crystal forms is given in Table 1.

The c-*myc* NLS–Kapα mutant co-crystals were grown at 4°C by the vapor diffusion hanging-drop method using a reservoir solution containing 100 mM Tris-HCl pH 8.5, 200 mM sodium acetate (NaAce), 30% (w/v) polyethylene glycol (PEG) 4000. The crystals appeared in two weeks as thin plates with uneven edges, and were flash-frozen in liquid propane after a quick transfer in a solution containing 100 mM Tris-HCl pH 8.5, 100 mM (NaAce), 25% (w/v) PEG 4000, 20% glycerol. Despite being 40 micron thick at best, these crystals are well ordered and diffracted X-rays to 2.1 Å resolution using synchrotron radiation. Their space group is $P2_1$, with cell dimensions $a = 42.7$ Å, $b = 85.9$ Å, $c = 117.5$ Å, $\beta = 93.4°$. The crystals have a solvent content of 43% ($V_m = 2.1$) with two protein molecules in the asymmetric unit related by a twofold axis parallel and almost adjacent to the crystallographic *b* axis.

The nucleoplasmin NLS–Kapα mutant complex was co-crystallized by vapor diffusion at 4°C using 16% (w/v) PEG 8000, 250 mM potassium di-hydrogen phosphate (KH_2PO_4), 50 mM n-octylglucoside (BOG). The crystals were cryo-protected for 2 min in 15% (w/v) PEG 8000, 100 mM KH_2PO_4, 20 mM BOG, 20% glycerol and frozen in liquid propane. Crystallographic data were measured to 2.4 Å resolution. The diffraction is consistent with space group *C*2, with cell dimensions $a = 157.6$ Å, $b = 63.7$ Å, $c = 62.1$ Å, $\beta = 97.3$. The asymmetric unit contains one molecule, for a solvent content of 60% ($V_m = 3.1$).

Structure determination and quality of the model

Phases were determined using the molecular replacement method with the program AmoRe [43] using the Kapα50 model [24]. Rotation functions were carried out between 8 and 3.5 Å resolution, and resulted in an unambiguous peak at 11 σ and 14 σ above the mean value for the c-*myc* and the nucleoplasmin crystal forms respectively, with the next highest peak being below 5.5 σ. Translation function calculations between 10 and 4 Å resolution yielded two solutions for the c-*myc* co-crystals, with a correlation coefficient of 48.6 (correlation coefficient of 14.0 for the next highest peak). In the case of the nucleoplasmin co-crystals, the translation function gave a peak with a correlation coefficient of 44.2 (correlation coefficient of 33.6 for the next highest peak). Examination of the solutions on the graphics using the program O [44] revealed good packing and no stereochemical clashes with symmetry-related molecules.

Table 1

Data collection and refinement statistics.

	c-*myc* NLS	Nucleoplasmin NLS
Data		
Space group	$P2_1$	*C*2
Maximum resolution (Å)	2.1	2.4
Independent reflections	47,949	21,907
No. of measurements	152,954	105,706
Mosaicity	0.42	0.20
Completeness (%)*	96.9 (96.9)	90.9 (80.0)
R_{sym} (%)*	8.5 (21.3)	5.7 (28.6)
I/σ*	15.8 (6.6)	32.7 (4.6)
Refinement		
Resolution limits (Å)	30–2.1	30–2.4
R_{free} %	27.7	26.5
$R_{working}$ %	24.6	24.1
$\phi\psi$ most favoured (%)	96.2	93.7
Rmsd bond (Å)	0.016	0.008
Rmsd angle (°)	1.74	3.45
Average B factor	21.4	67.2
Protein residues	423†	421
NLS peptide residues	13†	19
Water molecules	160†	32

*Values for the outermost resolution shell are given in parentheses.
†Values for one of the two complexes in the asymmetric unit.

The models were refined with the maximum-likelihood target in the program CNS [45]. No sigma cut-off was applied to the data and low-resolution terms to 30 Å were included in the refinement and a random sample containing 5% of the data was excluded. The progress of the refinement was judged by monitoring the agreement between calculated and observed structure factors for the excluded reflections (R_{free}).

In the c-*myc* crystal form, strict noncrystallographic twofold symmetry was initially enforced. Rounds of rigid-body refinement were carried out with Kapα50 subdivided in four domains, comprising residues 89–204 (ARM 1,–3), 205–330 (ARM 4–6), 331–417 (ARM 7, 8) and 418–509 (ARM 9, 10). The model was then subjected to cycles of positional refinement gradually increasing the resolution limits from 3.5 Å to 2.1 Å, including cycles of atomic temperature-factor refinement beyond 2.6 Å resolution. The placement of the c-*myc* peptide was straightforward at the two NLS binding sites present on the surface of each molecule, and unambiguous electron density was also present for the Y397D substitution of the Kapα mutant. Noncrystallographic constraints were released and subsequent model building and refinement were carried out independently for the two molecules in the asymmetric unit. The final model has a R_{free} is 27.7% to 2.1 Å resolution and good stereochemistry (Table 1).

In the nucleoplasmin crystal form, initial rigid-body refinement was performed with Kapα subdivided in the above-mentioned four domains and followed by cycles of positional and B-factor refinement with a stepwise increase in resolution. Electron-density maps calculated at this stage showed strong density corresponding to the bound peptide. Rounds of model building and refinement to 2.4 Å resolution finally resulted in a R_{free} of 26.5 with good statistics (Table 1). The model comprises 417 amino

acid residues for the yeast mutant (90–506), 19 amino acid residues for nucleoplasmin (153–171) and 31 water molecules. The average temperature factors are 61.3 Å^2 for all protein atoms and 72.4 Å^2 for the nucleoplasmin NLS, suggesting a partial occupancy of 85% for the peptide in this co-crystal. Despite the relatively high B factor values, the electron density for the nucleoplasmin NLS peptide is of good quality.

Accession numbers

The atomic coordinates of the human c-*myc* NLS–yeast Kapα50 complex and of the *Xenopus* nucleoplasmin NLS–yeast Kapα50 complex have been deposited in the Protein Data Bank with accession codes 1EE4 and 1EE5, respectively.

Acknowledgements

We thank Lore Leighton for the Y397D karyopherin α clone, Daniela Rhodes for the gift of nucleoplasmin peptide, Peter Kim for the Trp ΔLE plasmid, Satish Nair for advise on CNBr cleavage, Joseph Marcotrigiano for help with electrospray mass spectrometry and the staff at the Cornell High Energy Synchrotron Source for assistance. We also wish to thank the members of the Kuriyan lab for help and advise throughout, and Daniela Rhodes and Iain Mattaj for critical reading of the manuscript. EC was supported by a Human Frontier Science Program fellowship.

References

1. Nakielny, S., and Dreyfuss, G. (1999). Transport of proteins and RNAs in and out of the nucleus. *Cell* **99**, 677-690.
2. Mattaj, I.W. & Englmeier, L. (1998). Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.* **67**, 265-306.
3. Wozniak, R.W., Rout, M.P. & Aitchison, J.D. (1998). Karyopherins and kissing cousins. *Trends Cell Biol.* **8**, 184-188.
4. Adam, E.J. & 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.
5. Imamoto, N., Tachibana, Y., Matsubae, M. & Yoneda, Y. (1995). A karyophilic protein forms a stable complex with cytoplasmic components prior to nuclear pore binding. *J. Biol. Chem.* **270**, 8559-8565.
6. Enenkel, C., Blobel, G. & Rexach, M. (1995). Identification of a yeast karyopherin heterodimer that targets import substrate to mammalian nuclear pore complexes. *J. Biol. Chem.* **270**, 16499-16502.
7. Görlich, D., *et al*., & 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.
8. Moroianu, J., Hijikata, M., Blobel, G. & Radu, A. (1995). Mammalian karyopherin α1β and α2β heterodimers: α1 or α2 subunit binds nuclear localization signal and β subunit interacts with peptide repeat-containing nucleoporins. *Proc. Natl Acad. Sci. USA* **92**, 6532-6.
9. Görlich, D., Prehn, S., Laskey, R.A. & Hartmann, E. (1994). Isolation of a protein that is essential for the first step of nuclear protein import. *Cell* **79**, 767-778.
10. Weis, K., Mattaj, I.W. & Lamond, A.I. (1995). Identification of hSRP1α as a functional receptor for nuclear localization sequences. *Science* **268**, 1049-1053.
11. Moroianu, J., Blobel, G. & Radu, A. (1995). 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.
12. Rexach, M. & Blobel, G. (1995). Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. *Cell* **83**, 683-692.
13. Görlich, D., Pante, N., Kutay, U., Aebi, U. & Bischoff, F. R. (1996). Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J.* **15**, 5584-5594.
14. Kutay, U., Bischoff, F.R., Kostka, S., Kraft, R. & Görlich, D. (1997). Export of importin α from the nucleus is mediated by a specific nuclear transport factor. *Cell* **90**, 1061-1071.
15. Dingwall, C., Sharnick, S.V. & Laskey, R.A. (1982). A polypeptide domain that specifies migration into the nucleus. *Cell* **30**, 449-458.
16. Dingwall, C. & Laskey, R. A. (1991). Nuclear targeting sequences – a consensus? *Trends. Biol. Sci.* **16**, 178-181.
17. Kalderon, D., Roberts, B.L., Richardson, W.D. & Smith, A.E. (1984). A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499-509.
18. Lanford, R.E. & Butel, J.S. (1984). Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell* **37**, 801-813.
19. Colledge, W.H., Richardson, W.D., Edge, M.D. & Smith, A.E. (1986). Extensive mutagenesis of the nuclear location signal of simian virus 40 large-T antigen. *Mol. Cell Biol.* **6**, 4136-4139.
20. Dingwall, C., Robbins, J., Dilworth, S.M., Roberts, B. & 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.
21. Robbins, J., Dilworth, S.M., Laskey, R.A. & Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615-623.
22. Dang, C.V. & Lee, W.M.F. (1988). Identification of the human c-*myc* protein nuclear translocation signal. *Mol. Cell Biol.* **8**, 4048-4054.
23. Makkerh, J.P.S., Dingwall, C. & Laskey, R.A. (1996). Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. *Curr. Biol.* **6**, 1025-1027.
24. Conti, E., Uy, M., Leighton, L., Blobel, G. & Kuriyan, J. (1998). Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α. *Cell* **94**, 193-204.
25. Görlich, D., Henklein, P., Laskey, R.A. & Hartmann, E. (1996). A 41 amino acid motif in importin-α confers binding to importin-β and hence transit into the nucleus. *EMBO J.* **15**, 1810-1817.
26. Kobe, B. (1999). Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin α. *Nat. Struct. Biol.* **6**, 388-397.
27. Dingwall, C. & Laskey, R.A. (1998). Nuclear import: a tale of two sites. *Curr. Biol.* **8**, R922-R924.
28. Schreiber, V., Molinete, M., Boeuf, H., de Murcia, G. & Menissier-de Murcia, J. (1992). The human poly (ADP-ribose) polymerase nuclear localization signal is a bipartite element functionally separate from DNA binding and catalytic activity. *EMBO J.* **11**, 3263-3269.
29. Moll, T., Tebb, G., Surana, U., Robitsch, H. & Nasmyth, K. (1991). The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SWI5. *Cell* **66**, 743-758.
30. Chelsky, D., Ralph, R. & Jonak, G. (1989). Sequence requirements for synthetic peptide-mediated trasnslocation to the nucleus. *Mol. Cell Biol.* **9**, 2487-2492.
31. Köhler, M., *et al.* & Hartmann, E. (1999). Evidence for distinct substrate specificities of importin a family members in nuclear protein import. *Mol. Cell Biol.* **19**, 7782-7791.
32. Herold, A., Truant, R., Wiegand, H. & Cullen, B.R. (1998). Determination of the functional domain organization of the importin α nuclear import factor. *J. Cell Biol.* **143**, 309-318.
33. Sekimoto, T., Imamoto, N., Nakajima, K., Hirano, T. & 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.
34. Wang, P., Palese, P. & O'Neill, R. E. (1997). The NPI-1/NPI-3 (karyopherin α) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal. *J. Virol.* **71**, 1850-6.
35. Harrison, S.C. (1996). Peptide-surface association: the case of PDZ and PTB domains. *Cell* **86**, 341-343.
36. Müller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L. & Harrison, S.C. (1995). Structure of the NF-κB p50 homodimer bound to DNA. *Nature* **373**, 311-317.
37. Jacobs, M.D. & Harrison, S.C. (1998). Structure of an IκBα/NF-κB complex. *Cell* **95**, 749-758.
38. Cingolani, G., Petosa, C., Weis, K. & Müller, C.W. (1999). Structure of importin-β bound to the IBB domain of importin-α. *Nature* **399**, 221-229.
39. Schwabe, J.W., Neuhaus, D. & Rhodes, D. (1990). Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature* **348**, 458-461.
40. Ellenberger, T.E., Brandl, C.J., Struhl, K. & Harrison, S.C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the DNA complex. *Cell* **71**, 1223-1237.
41. Staley, J.P. & Kim, P.S. (1994). Formation of a native-like subdomain in a partially folded intermediate of bovine pancreatic trypsin inhibitor. *Protein Sci.* **3**, 1822-1832.
42. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
43. CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
44. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
45. Brünger, A.T., *et al.* & Warren, G.L. (1998). Crystallographic and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.
46. Carson, M. (1991). Ribbons 2.0. *Acta Crystallogr.* **24**, 958-961.
47. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and