Nuclear import: **A tale of two sites**

Colin Dingwall* and Ronald A. Laskey†

The recently determined crystal structure of a nuclear localization sequence receptor has revealed an exquisitely specific interaction between ligand and receptor, and explains how simple and complex nuclear localization signals can both be recognized specifically by the same molecule.

Addresses: *Department of Pharmacology, State University of New York at Stony Brook, Stony Brook, New York 11794-8651, USA. †Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, UK.

Current Biology 1998, 8:R922–R924
http://biomednet.com/elecref/09609822008R0922

Recent years have seen dramatic progress in elucidating the mechanisms by which specific proteins move between the cytoplasm and nucleus. These recent advances build on earlier studies that defined short stretches of amino acids in nuclear proteins that act as 'nuclear localization signals' (NLSs), specifying their import into the nucleus [1–5]. These early studies posed a puzzle, in that the sequenced NLSs were found to fall into two quite distinct classes: a simple type consisting of a cluster of positively charged lysines and arginines [2,3], and a more complex, bipartite type consisting of two separate basic clusters with an intervening spacer [1,4–7]. How could these two structurally quite different types of NLS be recognized by the same import receptor? The recently determined crystal structure of an import receptor [8] suggests a solution to this conundrum.

The essential elements of the bipartite NLS were found to be two clusters of basic amino acids: a smaller one of two basic amino acids and a larger one with at least three basic amino acids in a group of five, separated by a ten amino-acid spacer [6]. Surprisingly, the precise amino-acid sequence of the spacer was found not to be important, and certain much longer spacer segments could be tolerated and still promote efficient nuclear targeting [5,7]. Similar bipartite motifs have been found in more than half the nuclear proteins in a sequence database, but in only 4% of non-nuclear proteins [6].

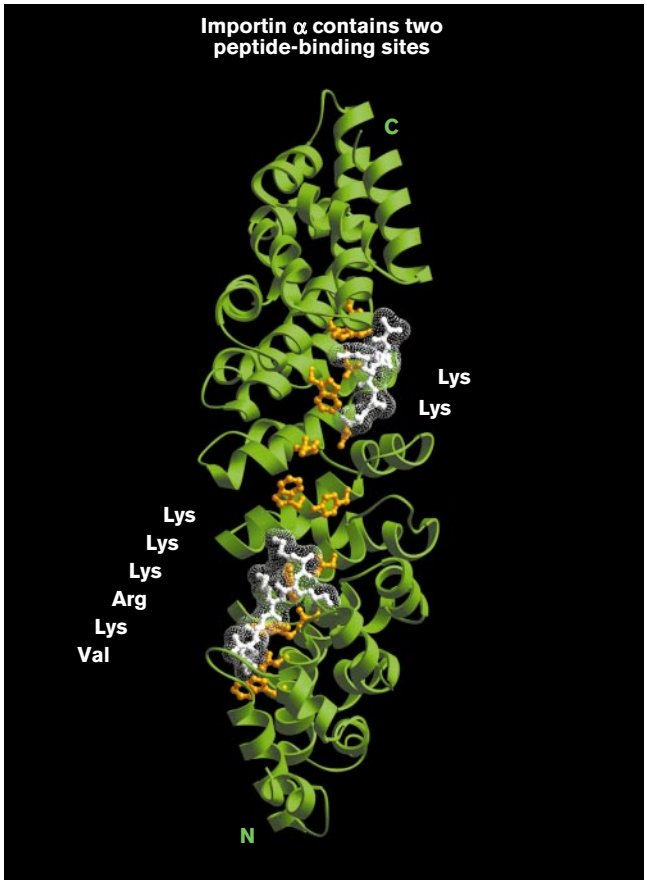
The recently reported [8] crystal structure of the budding yeast NLS receptor known as either importin α or karyopherin α has clearly revealed two distinct binding sites that can accommodate both essential elements of the bipartite NLS. In the crystal, a peptide corresponding to the simple NLS of simian virus 40 (SV40) T antigen is bound to both sites of the receptor independently, explaining how the same receptor can specifically bind the very different simple and bipartite NLSs. The structure has been determined both with and without soaking the crystal in the T antigen NLS. It reveals an extraordinarily beautiful receptor design, in which the NLS fits into its receptor like a hand fitting into a hand print in soft clay. The fingers are the lysine or arginine side chains; their hydrophobic stems are gripped between parallel hydrophobic tryptophan residues, and their positively charged tips are held electrostatically by negative charges at the ends of the finger imprints. The distance between the amino groups of the NLS and their interacting carboxyl groups in the receptor is relatively large, allowing either lysine or arginine to be recognized.

To put this in context, we need to review briefly what is now known about the mechanisms of nuclear protein import. It is now possible to reconstitute nuclear protein import fully using permeabilized cells and recombinant soluble import factors. The steps in the nuclear import of a protein that has a conventional basic NLS are now understood in some detail [9,10]. Such an NLS is recognized by a heterodimeric receptor in the cytoplasm, which docks at the cytoplasmic face of the nuclear pore complex. The two subunits of the receptor were named importin or karyopherin α and β (but from here on we shall just use the importin terminology) [11–13]. The specific recognition of the NLS in the target nuclear protein is the function of the α subunit, while the β subunit acts to dock the complex between the receptor dimer and its nuclear protein cargo with the nuclear pore [9–11].

The release of the nuclear protein cargo from the docking site, and its translocation through the channel of the nuclear pore complex, require the small GTPase Ran [9,10]. The docked state is stabilized by Ran–GDP, but destabilized by Ran–GTP. Translocation to the nuclear interior requires hydrolysis of the Ran-bound GTP; non-hydrolyzable GTP analogs and dominant-negative mutant forms of Ran have been shown to inhibit nuclear accumulation of nuclear proteins and other karyophiles [9,10]. The presence of Ran–GTP in the nucleus promotes dissociation of the importin α–β heterodimer, and release of the importin α–nuclear protein complex into the interior of the nucleus. Importins α and β are recycled to the cytoplasm by different routes for further rounds of nuclear protein import [9,10].

The sequence of importin α revealed that the protein is organized into three distinct domains [13,14]. An amino-terminal domain rich in basic amino acids is required for interaction with the β subunit [9,10]. A large central domain is made up of eight to ten repeats of a 42 residue sequence, similar to repeats that were originally identified in the

Figure 1



Two peptide-binding sites in the NLS receptor importin α (also known as karyopherin α). The figure shows a ribbon diagram of the importin α polypeptide, highlighting the small and large peptide-binding sites. The conserved residues in repeat helix 3 are shown in yellow, and the amino acids of the NLS peptides that make specific contacts are shown in white.

Drosophila Armadillo protein [15]. This is followed by a third, carboxy-terminal domain of unknown function. Analysis of human homologs of importin α indicated that the central repeat region forms the NLS-binding domain, with binding sites at each end of the repeat domain [15].

Related repeats have been found in a variety of functionally unrelated proteins, including the intracellular signaling molecule β-catenin. The structure of a proteolytic fragment consisting of the 12 repeats of β-catenin showed that the overall structure of this domain is a superhelix of alpha helices [16]. Each individual repeat folds into three alpha helices, which then pack into a right-handed superhelix, individual repeats making extensive interactions with adjacent repeats. A similar superhelical organization is seen in the crystal structure of the 10 repeats of budding yeast importin α [8]. The superhelix has a rod-like structure with a curved surface groove (Figure 1). This groove is formed by the third, and longest alpha helix of each repeat, and key amino acids in this helix are involved directly in NLS recognition.

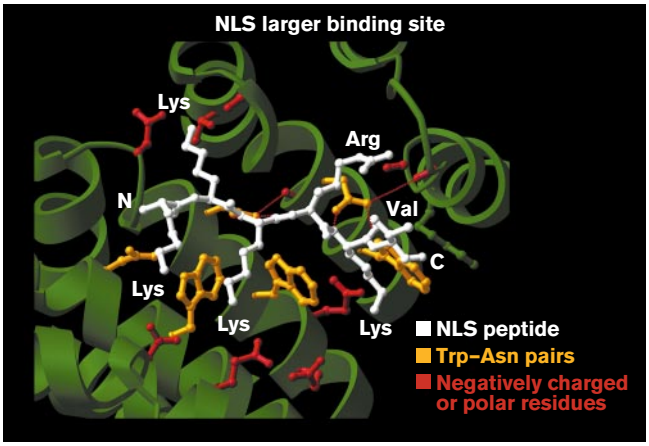
When the importin α crystals were soaked in the SV40 T antigen NLS peptide — which in the single-letter amino acid code has the sequence SPKKKRKVE — two distinct sites on the concave surface of the receptor superhelix were each seen to bind an NLS peptide in an extended beta-strand conformation (Figure 1). The larger site is at the amino-terminal end of the receptor between the second and fourth arm repeats. The smaller site is located at the carboxy-terminal end, between repeats seven and eight. At each site, the NLS peptide binds in an antiparallel direction via interactions with conserved tryptophan and asparagine pairs in repeat helix 3 (Figure 2). The antiparallel direction is determined by backbone interactions with the conserved asparagine residues.

Specific recognition of the NLS peptide is achieved through a combination of two types of interaction. There are hydrophobic interactions that involve the aliphatic portions of the peptide side chains and the conserved receptor tryptophan residues. And electrostatic interactions occur between the basic groups of the NLS peptide and conserved negatively charged residues at the ends of the hydrophobic grooves (Figure 2). At the smaller site, between the seventh and eighth arm repeats, only two residues in the bound NLS peptide are recognized specifically, but the same structural elements are involved in binding.

There are two particularly noteworthy features of the structure that are relevant to attempts to understand the unusual characteristics of bipartite NLSs. The first is that the smaller peptide-binding site allows optimal recognition of only two basic amino acids, whereas the larger site allows for the optimal recognition of five lysine or arginine residues (Figure 1). This structural information coincides exactly with the properties of bipartite NLSs that were inferred from detailed mutational analysis [5,7]. Furthermore, the nature of the donor–acceptor hydrogen bonding interactions at the two sites is consistent with the observation that, in bipartite NLSs, the larger cluster of basic residues must be carboxy-terminal to the smaller cluster.

The second relevant feature of the structure is that the separation of the two sites in the receptor structure would accommodate a 10 residue spacer linking the two basic NLSs, but a shorter spacer would not reach between the two clusters. In the bipartite NLS of nucleoplasmin, the essential clusters of basic amino acids are separated by a 10 residue spacer; the spacer can be longer and still direct a protein to the nucleus, but not shorter [5,7]. Further characterization of the bipartite NLS revealed that the sequence of the spacer is not important for nuclear import,

Figure 2



Molecular interactions at the larger peptide-binding site of importin α. The NLS peptide is shown in white and the tryptophan–asparagine pairs are shown in yellow. Negatively charged or polar residues are shown in red. Shallow apolar pockets, formed by the tryptophan residues, accommodate the aliphatic portions of the lysine side chains, while the negatively charged and polar residues are positioned to make specific electrostatic interactions with the nitrogen of the lysine side chains.

in fact it can be replaced by polyalanine or extended to 20 alanine residues without disrupting efficient nuclear import [7]. In the receptor structure [8], the regularity of the conserved asparagine–tryptophan repeat breaks down in the centre of the molecule (repeats 5 and 6), precisely where the spacer segment of a bipartite NLS would be positioned, consistent with a lack of specific sequence recognition in this part of the NLS.

The import of target proteins into the nucleus is, in many instances, modulated by phosphorylation of amino acids close to the basic residues of the NLS. Examples of this include the positive and negative effects of phosphorylating residues close to the NLS of SV40 large T antigen, and the negative effects of phosphorylating residues adjacent to the NLS of lamin B_2 [17,18]. The ability of a number of bipartite NLSs to target proteins to the nucleus is regulated in a similar manner [19]. Now we know the structure of importin α and how it interacts with NLS peptides [8], it is interesting to ask whether there is any indication that these specific phosphorylation events might have a direct influence on NLS binding.

In a number of cases, the phosphorylated amino acids would be predicted to contact importin α at sites where the bipartite spacer would be positioned. For example, this is where phosphorylated residues amino-terminal to the SV40 large T antigen NLS bound at the larger site in importin α would be located (Figure 1). In addition, many naturally occurring bipartite spacers contain acidic amino acids [6], suggesting that the simple presence or absence of a negatively charged, phosphorylated residue in this region may not be sufficient to affect directly the interaction with importin α.

The structure of the NLS receptor importin α [8] has revealed and explained much more than could possibly have been expected, and has also raised further interesting questions. Hopefully the structural analysis of members of the importin β family, together with the diverse signals that they recognize, will yield equally exciting and informative results.

Acknowledgements

We would like to thank Elena Conti and John Kuriyan for their helpful comments and generously supplying figures. C.D. would like to thank the Council for Tobacco Research for their generous support. R.A.L. is grateful to the Cancer Research Campaign for generous support.

References

1. Dingwall C, Sharnick SV, Laskey RA: **A polypeptide domain that specifies migration into the nucleus.** *Cell* 1982, **30:**449-458.
2. Lanford RE, Butel JS: **Construction and characterisation of an SV40 mutant defective in nuclear transport of T antigen.** *Cell* 1984, **37:**801-813.
3. Kalderon D, Roberts BL, Richardson WD, Smith AE: **A short amino acid sequence able to specify nuclear location.** *Cell* 1984, **39:**499-509.
4. Dingwall C, Robbins J, Dilworth SM, Roberts B, Richardson WD: **The nucleoplasmin nuclear localization sequence is larger and more complex than that of SV-40 large T anitgen.** *J Cell Biol* 1988, **107:**841-849.
5. Robbins J, Dilworth SM, Laskey RA, Dingwall C: **Two interdependent basic domains in nucleoplasmin targeting sequence: identification of a class of bipartite nuclear targeting sequence.** *Cell* 1991, **64:**615-623.
6. Dingwall C, Laskey RA: **Nuclear targeting sequences: a consensus?** *Trends Biochem Sci* 1991, **16:**178-181.
7. Makkerh JPS, Dingwall C, Laskey RA: **Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids.** *Curr Biol* 1996, **6:**1025-1027.
8. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J: **Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α.** *Cell* 1998, **94:**193-204.
9. Gorlich D, Mattaj IW: **Nucleocytoplasmic transport.** *Science* 1996, **271:**1513-1518.
10. Nigg EA: **Nucleocytoplasmic transport: signals, mechanisms and regulation.** *Nature* 1997, **386:**779-787.
11. Gorlich D, Kosta S, Kraft R, Dingwall C, Laskey RA, Hartmann E, Prehn S: **Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope.** *Curr Biol* 1995, **5:**383-392.
12. Wozniak RW, Rout MP, Aitchison JD: **Karyopherins and kissing cousins.** *Trends Cell Biol* 1998, **8:**184-188.
13. Gorlich D, Prehn S, Laskey RA, Hartmann E: **Isolation of a protein that is essential for the first step of nuclear protein import.** *Cell* 1994, **97:**767-778.
14. Peifer M, Berg S, Reynolds AB: **A repeating amino acid motif shared by proteins with diverse cellular roles.** *Cell* 1994, **76:**789-791.
15. Sekimoto T, Imamoto N, Nakajima K, Hirano T, Yoneda Y: **Extracellular signal-dependent nuclear import of STAT1 is mediated by nuclear pore-targeting complex formation with NPI-1, but not Rch1.** *EMBO J* 1997 **16:**7067-7077.
16. Huber AH, Nelson WJ, Weis WI: **Three-dimensional structure of the armadillo repeat region of β-catenin.** *Cell* 1997, **90:**871-882.
17. Hubner S, Xiao C-Y, Jans D: **The protein kinase CK2 site (Ser 111/112) enhances recognition of the simian virus 40 large T antigen nuclear localization sequence by importin.** *J Cell Biol* 1997, **272:**17191-17195.
18. Hennekes H, Peter M, Weber K, Nigg E: **Phosphorylation of protein kinase C sites inhibits nuclear import of lamin B2.** *J Cell Biol* 1993, **102:**1293-1394.
19. Jans D: **The regulation of protein transport to the nucleus by phosphorylation.** *Biochem J* 1995, **311:**715-716.