The distribution of different classes of nuclear localization signals (NLSs) in diverse organisms and the utilization of the minor NLS-binding site inplantnuclear import factor importin-α

Chiung-Wen Chang[1,2], Rafael Miguez Couñago[1,2], Simon J. Williams[1,2], Mikael Boden[1,3] and Bostjan Kobe[1,2,*]

[1]School of Chemistry and Molecular Biosciences and Institute for Molecular Bioscience, University of Queensland, Brisbane, Qld 4072, Australia

[2]Australian Infectious Diseases Research Centre, University of Queensland, Brisbane, Qld 4072, Australia

[3]School of Information Technology and Electrical Engineering, University of Queensland, Brisbane, Qld 4072, Australia.

*Correspondence to: Bostjan Kobe;Email: b.kobe@uq.edu.au

Addendum to:

Chang CW, Couñago RL, Williams SJ, Bodén M, Kobe B.Crystal structure of rice importin-α and structural basis of its interaction with plant-specific nuclear localization signals.Plant Cell 2012; 24:5074-88;PMID:23250448; doi: 10.1105/tpc.112.104422.

Submitted: 26 July 2013

Accepted:

Key words: importin-α, nuclear-cytoplasmic transport, nuclear localization signal, *Oryza sativa*, *Arabidopsis thaliana*

Correspondence to:

Bostjan Kobe

Email: b.kobe@uq.edu.au

Abstract

The specific recognition between the import receptor importin-α and the nuclear localization signals (NLSs) is crucial to ensure the selective transport of cargoes into the nucleus. NLSs contain one or two clusters of positively-charged amino-acids, which usually bind to the major (monopartite NLSs) or both minor and major NLS-binding sites (bipartite NLSs). In our recent study, we determined the structure of importin-α1a from rice (*Oryza sativa*), and made two observations that suggest an increased utilization of the minor NLS-binding site in this protein. Firstly, unlike the mammalian protein, both the major and minor NLS-binding sites are auto-inhibited in the unliganded rice protein. Secondly, we showed thatNLSs of the 'plant-specific' class preferentially bind to the minor NLS-binding site of rice importin-α. Here, we show that a distinct group of 'minor site-specific' NLSs also bind to the minor site of the rice protein. We further show a greater enrichment of proteins containing these "plant-specific' and 'minor site-specific' NLSs in the rice proteome. However, the analysis of the distribution of different classes of NLSs in diverse eukaryotes shows that in all organisms, the minor site-specific NLSs are much less prevalent than the classical monopartite and bipartite NLSs.

TEXT

In eukaryotes, selective import of proteins into the nucleus is an essential process and subject to stringent regulatory controls. The classical nuclear import pathway employs importin-α (Impα) as an adaptor protein that recognizes the nuclear localization signal (NLS) on the protein destined to the nucleus.[1-3] The NLSs contain one cluster (monopartite NLSs) or two clusters of basic residues (bipartite NLSs), connected by a linker region of ~10-12 residues. There are two separate NLS-binding sites on Impα, termed major and minor NLS-binding sites, which can accommodate these basic clusters. Although our knowledge of nuclear import in plants is less advanced than the understanding in mammals and yeast, an increasing number of components of the plant nuclear import machinery have been identified in recent years.[4,5]

While the general characteristics of the transport machinery are conserved between plants and other organisms, some differences have been observed in terms of preferences for NLSs,[6-8] the involvement of plant-specific components [9] and the functionality of the import receptors.[10] Impα from *Arabidopsis thaliana* (AtImpα) can recognize three different classes of NLSs: (i) the classical monopartite (exemplified by the NLS from the simian virus40 large T-antigen (SV40TAgNLS)) and (ii) bipartite (exemplified by NLS found in a maize Opaque-2 transcription factor) NLSs, and (iii) the NLSs related to the NLS from yeast Matα2 [11-14].The Matα2-like NLSs have been reported to be functional in yeast and plants, but have been shown not to bind to Impα1 from rice.[13,15,16]

A random peptide library screen applied to human, plant, and yeast Impα variants suggested six classes of NLS consensus sequences,[17] comprising classical monopartite (class-1 and -2) and bipartite (class-6) NLSs, and three new classes: minor site-specific (class-3 and -4) and plant-specific(class-5) NLSs (Table 1). The molecular basis of the binding of NLSs from these six classes to Impα has not been fully elucidated. We recently demonstrated that class-5 plant-specific NLSs show stronger binding to rice Impα1a (rImpα1a) than to the mouse (mImpα) and yeast (yImpα) proteins, and that they bind preferentially to the minor NLS-binding site of rImpα1a.[18] Interestingly, the consensus sequence of class-5 plant-specific NLSs shows only limited similarities to the consensus sequences of theclass-3 and -4 minor site-specific NLSs[17] (Table 1).

Here, we aimed to further characterize the distinct utilization of the minor NLS-binding site in rImpα1a. We first tested the binding of a class-3 minor site-specific NLS[17-19] to rImpα1a,and show that it binds with nM affinity and preferentially to the minor NLS-binding site. Structure analyses suggest that this NLS can bind to the minor NLS-binding site of rImpα1a in an analogous conformation as to mImpα, and that similar reasons prevent it from binding to the major site of both the rice and mouse proteins.We then analyzed bioinformatically the distribution of the six classes of NLSs in different yeast, plant and mammalian proteomes. These data indicate a greater prevalence

of proteins containing class-5 plant-specific NLSs as well as class-3 minor site-specific NLSs in the rice proteome, suggesting a greater usage of the minor NLS-binding site by rice Impα proteins. However, the class-5 and class-3 minor site-specific NLSs are rare in all organisms, and the classical monopartite (class-1 and -2) and bipartite NLSs account for the majority of identified NLSs.

Mutational analysis confirms the binding of class-3 minor site-specific NLSs to the minor site of rimpα1a

While we demonstrated that plant-specific NLSs bind to the minor NLS-binding site of rImpα1awith nM affinity,[18] their consensus sequence differs significantly from the class-3 and class-4 minor site-specific NLSs characterized by Kosugi and co-workers[17] (Table 1). Here, we investigated the binding of the peptide B6 ($S^{1}SHRKRKFSDAF^{12}$), a representative of the class-3 minor site-specific NLSs,[17] to rImpα1aΔIBB (rImpα1a lacking the importin-β-binding domain[18]). Our data indicate that B6 binds strongly to rImpα1aΔIBB, with an affinity of 23 nM (Table 2).This affinity falls in the range between 10 nM to 1μM proposed for functional NLSs.[20-22] B6 binding is only affected marginally when a major NLS-binding site mutant[18] ($rImp\alpha1a\Delta IBB^{D188K}$) is used. By contrast, a point mutation in the minor NLS-binding site[18] ($rImp\alpha1a\Delta IBB^{E388R}$) results in a 30-fold decrease in the binding affinity between B6 and rImpα1aΔIBB (Table 2). These results confirm that the class-3 minor site-specific NLSs utilize the minor NLS-binding site as a preferential binding site in rImpα1a, consistent with their interaction with mouse Impα.[19]

The structural basis of class-3 minor site-specific NLS binding to rImpα1a

To investigate the structural basis of class-3 minor site-specific NLS interacting withrImpα1a, we superimposed the structure of rImpα1aΔIBB (from the SV40TAgNLScomplex; PDB ID 4B8O)[18] onto the structure of mImpα ΔIBB in complex with the B6peptide (PDB ID3ZIQ)[19] (the root-mean-square distance (RMSD) for 374 Cα atoms is 1.62 Å). The superposition shows that the peptide

conformation in the mImpα complex is compatible with its binding to rImpα1a (Fig. 1A). The B6 peptide-binding determinants are conserved between the mouse and rice proteins. While the basic cluster (R^4KRK^7) in the B6 peptide binds in a conformation analogous to classical NLSs binding to mImpα, the C-terminal region of B6 and other class-3 minor site-specific NLSs forms a α-helical turn,[19] which is distinct from other conformations adopted by NLSs binding to Impα.[3,18,21,23-27] Superposition of the entire B6 peptide in its minor site-binding conformation onto the major NLS-binding site shows a steric clash with the N-terminal region of rImpα1aΔIBB (Fig. 1B), analogous to what is observed in mImpα.[19] The analysis supports our results on the binding to the rImpα1a ΔIBB mutant proteins with substitutions in the NLS-binding sites. The two residues in the minor NLS-binding site (Arg315 and Lys353 in mImpα) that are involved in stabilising the formation of the α-helical turn by forming cation-π interactions[28] with the B6 residues Phe8 and Phe12 in the structure of the B6:mImpα1aΔIBB complex are conserved in rImpα1a (Arg306 and Lys345).Our structural analysis supports the conclusion that class-3 minor site-specific NLSs can bind to rImpα in a manner analogous to their binding to mImpα.

Distribution of the six classes of NLSs in the proteomes from different organisms

NLS binding to rice Impα suggests an increased usage of the minor NLS-binding site for nuclear cargo proteins in rice and presumably plants in general. To compare the distribution of the six classes of NLS sequences in the proteomes from representative plants, yeast, and mammals, we performed bioinformatic analyses using two different approaches. The two approaches used regular expression patterns and position weight matrices (PWMs),[18,29] respectively, to describe the six classes of NLS, based on the data by Kosugi and coworkers[17](Fig. 2). The two approaches were used to screen the complete proteomes of plants (monocots *Oryza sativa* and *Sorghum vulgare*; dicots *A. thaliana*, *Vitis vinifera*, *Solanum tuberosum* and *Solanum lycopersicum*), yeast (*Saccharomyces cerevisiae*) and mammals (*Homo sapiens* and *Mus musculus*). The alignments provided by Kosugi and coworkers were modified to include data from their amino-acid

replacement analysis[17]. For bipartite NLSs, three PWMs were constructed based on different lengths of the linker region (and designated here as classes 6, 7, and 8 for 10, 11, and 12 residues in the linker region, respectively). The threshold for the PWM score of each class was determined to obtain the maximum Matthews correlation coefficient (MCC) for each organism. The MCC was calculated from the absolute counts of the protein sequences for true and false positives and negatives, to indicate the quality of the binary classification for each proteome, based on nuclear localization as annotated by the Gene Ontology (GO) in UniProt (cellular component "nucleus" or any of its sub-compartments).

Tables 3 and 4 show the results based on both approaches. Like simple consensus sequences, the limitation of the regular expression approach is that it is rigid (requires an exact match). Albeit limited in terms of the dependencies they capture, PWMs can model degrees of interaction between the NLS and Impα.[30] The PWM approach is therefore preferred, however to reach its full potential, it requires rich data.[29] In this particular case, the data that the representations of the six classes of NLSs are based on[17] are limited, which should be considered when interpreting the results. Overall, the analysis shows that across all the proteomes compared, proteins containing the 'classical' monopartite (class-1 and -2) and bipartite NLSs are much more prevalent than the non-classical NLSs (class-3 and -4 minor site-specific, and class-5 plant-specific NLSs). The data confirm the observations from our previous study[18] of a greater prevalence of class-5 plant-specific NLSs in the rice proteome. The rice proteome also shows a greater proportion of class-3 minor site-specific NLSs, compared to the other plant species, suggesting a greater usage of the minor NLS-binding site in rice Impα protein. However, even in rice, the class-5 and class-3 minor site-specific NLSs are the rarest NLS classes, with class-4 minor site specific NLSs bring significantly more common, and the classical monopartite (class-1 and -2) and bipartite NLSs accounting for the majority of identified NLSs.

Conclusions

In plants, the ancestral Impα1-like gene diversified to give rise to different Impα variants.[31,32] Phylogenetic studies indicate that distinct numbers of Impα variants exist in different plants.[4,18,33] Most studies of plant Impα proteins used *A. thaliana*, which contains nine different variants showing different functionalities. For example, AtImpα3 is suggested as a vital player in plant innate immunity[34] and AtImpα4 is involved in the *Agrobacterium*-mediated transformation.[35] Likewise, although only three variants have been identified in rice (α1a, α1b, and α2), differential tissue expression and light responsiveness in different isoforms have been reported.[12,15,36] These observations imply that plant Impα variants, analogous to their counterparts from yeast and mammals,[37] can interact with specific binding partners from plants and invading microbes[38]with roles in distinctive cellular activities. Our group has been using structural, biochemical and bioinformatic approaches to characterize the specific recognition between Impα and NLSs in an effort to improve our understanding of nuclear import and the composition of the nuclear proteome.

Here, together with our previous study,[18] we show that two groups of non-classical NLSs, class-3 minor site-specific and class-5 plant-specific NLSs, preferentially bind to the minor NLS-binding site of rImpα1a with nM affinity. The interaction between these NLS peptides and rImpα1aΔIBB falls within the functional affinity limits,[18,20-22] which suggests these non-classical NLSs are able to act as functional NLSs. Although both classes of NLSs bind to the minor NLS-binding site preferentially, the binding conformations of the NLSs are different to each other. The class-5 plant-specific NLSs display an extended conformation in the C-terminal region of the peptide when bound to rImpα1a[18], whileclass-3 minor site-specific NLSs display an α-helical turn in their C-terminal region, stabilized by cation-π interactions with basic residues from mImpα.[19] We show here that that class-3 minor site-specific NLSs are likely to use an analogous binding mode when binding to rImpα1a. Both binding conformations are distinct from the binding of the other monopartite NLSs to Impα proteins characterized structurally to date.[3,23-25,27]

Our structural studies support an increased usage of the minor site in rice Impα, compared to mammalian and yeast proteins. We have previously shown that the proportion of proteins containing plant-specific NLSs is higher in rice compared to any yeast and mammalian proteomes analyzed, and also compared to Arabidopsis.[18] Here, we analysed the distribution of the six classes of NLS sequences in the proteomes from plants, yeast, and mammals. Interestingly, rice proteome not only has a greater prevalence of proteins containing the class-5 NLSs, but also the class-3 NLSs, compared to the other species investigated here. However, the classical (monopartite (class-1 and -2) and bipartite (class-6) NLSs are much more prevalent in all proteomes than the non-classical NLSs (classes 3-5).

The different classes of NLSs employ different binding modes to bind to the NLS-binding sites of Impα. Notably, the preferential binding site for the non-classical NLSs (class-3 and class-5) is the minor NLS-binding site, whereas the class-1 NLSs bind to the major NLS-binding site. Intriguingly, the binding of plant-specific NLSs to the minor NLS-binding site coincides with the additional auto-inhibitory segment found in the minor site of rImpα1a.[18] This may be important in terms of the regulation of the nuclear import cycle, in particular the release of cargo proteins with minor site-specific NLS from Impα in the plant nucleus.

Acknowledgments

We thank Shunichi Kosugi (Iwate Biotechnology Research Center, Kitakami, Japan) forthe rImpα1aΔIBB expression plasmid; and Mary Marfori and members of the Kobe lab for help and discussions. BK is a National Health and Medical Research Council Research Fellow.

References

1. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH. Classical nuclear localization signals: definition, function, and interaction with importin alpha. J Biol Chem 2007; 282:5101-5.
2. Stewart M. Molecular mechanism of the nuclear protein import cycle. Nat Rev Mol Cell Biol 2007; 8:195-208.
3. Marfori M, Mynott A, Ellis JJ, Mehdi AM, Saunders NF, Curmi PM, et al. Molecular basis for specificity of nuclear import and prediction of nuclear localization. Biochim Biophys Acta 2011; 1813:1562-77.
4. Merkle T. Nucleo-cytoplasmic transport of proteins and RNA in plants. Plant Cell Rep 2011; 30:153-76.
5. Merkle T. Nuclear import and export of proteins in plants: a tool for the regulation of signalling. Planta 2001; 213:499-517.
6. Tzfira T, Rhee Y, Chen MH, Kunik T, Citovsky V. Nucleic acid transport in plant-microbe interactions: the molecules that walk through the walls. Annu Rev Microbiol 2000; 54:187-219.
7. Hicks GR, Raikhel NV. Protein import into the nucleus: an integrated view. Annu Rev Cell Dev Biol 1995; 11:155-88.
8. Wirthmueller L, Roth C, Banfield MJ, Wiermer M. Hop-on hop-off: importin-alpha-guided tours to the nucleus in innate immune signaling. Front Plant Sci 2013; 4:149.
9. Deng W, Chen L, Wood DW, Metcalfe T, Liang X, Gordon MP, et al. Agrobacterium VirD2 protein interacts with plant host cyclophilins. Proc Natl Acad Sci U S A 1998; 95:7040-5.
10. Hubner S, Smith HM, Hu W, Chan CK, Rihs HP, Paschal BM, et al. Plant importin alpha binds nuclear localization sequences with high affinity and can mediate nuclear import independent of importin beta. J Biol Chem 1999; 274:22610-7.
11. Smith HM, Hicks GR, Raikhel NV. Importin alpha from Arabidopsis thaliana is a nuclear import receptor that recognizes three classes of import signals. Plant Physiol 1997; 114:411-7.

12. Yamamoto N, Deng XW. Protein nucleocytoplasmic transport and its light regulation in plants. Genes Cells 1999; 4:489-500.
13. Hicks GR, Smith HM, Shieh M, Raikhel NV. Three classes of nuclear import signals bind to plant nuclei. Plant Physiol 1995; 107:1055-8.
14. Lassner MW, Jones A, Daubert S, Comai L. Targeting of T7 RNA polymerase to tobacco nuclei mediated by an SV40 nuclear location signal. Plant Mol Biol 1991; 17:229-34.
15. Jiang CJ, Imamoto N, Matsuki R, Yoneda Y, Yamamoto N. Functional characterization of a plant importin alpha homologue. Nuclear localization signal (NLS)-selective binding and mediation of nuclear import of nls proteins in vitro. J Biol Chem 1998; 273:24083-7.
16. Chelsky D, Ralph R, Jonak G. Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol Cell Biol 1989; 9:2487-92.
17. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, et al. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J Biol Chem 2009; 284:478-85.
18. Chang CW, Counago RL, Williams SJ, Boden M, Kobe B. Crystal structure of rice importin-alpha and structural basis of Its interaction with plant-specific nuclear localization signals. Plant Cell 2012; 24:5074-88.
19. Chang CW, Counago RL, Williams SJ, Boden M, Kobe B. Distinctive conformation of minor site-specific nuclear localization signals bound to importin-α. Traffic 2013.
20. Hu W, Kemp BE, Jans DA. Kinetic properties of nuclear transport conferred by the retinoblastoma (Rb) NLS. J Cell Biochem 2005; 95:782-93.
21. Marfori M, Lonhienne TG, Forwood JK, Kobe B. Structural basis of high-affinity nuclear localization signal interactions with importin-alpha. Traffic 2012; 13:532-48.
22. Hodel MR, Corbett AH, Hodel AE. Dissection of a nuclear localization signal. J Biol Chem 2001; 276:1317-25.

23. Conti E, Kuriyan J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure 2000; 8:329-38.
24. Fontes MR, Teh T, Kobe B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. J Mol Biol 2000; 297:1183-94.
25. Takeda AA, de Barros AC, Chang CW, Kobe B, Fontes MR. Structural basis of importin-alpha-mediated nuclear transport for Ku70 and Ku80. J Mol Biol 2011; 412:226-34.
26. de Barros AC, Takeda AA, Chang CW, Kobe B, Fontes MR. Structural basis of nuclear import of flap endonuclease 1 (FEN1). Acta Crystallogr D Biol Crystallogr 2012; 68:743-50.
27. Lott K, Bhardwaj A, Sims PJ, Cingolani G. A minimal nuclear localization signal (NLS) in human phospholipid scramblase 4 that binds only the minor NLS-binding site of importin alpha1. J Biol Chem 2011; 286:28160-9.
28. Dougherty DA. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science 1996; 271:163-8.
29. Kobe B, Boden M. Computational modelling of linear motif-mediated protein interactions. Curr Top Med Chem 2012; 12:1553-61.
30. Benos PV, Lapedes AS, Stormo GD. Is there a code for protein-DNA recognition? Probab(ilistical)ly. BioEssays : news and reviews in molecular, cellular and developmental biology 2002; 24:466-75.
31. Tewari R, Bailes E, Bunting KA, Coates JC. Armadillo-repeat protein functions: questions for little creatures. Trends Cell Biol 2010; 20:470-81.
32. Mason DA, Stage DE, Goldfarb DS. Evolution of the metazoan-specific importin alpha gene family. J Mol Evol 2009; 68:351-65.
33. Kanneganti TD, Bai X, Tsai CW, Win J, Meulia T, Goodin M, et al. A functional genetic assay for nuclear trafficking in plants. Plant J 2007; 50:149-58.
34. Palma K, Zhang Y, Li X. An importin alpha homolog, MOS6, plays an important role in plant innate immunity. Curr Biol 2005; 15:1129-35.

35. Bhattacharjee S, Lee LY, Oltmanns H, Cao H, Veena, Cuperus J, et al. IMPa-4, an Arabidopsis importin alpha isoform, is preferentially involved in agrobacterium-mediated plant transformation. Plant Cell 2008; 20:2661-80.
36. Shoji K, Iwasaki T, Matsuki R, Miyao M, Yamamoto N. Cloning of a cDNA encoding an importin-alpha and down-regulation of the gene by light in rice leaves. Gene 1998; 212:279-86.
37. Goldfarb DS, Corbett AH, Mason DA, Harreman MT, Adam SA. Importin alpha: a multipurpose nuclear-transport receptor. Trends Cell Biol 2004; 14:505-14.
38. Deslandes L, Rivas S. The plant cell nucleus: a true arena for the fight between plants and pathogens. Plant Signal Behav 2011; 6:42-8.
39. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res 2004; 14:1188-90.

Figure Legends

Figure 1

The α-helical turn prevents the minor site-specific NLSs binding to the major NLS-binding sitein rice rImpα. The structure of the B6:mImpα ΔIBB (PDB ID 3ZIQ[19])complex was superimposed onto the structure of SV40TAgNLS:rImpα1aΔIBB complex(PDB ID 4B8O[18]). (a) The structure of B6 peptide in the minor NLS-binding site (magenta in cartoon representation) from its complex with mImpα ΔIBB (not shown) superimposed onto the structure of rImpα1aΔIBB (in green cartoon and surface representations). (b) The structure of B6 peptide in the conformation as it is found bound to the minor site (magenta in cartoon representation), but superimposed onto the peptide in the major site ofmImpα ΔIBB (not shown) and the structure of rImpα1aΔIBB (in green cartoon and surface representations). There is a steric clash (magnified at the right-hand corner) between the B6 peptide and N-terminal region of rImpα1aΔIBB. The images in (a) and (b) are related by a 90° rotation around the *x*-axis.

Figure 2

Logos for the NLS sequence alignments, and the regular expression patterns of the NLS classes used in this study. Aligned sequences identified by Kosugi and co-workers,[17] including the data from their amino acid replacement analysis were used to derive the regular expression patterns. Classes 6_{10}, 6_{11}, and 6_{12} are class-6 classical bipartite NLSs, but with different linker lengths (10, 11, and 12 residues, respectively). The logos were created by WebLogo 3.3.[39]

Table 1

Consensus sequences of six classes of NLSs.[17]

NLS class	Consensus sequence[a]
Class-1	KR(K/R)R, K(K/R)RK
Class-2	(P/R)*XX*KR(^DE)(K/R)
Class-3	KR*X*(W/F/Y)*XX*AF
Class-4	(R/P)*XX*KR(K/R)(^DE)
Class-5	LGKR(K/R)(W/F/Y)
Class-6	KR$X_{10–12}$K(KR)(KR) or KR$X_{10–12}$K(KR)*X*(K/R)

[a]*X*, any amino-acid; ^D/E, any amino-acid except Asp or Glu.

Table 2

The dissociation constants (K_d; μM) for rImpα1aΔIBB:NLS interactions[a]

	NLS	
	SV40Tag	B6
rImpα1aΔIBB	0.007±0.001	0.023 ±0.004
rImpα1aΔIBB^{D188K} (major-site mutant)	0.68±0.094[b]	0.025 ±0.005
rImpα1aΔIBB^{E388R} (minor-site mutant)	0.047±0.008[b]	0.81 ±0.22

[a]The $K_d$values (presented in μM) were calculated using program GraphPad (Prism). Each assay was performed in triplicate and the values with standard error correspond to the best fit to the one-site specific binding equation [$Y = B_{max}*X/(K_d + X)$, B_{max} is the maxium specific binding with the same unit as Y, $K_d$is the equilibrium binding constant and X is ligand concentration].

[b] values from [18].

Table 3

Distribution of the six classes of NLS sequences in the proteomes from different organisms, using the regular expression approach[a]

		Numbers of proteins						Count of proteins with NLS class						Proportions of NLS class in NLS count (%						Proportion of proteins with NLS class (%)					
	MCC	TP	TN	FP	FN	Total	Nuclea	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
O. sativa	0.071	550	51150	5412	2249	59361	2799	3212	1985	17	1830	36	837	40.57	25.07	0.21	23.11	0.45	10.57	5.41	3.34	0.03	3.08	0.06	1.41
S. vulgare	0.068	358	21911	2594	1541	26404	1899	1680	847	8	836	4	460	43.81	22.09	0.21	21.80	0.10	11.99	6.36	3.21	0.03	3.17	0.02	1.74
A. thaliana	0.167	993	25297	2608	2902	31800	3895	1937	1063	8	943	0	753	41.18	22.60	0.17	20.05	0.00	16.01	6.09	3.34	0.03	2.97	0.00	2.37
V. vinifera	0.11	332	25428	2150	1213	29123	1545	1347	702	9	681	0	418	42.67	22.24	0.29	21.57	0.00	13.24	4.63	2.41	0.03	2.34	0.00	1.44
S. tuberosum	0.093	291	47385	4214	813	52703	1104	2346	1254	7	1163	7	931	41.10	21.97	0.12	20.37	0.12	16.31	4.45	2.38	0.01	2.21	0.01	1.77
S. lycopersicum	0.112	344	30318	2715	1074	34451	1418	1712	913	11	898	5	569	41.67	22.22	0.27	21.86	0.12	13.85	4.97	2.65	0.03	2.61	0.01	1.65
S. cerevisiae	0.239	503	4146	287	1692	6628	2195	412	197	1	184	0	234	40.08	19.16	0.10	17.90	0.00	22.76	6.22	2.97	0.02	2.78	0.00	3.53
H. sapiens	0.194	1866	32751	2774	6180	43571	8046	2473	1396	13	1528	4	998	38.57	19.93	0.20	23.83	0.06	15.56	5.68	3.20	0.03	3.51	0.01	2.29
M. musculus	0.211	1848	23677	2299	5313	33137	7161	2244	1278	15	1371	0	920	38.50	21.93	0.26	23.52	0.00	15.79	6.77	3.86	0.05	4.14	0.00	2.78

[a]Different plant, yeast, and mammalian reference proteomes were taken from UniProt (June, 2013) as representatives of monocot plants (*Oryza sativa, subsp. Japonica, Sorghum vulgare*), dicot plants (*A. thaliana, Vitis vinifera, Solanum tuberosum, Solanum lycopersicum*), yeast (*Saccharomyces cerevisiae*) and mammals (*Homo sapiens, Mus musculus*). MCC: Matthews correlation coefficient; TP and TN: true positive and true negative numbers of proteins; FP and FN: false positive and negative numbers of proteins. PWM scoring thresholds were selected to maximize MCC. Total: number of sequences in each complete proteome. Proteomes are identified by UniProt as "Reference Proteomes" and exclude fragments. Nuclear: the numbers of sequences annotated as nuclear proteins in the Gene Ontology as made available in UniProt. NLS1-5: class-1 to -5 monopartite NLSs.[17] Counts and proportions are based on each full proteome.

Table 4

Distribution of the six classes of NLS sequences in the proteomes from different organisms, using the PWM approach[a]

		Numbers of proteins						Count of proteins with NLS class								Proportions of NLS class in NLS count (%								Proportion of proteins with NLS class (%)							
	MCC	TP	TN	FP	FN	Total	Nuclear	1	2	3	4	5	6_{10}	6_{11}	6_{12}	1	2	3	4	5	6_{10}	6_{11}	6_{12}	1	2	3	4	5	6_{10}	6_{11}	6_{12}
O. sativa	0.043	1522	31561	25001	1277	59361	2799	2493	03108	393	853	46	1489	1684	1741	72.80	9.08	1.15	2.49	0.13	4.35	4.92	5.08	42.00	5.24	0.66	1.44	0.08	2.51	2.84	2.93
S. vulgare	0.071	1058	14198	10307	841	26404	1899	1070	81259	59	409	11	379	444	442	78.10	9.18	0.43	2.98	0.08	2.76	3.24	3.22	40.55	4.77	0.22	1.55	0.04	1.44	1.68	1.67
A. thaliana	0.144	2281	17563	10342	1614	31800	3895	1213	2947	61	443	17	247	311	292	83.96	6.55	0.42	3.07	0.12	1.71	2.15	2.02	38.15	2.98	0.19	1.39	0.05	0.78	0.98	0.92
V. vinifera	0.126	894	18911	8667	651	29123	1545	919	3682	62	325	7	159	143	192	85.41	6.34	0.58	3.02	0.07	1.48	1.33	1.78	31.57	2.34	0.21	1.12	0.02	0.55	0.49	0.66
S. tuberosum	0.088	644	36101	15498	460	52703	1104	1544	81213	84	613	30	270	291	291	84.69	6.65	0.46	3.36	0.16	1.48	1.60	1.60	29.31	2.30	0.16	1.16	0.06	0.51	0.55	0.55
S. lycopersicum	0.121	855	22539	10494	563	34451	1418	1090	1786	81	415	20	197	228	238	84.73	6.11	0.63	3.23	0.16	1.53	1.77	1.85	31.64	2.28	0.24	1.20	0.06	0.57	0.66	0.69
S. cerevisiae	0.191	1100	3082	1351	1095	6628	2195	2383	116	6	89	2	32	46	38	87.87	4.28	0.22	3.28	0.07	1.18	1.70	1.40	35.95	1.75	0.09	1.34	0.03	0.48	0.69	0.57
H. sapiens	0.188	4642	23403	12122	3404	43571	8046	1603	81707	79	640	25	580	547	578	79.42	8.45	0.39	3.17	0.12	2.87	2.71	2.86	36.81	3.92	0.18	1.47	0.06	1.33	1.26	1.33
M. musculus	0.175	4218	16120	9856	2943	33137	7161	1351	71419	84	560	18	474	478	504	79.26	8.32	0.49	3.28	0.11	2.78	2.80	2.96	40.79	4.28	0.25	1.69	0.05	1.43	1.44	1.52

[a] See footnote to Table 3. NLS class 6_{10}, 6_{11} 6_{12}: bipartite class-6 NLSs with linker lengths 10, 11, and 12 residues, respectively.

Figures

Fig. 1

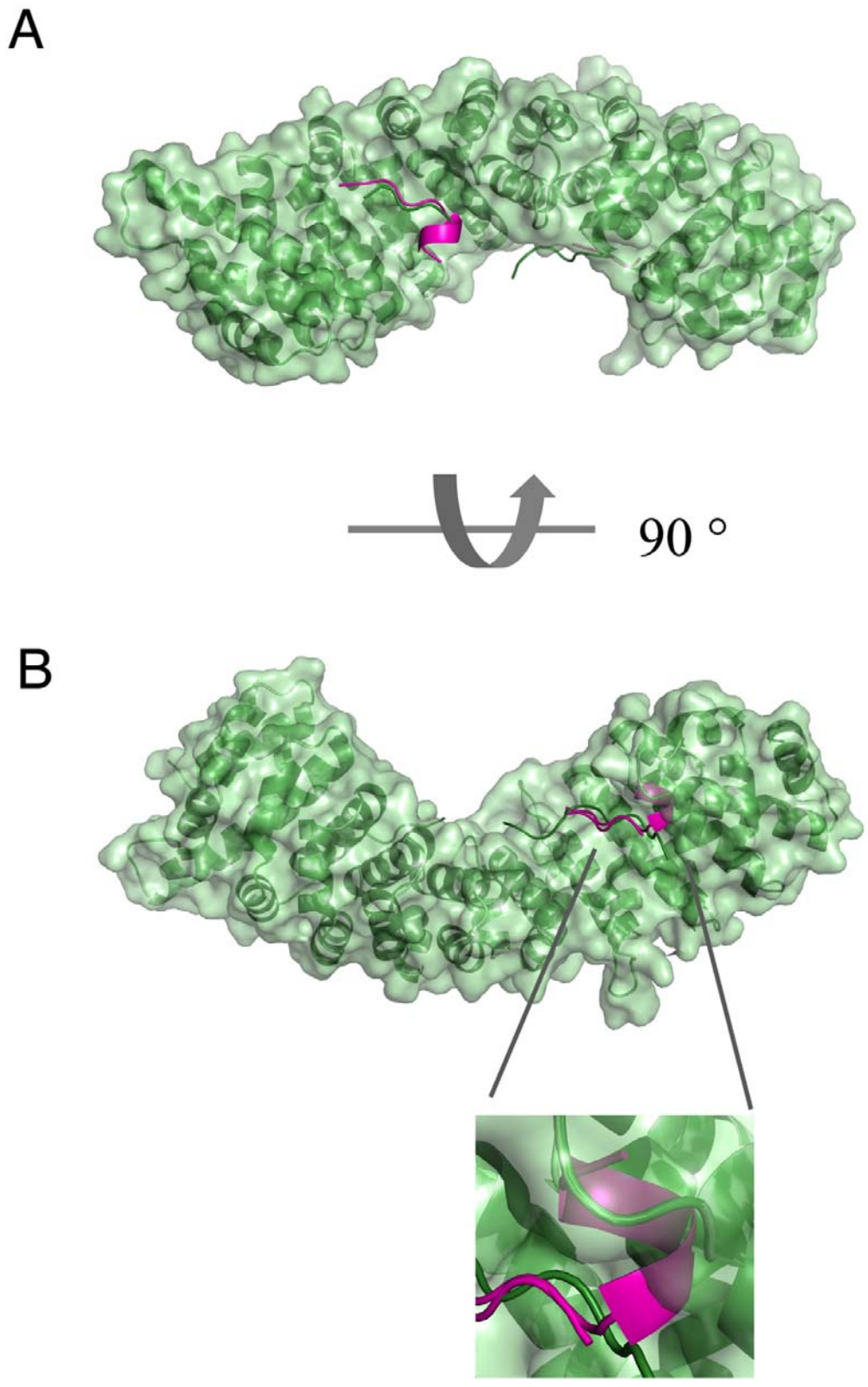
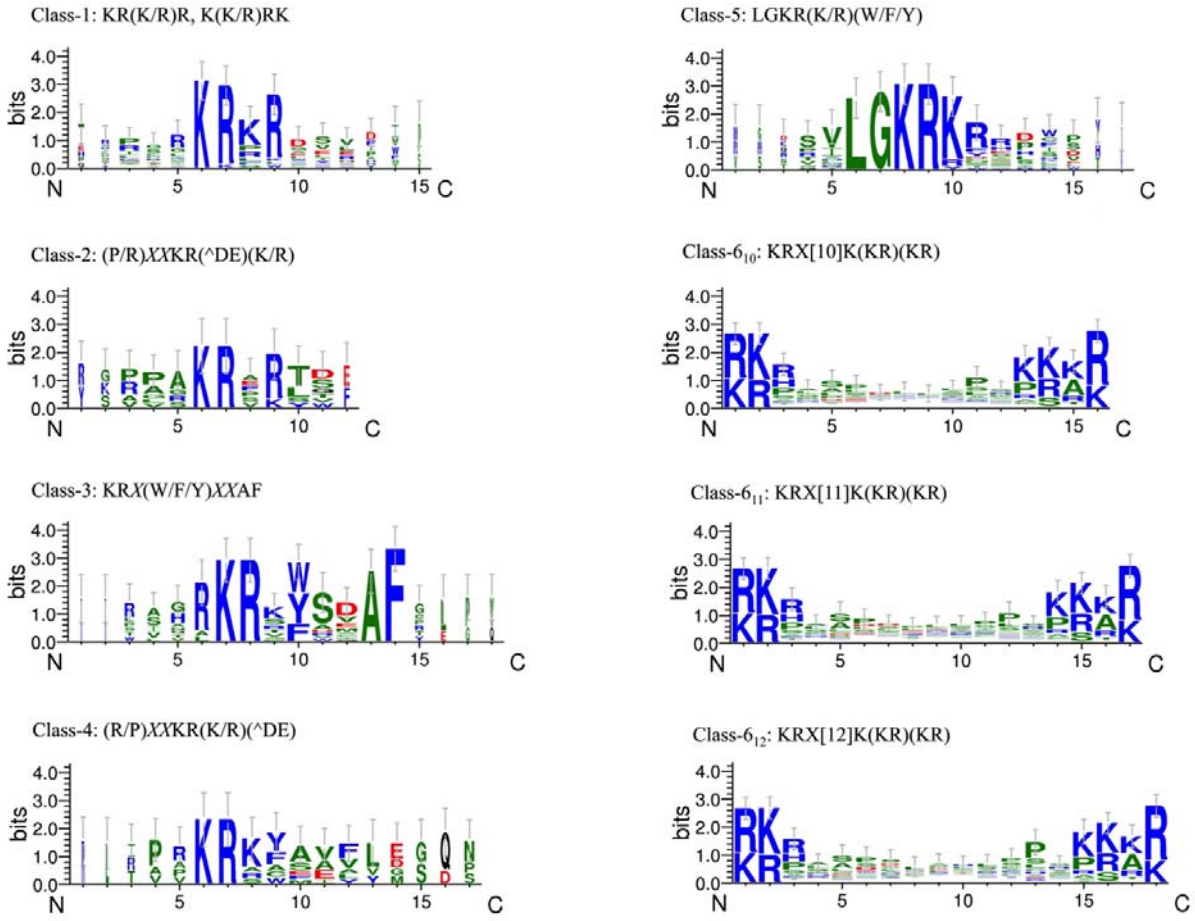


Fig. 2


Class-1: KR(K/R)R, K(K/R)RK
Class-5: LGKR(K/R)(W/F/Y)
Class-2: (P/R)XXKR(^DE)(K/R)
Class-6$_{10}$: KRX[10]K(KR)(KR)
Class-3: KRX(W/F/Y)XXAF
Class-6$_{11}$: KRX[11]K(KR)(KR)
Class-4: (R/P)XXKR(K/R)(^DE)
Class-6$_{12}$: KRX[12]K(KR)(KR)
bits
N
C