Structural basis of nuclear import of flap endonuclease 1 (FEN1)

Andrea C. de Barros,[a] Agnes A. S. Takeda,[a] Chiung-Wen Chang,[b] Boštjan Kobe[b] and Marcos R. M. Fontes[a]*

[a]Departamento de Física e Biofísica, Instituto de Biociências, Universidade Estadual Paulista, Botucatu, SP 18618-970, Brazil, and [b]School of Chemistry and Molecular Biosciences, Institute for Molecular Bioscience and Australian Infectious Diseases Research Centre, University of Queensland, Brisbane, Queensland 4072, Australia

Correspondence e-mail: fontes@ibb.unesp.br

Flap endonuclease 1 (FEN1) is a member of the nuclease family and is structurally conserved from bacteriophages to humans. This protein is involved in multiple DNA-processing pathways, including Okazaki fragment maturation, stalled replication-fork rescue, telomere maintenance, long-patch base-excision repair and apoptotic DNA fragmentation. FEN1 has three functional motifs that are responsible for its nuclease, PCNA-interaction and nuclear localization activities, respectively. It has been shown that the C-terminal nuclear localization sequence (NLS) facilitates nuclear localization of the enzyme during the S phase of the cell cycle and in response to DNA damage. To determine the structural basis of the recognition of FEN1 by the nuclear import receptor importin α, the crystal structure of the complex of importin α with a peptide corresponding to the FEN1 NLS was solved. Structural studies confirmed the binding of the FEN1 NLS as a classical bipartite NLS; however, in contrast to the previously proposed $^{354}\mathrm{KRK}X_8\mathrm{KKK}^{367}$ sequence, it is the $^{354}\mathrm{KR}X_{10}\mathrm{KK\text{-}AK}^{369}$ sequence that binds to importin α. This result explains the incomplete inhibition of localization that was observed on mutating residues $^{365}\mathrm{KKK}^{367}$. Acidic and polar residues in the X_{10} linker region close to the basic clusters play an important role in binding to importin α. These results suggest that the basic residues in the N-terminal basic cluster of bipartite NLSs may play roles that are more critical than those of the many basic residues in the C-terminal basic cluster.

1. Introduction

Flap endonuclease 1 (FEN1) is a member of the 5′ nuclease family and is structurally conserved from bacteriophages to humans (Mase *et al.*, 2011). FEN1 is a 40 kDa protein and contains two metal-ion-binding sites (Hwang *et al.*, 1998; Hosfield *et al.*, 1998). This protein is involved in multiple DNA-processing pathways, including Okazaki fragment maturation, stalled replication-fork rescue, telomere maintenance, long-patch base-excision repair and apoptotic DNA fragmentation (Zheng *et al.*, 2011).

FEN1 has three functional motifs which are responsible for its nuclease activity, interaction with PCNA (proliferating cell nuclear antigen) and nuclear localization, respectively. Several crystal structures of FEN1 have been elucidated from archaea (Mase *et al.*, 2011) and humans (Sakurai *et al.*, 2005; Tsutakawa *et al.*, 2011). FEN1–DNA structural complexes revealed that motifs that are conserved in the nuclease family are responsible for binding and cleaving the DNA (Tsutakawa *et al.*, 2011). The FEN1 construct that was used in the structure determination of the human FEN1–DNA complex was truncated at residue 336 owing to the high flexibility of the C-terminal region.

Nuclear localization sequences (NLSs) were identified in the C-terminal region of FEN1 following the PCNA-binding motif; in the case of human FEN1 the NLS corresponds to residues 354–370 (Qiu *et al.*, 2001). It has been shown that this signal facilitates nuclear localization of the enzyme during the S phase of the cell cycle and in response to DNA damage. Truncation of the NLS motif prevents translocation of the protein from the cytoplasm to the nucleus, but does not affect its nuclease activity. Site-directed mutagenesis experiments demonstrated that the mutation of ^{354}KRK^{356} to alanine residues completely blocked the localization of enzyme into the nucleus, while analogous mutation of ^{365}KKK^{367} only led to a partial block of translocation (Qiu *et al.*, 2001). This NLS has been proposed to involve the classical importin α (Impα)/ importin β (Impβ) mediated nuclear import pathway (Koike *et al.*, 1999, 2001; Bertinato *et al.*, 2001).

The classical NLSs (cNLSs) contain one or two clusters of positively charged amino acids and are therefore usually divided into monopartite cNLSs, which contain a single cluster of basic residues, and bipartite cNLSs, which contain two clusters of basic residues separated by 10–12 variant residues (termed linker residues; Fontes, Teh, Jans *et al.*, 2003). Structural studies have shown that both classes are recognized by the nuclear import receptor Impα (Conti & Kuriyan, 2000; Fontes *et al.*, 2000). This protein has two NLS-binding regions formed by conserved residues in its armadillo (ARM) repeat domain, termed the major and minor NLS-binding sites. The N-terminal (positions P1$'$–P2$'$) and C-terminal (positions P1–P6) clusters of a bipartite NLS interact with the minor and major NLS-binding sites, which correspond to ARM repeats 4–8 and 1–4, respectively, while a monopartite NLS primarily interacts with the major binding site (Conti & Kuriyan, 2000; Fontes *et al.*, 2000; Rexach & Blobel, 1995). Taking the mutagenesis studies into account in combination with nuclear localization experiments on FEN1 led to the hypothesis that the enzyme has a classical bipartite NLS (Qiu *et al.*, 2001).

Nuclear localization is a fundamental control mechanism that allows the cell to regulate DNA replication, DNA repair and many other biological nuclear functions. Although many NLSs have been identified, understanding of the specificity of NLS recognition and regulation remains limited (Marfori *et al.*, 2011). The abolition of nuclear localization of FEN1 resulted in *in vivo* defects in DNA synthesis and in the repair of DNA damage (Qiu *et al.*, 2001), highlighting its key role in regulating physiological nuclear functions.

With the aim of testing the prediction of a bipartite NLS in FEN1 and gaining insight into both the function of FEN1 and the nuclear import process, we solved the crystal structure of the peptide corresponding to the FEN1 NLS bound to Impα. Our results confirmed that the proposed sequence binds to both the major and the minor sites and is consequently a bipartite NLS. Previous studies proposed that the bipartite NLS comprised the sequence KRKX_8KKK, but we found that the sequence instead corresponded to KRX_{10}KKAK. This result may explain why only partial blockage of localization resulted from the mutation of residues ^{365}KKK^{367}; the major-site-binding residues instead correspond to ^{366}KKAK^{369}. We

also show that the FEN1 NLS peptide has an affinity for Impα that is comparable to those of other bipartite NLSs as measured by a solid-phase binding assay. Our study highlights the importance of interactions with Impα involving the NLS linker region and the key role of the minor NLS-binding site.

2. Materials and methods

2.1. Synthesis of NLS peptides

The peptide corresponding to human FEN1 NLS (^{351}SSA-KRKEPEPKGSTKKKAKT^{370}; FEN1NLS) was synthesized by Proteimax (Brazil) with a purity of higher than 99%. The peptides contained additional residues at the N- and C-termini compared with the minimal identified NLS (Knudsen *et al.*, 2009) in order to avoid artifactual binding at the termini (Fontes *et al.*, 2000; Marfori *et al.*, 2012).

2.2. Protein expression and purification

Hexa-His-tagged truncated *Mus musculus* importin α2, comprising amino acids 70–529 (mImp$\alpha\Delta$IBB), was expressed and purified by nickel-affinity chromatography as described previously (Teh *et al.*, 1999). The protein was eluted using an imidazole gradient, followed by dialysis. The Impα sample was stored in a buffer composed of 20 mM Tris–HCl pH 8.0, 100 mM sodium chloride, 10 mM DTT at 253 K. The purity was estimated to be 98% by SDS–PAGE.

2.3. Quantitative binding assay

GST-FEN1NLS fusion protein was expressed in *Escherichia coli* BL21 (DE3) for 6 h at 310 K followed by 18 h at 293 K in auto-induction medium (Studier, 2005). The bacterial pellet was resuspended in GST-*A* buffer (50 mM Tris–HCl, 1 mM DTT, 125 mM NaCl at pH 7.8) and underwent three freeze-thaw cycles before purification. 1 mg ml^{-1} lysozyme, 1 mg DNAase and 1 mM PMSF (Sigma) were added to the cell crude extract and mixed thoroughly. After centrifugation, the soluble fraction was loaded onto a GSTrap column (5 ml; GE Healthcare) and washed with 100 ml GST-*A* buffer. After washing, the GST-FEN1NLS protein was eluted with GST-*B* buffer (50 mM Tris–HCl, 125 mM NaCl, 1 mM DTT, 10 mM reduced glutathione at pH 7.8). The eluted fractions were pooled and concentrated prior to gel-filtration purification using a Superdex 200 (20/60) gel-filtration column (GE Healthcare). Pure GST-FEN1NLS protein was concentrated using an Amicon Ultra centrifugal filter device (10 kDa cutoff).

The solid-phase binding assay was performed essentially as described previously by Takeda *et al.* (2011). The assay was carried out in an Immuno MaxiSorp 96-well plate (Thermo Fisher Scientific). The wells were coated with 50 nM GST-FEN1NLS or GST. Binding reactions were carried out for 2 h at 277 K with 100 µl S-tagged mImp$\alpha\Delta$IBB (0–20 µM in a 2$\times$ serial dilution) in binding buffer per well. After binding and a few steps of washing, the plates were incubated in S-protein-horseradish peroxidase conjugate (Novagen). Horseradish peroxidase substrate (100 µg ml^{-1} 3,3$'$,5,5$'$-tetramethylbenzi-

Table 1
X-ray data-collection and refinement statistics for mImpαΔIBB–FEN1NLS.

Values in parentheses are for the highest resolution shell.

Diffraction data statistics	
Unit-cell parameters (Å)	$a = 78.87$, $b = 89.56$, $c = 100.34$
Space group	$P2_12_12_1$
Resolution (Å)	40.0–2.38 (2.44–2.38)
Unique reflections	28970
Multiplicity	3.0 (2.5)
Completeness (%)	98.0 (95.6)
R_{merge}† (%)	10.5 (68.5)
Average $I/\sigma(I)$	15.0 (2.0)
Refinement statistics	
Resolution (Å)	40.0–2.38 (2.44–2.38)
No. of reflections	28970
R_{cryst}‡ (%)	16.78
R_{free}§ (%)	21.50
No. of non-H atoms	
Protein	3237
Peptide	146
Solvent	157
Mean B factor (Å^2)	48.65
Coordinate error¶ (Å)	0.213
R.m.s.d. from ideal values¶	
Bond lengths (Å)	0.023
Bond angles (°)	1.961
Ramachandran plot††	
Residues in most favoured regions (%)	98
Residues in disallowed regions (%)	0.2

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual measurement of the reflection with Miller indices hkl and $\langle I(hkl)\rangle$ is the mean intensity of this reflection. Calculated for $I > -3\sigma(I)$ (Otwinowski & Minor, 1997). ‡ $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl} |F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure-factor amplitudes, respectively. § R_{free} is equivalent to R_{cryst} but was calculated with reflections (5%) omitted from the refinement process. ¶ Calculated based on the Luzzati plot with the program *SFCHECK* (Laskowski *et al.*, 1996). †† Calculated with the program *PROCHECK* (Laskowski *et al.*, 1996).

dine; Sigma) was added for 20 min at room temperature and the reaction was stopped by adding an equal volume of 0.5 *M* H_2SO_4. The signal was determined at 450 nm with a Molecular Devices plate reader (Spectra Max 250). Average absorbance values at OD_{450} were determined for GST and GST-FEN1NLS at each S-tagged mImpαΔIBB concentration. Background absorbance values (without mImpαΔIBB) were subtracted. The average absorbance values were used to generate binding curves by nonlinear regression using the *GraphPad Prism* software. The apparent dissociation constants (K_d) for the binding of mImpαΔIBB to GST and GST-FEN1 were also calculated using the *GraphPad Prism* software. The curves were plotted against the serial concentration of mImpαΔIBB and were fitted using the equation

$$Y = \frac{B_{max}X}{(K_d + X)}, \quad (1)$$

where X is the concentration of the protein, B_{max} is the maximum specific binding and K_d is the apparent dissociation constant representing the protein concentration yielding half-maximal binding.

2.4. Crystallization and crystal structure determination

mImpαΔIBB was concentrated to 20 mg ml^{-1} using an Amicon 30 kDa cutoff filter unit (Millipore) and was stored at

253 K. Crystallization conditions were screened by systematically altering various parameters, using as a starting point the crystallization conditions that had been successful for other peptide complexes (Fontes, Teh, Jans *et al.*, 2003; Fontes *et al.*, 2000). Crystals were obtained by cocrystallization by combining 1 µl protein solution, 0.5 µl peptide solution (peptide:protein molar ratio of 8:1 for the FEN1NLS peptide) and 1 µl reservoir solution on a cover slip and suspending the mixture over 0.5 ml reservoir solution. Single crystals were obtained using a reservoir solution consisting of 0.60–0.65 *M* sodium citrate pH 5.8, 10 m*M* DTT after 15–20 d.

X-ray diffraction data were collected at a wavelength of 1.45 Å on the MX-2 beamline at the Laboratório Nacional de Luz Síncrotron synchrotron-radiation source (Campinas, Brazil) with a MAR Mosaic 225 imaging-plate detector (MAR Research). The crystals were mounted in nylon loops, transiently soaked in reservoir solution supplemented with 25% glycerol and flash-cooled at 100 K in a nitrogen stream (Oxford Nitrogen Cryojet XL, Oxford Cryosystems). Data were processed using the *HKL*-2000 package (Otwinowski & Minor, 1997). The crystals had the symmetry of space group $P2_12_12_1$ and were isomorphous to those of other mImpαΔIBB–NLS peptide complexes (Table 1). The structure of the complex with CN-SV40TAg NLS (PDB entry 1q1s; Fontes, Teh, Toth *et al.*, 2003) with the NLS peptide omitted was employed as a starting model for crystallographic refinement. After rigid-body refinement using the program *REFMAC*5 (Murshudov *et al.*, 2011), inspection of the electron-density map confirmed the presence of the peptide in both the major and minor NLS-binding sites of the protein. Rounds of crystallographic refinement with *REFMAC*5 (positional and restrained isotropic individual *B* factors with an overall anisotropic temperature factor and bulk-solvent correction) and manual modelling using the program *Coot* (Emsley & Cowtan, 2004) were used to improve the model, considering free *R* factors. The final model of the mImpαΔIBB–FEN1NLS complex consisted of 425 residues of Impα (72–496), one peptide ligand (19 residues could be modelled) and 157 water molecules (Table 1). Structure quality was checked using the program *PROCHECK* (Laskowski *et al.*, 1996) and the contacts were analyzed using the program *LIGPLOT* (Wallace *et al.*, 1995).

3. Results

3.1. Structure of importin *a* complexed with a peptide corresponding to FEN1 NLS

A peptide corresponding to human FEN1 NLS (FEN1NLS) was cocrystallized with N-terminally truncated mouse Impα lacking residues 1–69 (mImpαΔIBB); the truncated residues are responsible for its autoinhibition (Kobe, 1999). Mouse Impα is a suitable model for the human protein as their highly interacting residues with NLS peptides are strictly conserved (including the Trp-Asn array and negatively charged residues of the major and minor binding sites and Tyr277 and Arg315 of the linker region; Fontes, Teh, Jans *et al.*, 2003; Marfori *et al.*, 2011).

The cocrystals were grown under similar conditions to and were isomorphous to other mouse Impα crystals (Fontes, Teh, Jans *et al.*, 2003; Fontes *et al.*, 2000). Electron-density maps based on the Impα model clearly showed electron density corresponding to the peptide following rigid-body refinement (Fig. 1). The structure was refined at 2.38 Å resolution (Table 1).

Impα is an elongated protein composed of ten ARM motifs displayed in tandem (Kobe, 1999; Catimel *et al.*, 2001), each containing three α-helices (H1, H2 and H3; Fontes *et al.*, 2000). The NLS-binding sites are located in a concave groove on the surface of Impα. The major binding site is mainly formed by the H3 helices of ARM repeats 2–4, while the minor site is located at ARM repeats 7–8. The structure of Impα in the FEN1NLS complex is essentially identical to those of full-length Impα and of Impα in previously reported complexes with bipartite NLS-like peptides. The root-mean-square deviation (r.m.s.d.) of the C^α atoms of Impα residues 72–496 is 0.39 Å between full-length Impα and the Impα–FEN1NLS complex.

3.2. Binding of FEN1NLS to importin α

The peptide corresponding to the FEN1 NLS, [351]SSAKR-KEPEPKGSTKKKAKT[370], binds to both the minor and the major sites of Impα, with the main chain positioned in an antiparallel configuration when compared with the direction of the ARM repeats. Residues 352–370 of the FEN1NLS peptide (residue 351 had no interpretable electron density) could be unambiguously identified in the electron-density

maps (Fig. 1a). The buried surface between the protein and the peptide was 1334.7 $Å^2$.

The average *B* factor of the FEN1NLS peptide (50.5 $Å^2$) was similar to the average *B* factor of the entire structure (49.5 $Å^2$). The major-site-binding residues (residues 365–369, positions P1–P5) presented lower average *B* factors (43.5 $Å^2$) compared with the whole peptide, and residues Lys366 and Ala368 (positions P2 and P4) had the lowest *B* factors (39.2 and 41.0 $Å^2$, respectively) and the largest number of interactions between the peptide and the conserved residues of Impα. The residues bound to the minor site (residues 354–357, positions P1′–P4′) had average *B* factors (51.5 $Å^2$) that were slightly higher than those of the peptide; Lys355 (position P2′) had the lowest *B* factor in this region (46.40 $Å^2$), reflecting the strong interaction of this residue with the protein. The linker sequence connecting the residues bound to the major and the minor sites (residues 356–365, positions P3′–P1) also showed average *B* factors that were slightly higher (51.44 $Å^2$) than the average *B* factor of the entire FEN1NLS peptide; residues Lys356, Gly362, Ser363, Thr364 and Lys365 had the lowest *B* factors (lower than 48 $Å^2$). Because most peptide side chains interact with the protein, they had average *B* factors in the range 40–54 $Å^2$ (only two residues had *B* factors higher than 54 $Å^2$). These relatively similar *B*-factor values along the entire peptide are not an unusual feature compared with other bipartite NLSs (Fontes, Teh, Jans *et al.*, 2003; Fontes *et al.*, 2000).

Comparison of FEN1NLS with other bipartite NLSs (nucleoplasmin, RB and N1N2; Fontes, Teh, Jans *et al.*, 2003; Fontes *et al.*, 2000) showed high structural similarity in both the major and the minor binding sites. The r.m.s.d.s of C^α atoms of NLS residues in positions P1–P5 are 0.36, 0.32 and 0.18 Å, respectively, and the r.m.s.d.s of the residues in positions P1′–P3′ are 0.50, 0.09 and 0.10 Å, respectively, compared with nucleoplasmin, RB and N1N2. However, the residues of the NLS peptides in the linker region presented very different conformations in the different peptides. Comparison of the C^α atoms of the linker residues of FEN1NLS (linker positions P3′–P1) with the nucleoplasmin, RB and N1N2 NLSs resulted in r.m.s.d.s higher than 2.5 Å (2.57, 2.55 and 2.57 Å, respectively; Fig. 2).
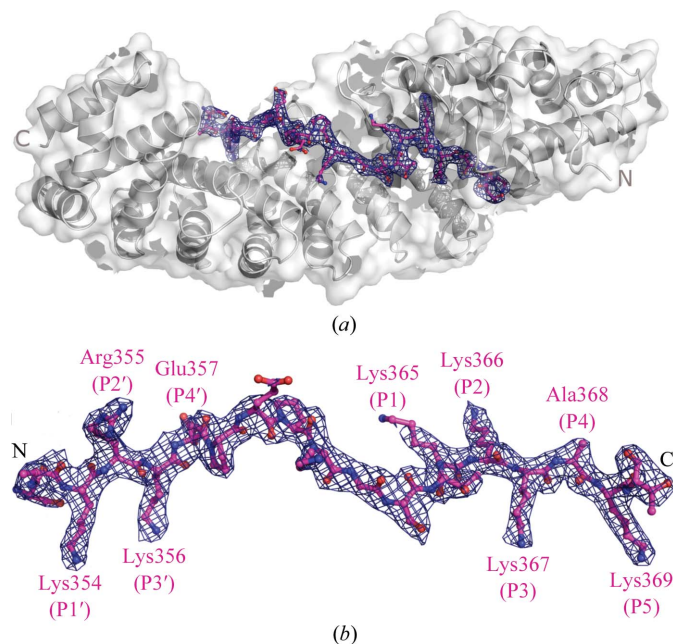

Figure 1
Crystal structure of the Impα–FEN1NLS complex. (*a*) Overall structure of the Impα–FEN1NLS complex. Impα is shown as a ribbon diagram. FEN1NLS (pink) is shown in stick representation. (*b*) Electron-density map (coefficients $3|F_{obs}| - 2|F_{calc}|$) of the Impα–FEN1NLS complex in the area corresponding to the peptide (contoured at 1.5 s.d.). All peptide residues were omitted from the model and simulated annealing was run with a starting temperature of 1000 K.
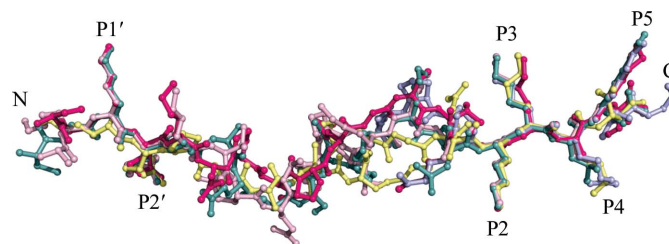

Figure 2
Comparison of NLS peptide binding to the minor and major NLS-binding sites. FEN1 (dark pink), N1N2 (light pink; PDB entry 1pjn; Fontes, Teh, Jans *et al.*, 2003), nucleoplasmin (yellow; PDB entry 1ejy; Fontes *et al.*, 2000), RB (green; PDB entry 1pjm) and CN-SV40TAg (light blue; PDB entry 1q1t; Fontes, Teh, Jans *et al.*, 2003) NLSs were superimposed using the C^α atoms of the peptides. Positions binding to the major (P1–P5) and minor binding sites (P1′–P4′) are identified along the chains.

Despite the structural differences in the linker regions, conserved residues of Impα have been observed to interact with different bipartite NLSs in this region. The strictly conserved residues Arg238, Tyr277 (both in ARM repeat 5) and Arg315 (ARM repeat 6) of Impα interact with the main chains or side chains of the residues of all bipartite peptide linker regions (FEN1, nucleoplasmin, RB and N1N2 NLSs). In the case of FEN1NLS these three residues interact with the protein main chain only via hydrogen bonds or hydrophobic interactions (Fig. 3). These interactions with the peptide main chain are likely to be responsible for its high stability, conferring low B-factor values and high-quality electron density compared with other bipartite NLSs.

In addition, the majority of FEN1 linker residues make interactions with Impα (except for residues Glu357 and Pro358; Fig. 3). Again, this is an unusual feature compared with other bipartite NLS structures solved to date.

3.3. Affinity of FEN1NLS binding to importin α

A solid-phase binding assay was used to measure the affinity of FEN1NLS for Impα (Fig. 4). The apparent dissociation constant (K_d) of FEN1NLS for mImpαΔIBB corresponded to 42 ± 6.5 nM. This value is comparable with an optimal monopartite NLS peptide binding to mImpαΔIBB (53 ± 15 nM; Yang et al., 2010) and is very similar to the K_d obtained for the 'extended' monopartite Ku70 and CN-SV40TAg NLS peptides binding to Impα (29 ± 4 and 24 ± 8 nM, respectively; Takeda et al., 2011; Fontes, Teh, Toth et al., 2003); these NLSs make critical contacts distinct from those of bipartite NLSs in the region N-terminal to the basic cluster (Takeda et al., 2011). All these affinities were measured using techniques similar to that used here. The K_d values for the binding of the N1N2 and RB bipartite NLSs were also similar to those for FEN1NLS

(5.4 and 45 nM for N1N2 and RB, respectively; Efthymiadis et al., 1997; Hu & Jans, 1999).

4. Discussion

4.1. FEN1NLS is a classical bipartite NLS

The two basic clusters at positions 354–356 and 365–367 (KRKX_8KKK) have previously been proposed to be responsible for translocation of FEN1 to cell nuclei (Qiu et al., 2001) by the classical Impα/Impβ-mediated nuclear import pathway (Koike et al., 1999, 2001; Bertinato et al., 2001). In our crystal structure of the FEN1NLS–Impα complex the FEN1NLS peptide binds to Impα with Lys354-Arg355-Lys356 residues at positions P1′-P2′-P3′ (minor site) and Lys365-Lys366-Lys367 at positions P1-P2-P3 (major site). This further places Lys369 at the P5 position. This is in agreement with the bipartite NLS consensus sequence KRX_{10-12}KK/RXK (Fontes, Teh, Jans et al., 2003; Görlich et al., 1994).

The affinity of FEN1NLS in the low-nanomolar range is comparable with those of other bipartite NLSs and extended monopartite NLSs. In contrast, shorter monopartite NLSs comprising only the basic cluster have affinities in the high-nanomolar range [for example, the extended SV40TAg (CN-SV40Tag) and Ku70 NLSs have fourfold higher affinities compared with the shorter SV40TAg and Ku80 NLSs, respectively (Fontes, Teh, Toth et al., 2003; Takeda et al., 2011)].

4.2. Comparison of bipartite NLSs and the role of their linker regions

The structures of three other regular-affinity bipartite NLS–Impα complexes have been solved to date (nucleoplasmin, RB and N1N2; Fontes, Teh, Jans et al., 2003). A detailed comparison of these structures with FEN1NLS–Impα sheds light on the specificity of NLS binding. All NLS peptides present very similar conformations in the minor and major sites, but adopt very different conformations in the linker region (between positions P3′ and P1).

Impα uses a Trp-Asn array to bind the basic clusters in the minor and major NLS-binding sites (Fontes et al., 2000). Analogously, the key feature responsible for binding the linker region involves the strictly conserved residues Arg315 (ARM repeat 6) and Tyr277 (ARM repeat 5) that interrupt the regularity of the Trp-Asn array. These residues bind to the main chain (or the side chain for RB and N1N2 NLSs) of all NLS peptides and
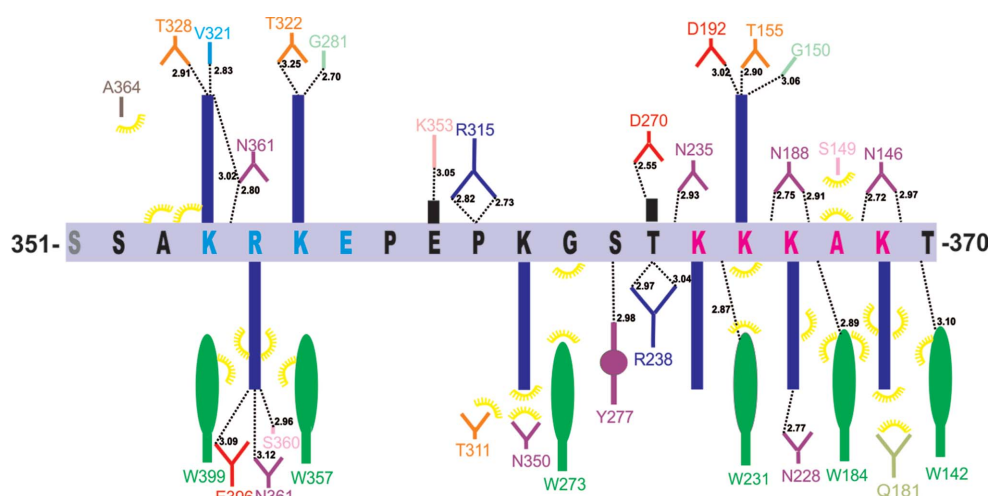
Figure 3
Schematic diagram of the interactions between FEN1NLS and Impα. The peptide backbone is drawn in light blue, with the residues identified by their one-letter codes (residues in blue and pink represent the minor and major binding sites, respectively). FEN1NLS peptide side chains interacting with the protein are represented in dark blue or black. Impα side-chain residues interacting with the peptide are indicated by their names and different colours. Polar contacts are shown as dashed lines and hydrophobic contacts are indicated by arcs with radiating spokes.

appear to be important for bipartite NLS binding independent of the linker length (ten, 11 or 12 residues). Additionally, the conserved residue Arg238 (ARM repeat 4) always binds to the main chain of all NLS peptides. These residues are conserved in all Impα proteins from different species (Fontes, Teh, Jans *et al.*, 2003).

Another important feature of bipartite NLSs (Fontes, Teh, Jans *et al.*, 2003) and extended monopartite NLSs (Fontes, Teh, Toth *et al.*, 2003; Cutress *et al.*, 2008; Takeda *et al.*, 2011) is the region just preceding position P1. In the CN-SV40TAg, Ku70 and human androgen receptor monopartite NLS peptides additional residues N-terminal to the major-site basic cluster form favourable interactions with Impα which increase their affinity for the receptor compared with shorter basic cluster sequences (Fontes *et al.*, 2000; Takeda *et al.*, 2011; Cutress *et al.*, 2008; Hübner *et al.*, 1997; Miyamoto *et al.*, 1997). FEN1NLS Ser364 and Thr365 (two residues preceding the P1 position) form several hydrogen bonds with Impα. Lys and Asn residues preceding the P1 position similarly interact with Impα in RB and N1N2 NLSs. Low values of crystallographic B factors and the quality of the electron-density maps further suggest that the C-terminal portion of the linker (closest to the major site) plays an important role in binding to Impα. Similarly, the N-terminal portion of the linker (closest to the minor site) also makes favourable interactions with Impα. Charged and polar residues in positions P3' and P4' of the FEN1 and N1N2 NLS peptides interact with Thr322 and Asn280 of Impα, respectively.

Differences in the length of the linker region (defined here as between positions P2' and P2) in nucleoplasmin and FEN1 (ten residues), RB (11 residues) and N1N2 (12 residues) NLSs and the presence of specific residues (charged, polar and proline residues) seem to be the main reasons for the different conformations of the linkers. The ten-residue linker from FEN1NLSs has the best-defined electron-density maps and

Figure 4
Quantitative binding assays. The affinity of Impα for FEN1NLS peptide was determined using a solid-phase binding assay. The measurements were performed in duplicate in an mImpαΔIBB concentration range of 0–50 μM. The average absorbance value (OD$_{450\,nm}$) of individual concentration points (μM) was plotted against the log of the concentration of mImpαΔIBB. The curve was fitted using the one-site binding equation, yielding a K_d of 42 ± 6.5 nM. This value corresponds to the high-affinity binding mode (further low-affinity binding modes may be present).

the largest number of favourable contacts and appears to be more favourable for Impα binding than longer linkers. In the N1N2 NLS two linker side-chain residues could not be modelled owing to a lack of electron density (Fontes, Teh, Jans *et al.*, 2003). The N1N2 NLS linker requires some short turns and forces the chain away from Impα, precluding some favourable interactions. The linker from RB NLS has an intermediate quality compared with the FEN1 and N1N2 NLSs. In contrast to these results, the ten-residue linker of the nucleoplasmin NLS structure presents a poorly ordered linker region, particularly in its central region. This could be a consequence of the lower resolution of the structure (2.9 Å) compared with the other bipartite NLSs and the absence of acidic and proline residues in the linker region. In summary, we suggest that shorter linkers appear to be more favourable because their extended conformations allow the largest number of favourable contacts with the receptor.

4.3. The relative importance of minor-site and major-site binding

Mutagenesis studies combined with nuclear localization assays of FEN1 suggested that this enzyme contains a classical bipartite NLS. Mutation of [354]KRK[356] to alanines completely blocked the localization of the enzyme into the nucleus, while an analogous experiment for the cluster [365]KKK[367] led to only partial blockage of translocation to the nucleus (Qiu *et al.*, 2001). This result appeared to be paradoxical because binding of the C-terminal cluster to the major NLS-binding site is usually considered to make a more significant contribution energetically (Hodel *et al.*, 2001). This result can now be explained by the FEN1NLS–Impα structure reported here. The structure shows that the mutated residues corresponded to the P1–P3 positions and that the P5 Lys residue remained intact in the described mutagenesis experiment. The optimal consensus in the C-terminal major-site-binding basic cluster has been proposed to be K(K/R)X(K/R) (for positions P2–P5; Dingwall & Laskey 1991); however, other studies with different NLS sequences have demonstrated that not all of these positions must be occupied by Lys or Arg residues. A number of studies established that Lys in position P2 abolished nuclear import, whereas the mutation of one of the other basic residues at P1, P3, P4 or P5 had a smaller effect (Colledge *et al.*, 1986). A structural study with a nonclassical NLS (phospholipid scramblase 1 NLS) demonstrated that just two Lys residues at positions P2 and P5 were sufficient for nuclear translocation of the peptide (Chen *et al.*, 2005) and *in vivo* experiments with the same NLS peptide showed that a Lys at position P2 was dispensable (Chen *et al.*, 2005). In addition, a recent study involving the structure of phospholipid scramblase 4 NLS showed that this peptide bound exclusively to the minor binding site (Lott *et al.*, 2011), suggesting that the minor site does not merely play a secondary role. Consistent with this suggestion, mutagenesis of nucleoplasmin (Fontes, Teh, Jans *et al.*, 2003), RB (Rexach & Blobel, 1995; Fontes, Teh, Jans ,*et al.*, 2003) and N1N2 (Hu & Jans, 1999; Fontes, Teh, Jans *et al.*, 2003; Fontes, Teh, Toth *et al.*,
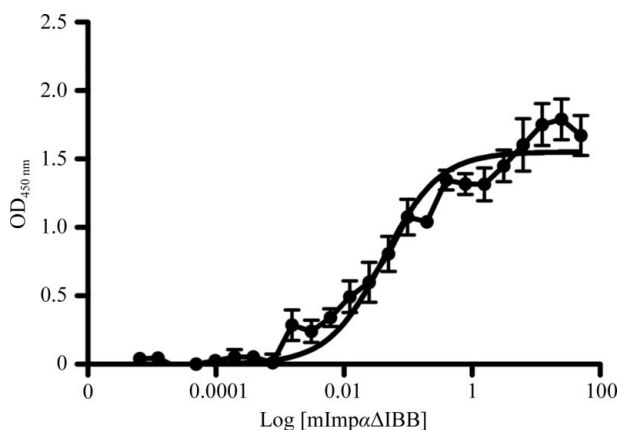
Figure 5

Protein				P1′	P2′	P3′	P4′	Linker	P1	P2	P3	P4	P5	
FEN1	S	S	A	K	R	K	E	P E P K G S T	K	K	K	A	K	T
Ku70								N E G S G	S	K	R	P	K	
SV40				K	K	R	K	V A A P P	K	K	K	R	K	V E
CN-SV40								D A Q H A A P	P	K	K	K	R	
AR								E A G M T L G A	R	K	L	K	K	L
PLSCR4	G	S	I	I	R	K	W	N G L						
Nucleop1		A	V	K	R	P	A	A T K K A G Q	A	K	K	K	K	L
RB				K	R	S	A	E G S N P P K P	L	K	K	L	R	G
N1N2		R	K	K	R	K	T	E E E S P L K D K	A	K	K	S	K	G
Consensus				K	R			$X_{10\text{–}12}$		K	K/R	X	K/R	

Binding to specific binding pockets of Impα based on structural data. NLSs are aligned as observed to bind to the NLS-binding sites (P1′–P4′, minor binding site; P1–P6, major binding site, as defined in Conti & Kuriyan, 2000). The consensus sequences for monopartite and bipartite NLSs were defined by Chelsky et al. (1989) and Dingwall & Laskey (1991), respectively. Ku70, Ku70 protein (Takeda et al., 2011); SV40, simian virus antigen T (Fontes et al., 2000); CN-SV40, simian virus antigen T phosphorylated on residue Ser112 (Fontes, Teh, Toth et al., 2003); AR, androgen receptor (Cutress et al., 2008); PLSCR4, phospholipid scramblase 4 (Lott et al., 2011); Nucleop1, nucleoplasmin (Fontes et al., 2000); RB, retinoblastoma protein (Fontes, Teh, Jans et al., 2003); N1N2, Xenopus laevis phosphoprotein N1N2 (Fontes, Teh, Jans et al., 2003).

2003) NLSs revealed that the substitution of the P1′ and P2′ residues by amino acids other than Arg or Lys abolished nuclear localization. In the light of the available data, our structure suggests that in the context of FEN1NLS the minor-site binding is critical and mutation of the residues in positions P1–P3 in the C-terminal cluster, while retaining the lysine in the P5 position, cannot completely abolish binding to Impα and nuclear import.

4.4. The importance of the linker region for NLS binding

The linker region is variable in both length and amino-acid composition, with this variability probably being relevant to the specificity of cargo-protein binding to Impα from different organisms and different Impα isoforms. Highly reactive inhibitors have been designed for both yeast and mammalian Impα by systematic replacement of each residue in a 12-residue linker bipartite NLS (Kosugi et al., 2008). The authors were able to propose two optimized bipartite NLSs which had a high content of acidic and proline residues in the central linker region. However, different linker lengths were not tested in this experiment. Kosugi et al. (2009) recently proposed six different Impα-specific NLS classes and found specific patterns in the bipartite NLS linker region which also included the prevalence of prolines and acidic residues and the lack of hydrophobic residues. These results are in agreement with the polar interactions observed in the structures of the FEN1 and N1N2 NLSs (Fig. 5). The presence of polar and charged residues may be responsible for the fivefold higher affinity for Impα of N1N2 NLS compared with RB NLS, despite N1N2 having a longer linker that lacks some contacts with Impα (Fontes, Teh, Jans et al., 2003). FEN1 NLS also has polar and charged residues adjacent to the basic clusters and in the central linker region which interact with the receptor, has two proline residues that provide rigidity and has an optimal ten-residue linker length. The result of the combination of these factors is that the peptide can make favourable interactions all along its length, resulting in a well ordered structure and excellent-quality electron density throughout the peptide compared with other bipartite NLSs with available structures (nucleoplasmin, N1N2 and RB). We propose that the linker region is not only a region that links the major and minor binding sites but also contributes significantly to specificity of transport receptor binding.

5. Conclusions

We confirmed that the FEN1 NLS peptide can bind to Impα as a classical bipartite NLS but using different residues to those proposed previously. We also reaffirmed the importance of Impα residues Arg315, Tyr277 and Arg238 in bipartite NLS binding and additionally identified polar residues in the NLS linker region adjacent to the major and minor NLS-binding sites that contribute to the specificity of bipartite NLS binding. Lys and Arg residues in positions P1′ and P2′ may play a more critical role than some of the basic residues that bind to the major NLS-binding site in FEN1 NLS and bipartite NLSs in general.

References

Bertinato, J., Schild-Poulter, C. & Haché, R. J. (2001). *J. Cell Sci.* **114**, 89–99.

Catimel, B., Teh, T., Fontes, M. R., Jennings, I. G., Jans, D. A., Howlett, G. J., Nice, E. C. & Kobe, B. (2001). *J. Biol. Chem.* **276**, 34189–34198.

Chelsky, D., Ralph, R. & Jonak, G. (1989). *Mol. Cell. Biol.* **9**, 2487–2492.

Chen, M.-H., Ben-Efraim, I., Mitrousis, G., Walker-Kopp, N., Sims, P. J. & Cingolani, G. (2005). *J. Biol. Chem.* **280**, 10599–10606.

Colledge, W. H., Richardson, W. D., Edge, M. D. & Smith, A. E. (1986). *Mol. Cell. Biol.* **6**, 4136–4139.

Conti, E. & Kuriyan, J. (2000). *Structure*, **8**, 329–338.

Cutress, M. L., Whitaker, H. C., Mills, I. G., Stewart, M. & Neal, D. E. (2008). *J. Cell Sci.* **121**, 957–968.

Dingwall, C. & Laskey, R. A. (1991). *Trends Biochem. Sci.* **16**, 478–481.

Efthymiadis, A., Shao, H., Hübner, S. & Jans, D. A. (1997). *J. Biol. Chem.* **272**, 22134–22139.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fontes, M. R., Teh, T., Jans, D., Brinkworth, R. I. & Kobe, B. (2003). *J. Biol. Chem.* **278**, 27981–27987.

Fontes, M. R., Teh, T. & Kobe, B. (2000). *J. Mol. Biol.* **297**, 1183–1194.

Fontes, M. R., Teh, T., Toth, G., John, A., Pavo, I., Jans, D. A. & Kobe, B. (2003). *Biochem. J.* **375**, 339–349.

Görlich, D., Prehn, S., Laskey, R. A. & Hartmann, E. (1994). *Cell*, **79**, 767–778.

Hodel, M. R., Corbett, A. H. & Hodel, A. E. (2001). *J. Biol. Chem.* **276**, 1317–1325.

Hosfield, D. J., Frank, G., Weng, Y., Tainer, J. A. & Shen, B. (1998). *J. Biol. Chem.* **273**, 27154–27161.

Hu, W. & Jans, D. A. (1999). *J. Biol. Chem.* **274**, 15820–15827.

Hübner, S., Xiao, C.-Y. & Jans, D. A. (1997). *J. Biol. Chem.* **272**, 17191–17195.

Hwang, K. Y., Baek, K., Kim, H.-Y. & Cho, Y. (1998). *Nature Struct. Biol.* **5**, 707–713.

Knudsen, N. Ø., Andersen, S. D., Lützen, A., Nielsen, F. C. & Rasmussen, L. J. (2009). *DNA Repair*, **8**, 682–689.

Kobe, B. (1999). *Nature Struct. Biol.* **6**, 388–397.

Koike, M., Ikuta, T., Miyasaka, T. & Shiomi, T. (1999). *Exp. Cell Res.* **250**, 401–413.

Koike, M., Shiomi, T. & Koike, A. (2001). *J. Biol. Chem.* **276**, 11167–11173.

Kosugi, S., Hasebe, M., Entani, T., Takayama, S., Tomita, M. & Yanagawa, H. (2008). *Chem. Biol.* **15**, 940–949.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M. & Yanagawa, H. (2009). *J. Biol. Chem.* **284**, 478–485.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). *J. Biomol. NMR*, **8**, 477–486.

Lott, K., Bhardwaj, A., Sims, P. J. & Cingolani, G. (2011). *J. Biol. Chem.* **286**, 28160–28169.

Marfori M., Lonhienne T. G., Forwood, J. K. & Kobe, B. (2012). *Traffic*, doi: 10.1111/j.1600-0854.2012.01329.x.

Marfori, M., Mynott, A., Ellis, J. J., Mehdi, A. M., Saunders, N. F., Curmi, P. M., Forwood, J. K., Bodén, M. & Kobe, B. (2011). *Biochim. Biophys. Acta*, **1813**, 1562–1577.

Mase, T., Kubota, K., Miyazono, K., Kawarabayasi, Y. & Tanokura, M. (2011). *Acta Cryst.* F**67**, 209–213.

Miyamoto, Y., Imamoto, N., Sekimoto, T., Tachibana, T., Seki, T., Tada, S., Enomoto, T. & Yoneda, Y. (1997). *J. Biol. Chem.* **272**, 26375–26381.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Qiu, J., Li, X., Frank, G. & Shen, B. (2001). *J. Biol. Chem.* **276**, 4901–4908.

Rexach, M. & Blobel, G. (1995). *Cell*, **83**, 683–692.

Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H. & Hakoshima, T. (2005). *EMBO J.* **24**, 683–693.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Takeda, A. A. S., de Barros, A. C., Chang, C.-W., Kobe, B. & Fontes, M. R. M. (2011). *J. Mol. Biol.* **412**, 226–234.

Teh, T., Tiganis, T. & Kobe, B. (1999). *Acta Cryst.* D**55**, 561–563.

Tsutakawa, S. E., Classen, S., Chapados, B. R., Arvai, A. S., Finger, L. D., Guenther, G., Tomlinson, C. G., Thompson, P., Sarker, A. H., Shen, B., Cooper, P. K., Grasby, J. A. & Tainer, J. A. (2011). *Cell*, **145**, 198–211.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 127–134.

Yang, S. N. Y., Takeda, A. A. S., Fontes, M. R. M., Harris, J. M., Jans, D. A. & Kobe, B. (2010). *J. Biol. Chem.* **285**, 19935–19946.

Zheng, L., Jia, J., Finger, L. D., Guo, Z., Zer, C. & Shen, B. (2011). *Nucleic Acids Res.* **39**, 781–794.