

# SCIENTIFIC REPORTS



OPEN

## Structural Basis for Importin- $\alpha$ Binding of the Human Immunodeficiency Virus Tat

K. M. Smith, Z. Himiari, S. Tsimbalyuk &amp; J. K. Forwood

HIV-1 has caused 35 million deaths globally, and approximately the same number is currently living with HIV-1. The trans-activator of transcription (Tat) protein of HIV-1 plays an important regulatory function in the virus life cycle, responsible for regulating the reverse transcription of the viral genome RNA. Tat is found in the nucleus of infected cells, but can also invade uninfected neighbouring cells. Regions within Tat responsible for these cellular localisations are overlapping and include a nuclear localisation signal (NLS) spanning <sup>48</sup>GRKKRR, and a cell penetrating peptide (CPP) signal spanning <sup>48</sup>GRKKRRQRRRAPQN. However, the mechanism by which this NLS/CPP region mediates interaction with the nuclear import receptors remains to be resolved structurally. Here, we establish that the HIV-1 Tat:NLS/CPP is able to form a stable and direct interaction with the classical nuclear import receptor importin- $\alpha$  and using x-ray crystallography, we have determined the molecular interface and binding determinants to a resolution of 2.0 Å. We show for the first time that the interface is the same as host factors such as Ku70 and Ku80, rather than other virus proteins such as Ebola VP24 that bind on the outer surface of importin- $\alpha$ .

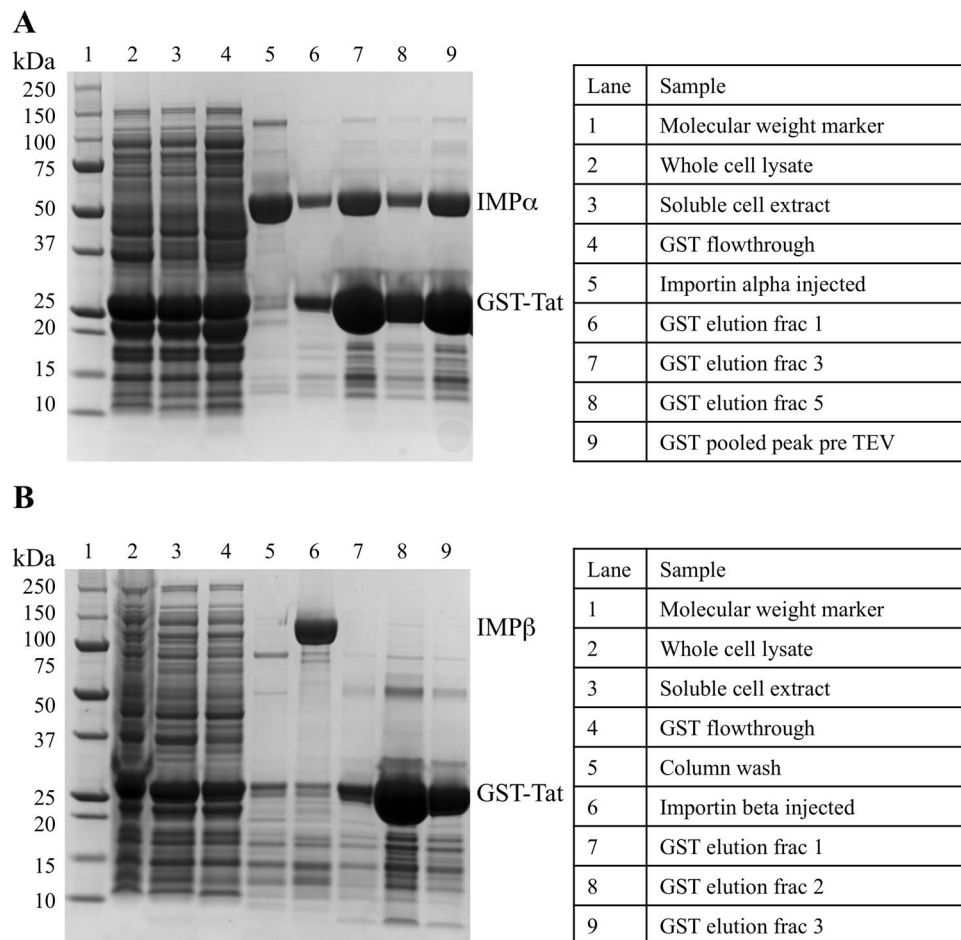
The HIV-1 virus has spread worldwide, infecting 60 million people, and causing more than 25 million deaths. More than 30 million people currently live with the disease<sup>1</sup>, but despite highly active antiretroviral therapy (HAART) reducing the effects of the virus, these antivirals do not clear the virus from infected patients. HIV-1 encodes three groups of proteins that are common in all retroviruses. The gag polyprotein, pol polyprotein and gp160 precursors are structural proteins that form the outer shell of the virus particle, and are processed to produce proteins for the virion interior. The accessory regulatory proteins, Vif, Vpr, Vpu and Nef, interact with cellular ligands and function as adapter molecules or to inhibit normal host function. The third group are the essential regulatory elements, Tat and Rev. The primary role of Tat is in regulating the reverse transcription of viral genome RNA, whilst Rev is responsible for the synthesis of major viral proteins for viral replication<sup>2</sup>.

Tat is a transcriptional trans-activator and plays an important role during HIV-1 replication by binding to a short-stem loop structure, known as the transactivation response element (TAR) located at the 5' end of HIV RNAs. It assists in the elongation phase of HIV-1 transcription so that full-length transcripts can be produced<sup>3</sup>, and these functions occur within the nucleus of infected cells. Tat has been shown to localise to the nucleus in many studies, however, the mechanism by which it interacts with the nuclear import receptors has not been elucidated structurally<sup>4,5</sup>.

Nuclear import can occur through passive diffusion (<45 kDa) or by energy dependent nuclear import receptors. The classical nuclear import pathway is the best characterised mechanism and is mediated by an adaptor molecule, importin- $\alpha$ , also known as the classical nuclear import receptor, binding cargo that can display a nuclear localisation signal (NLS). The transport carrier importin- $\beta$  interacts with importin- $\alpha$ , and mediates translocation across the nuclear envelope through interactions with the nucleoporin proteins lining the nuclear pore complex<sup>6,7</sup>. Upon entry to the nucleus, the heterotrimer transport complex is dissociated by the small GTPase Ran, releasing the NLS-containing cargo, and allowing recycling of the import receptors back to the cytoplasm<sup>8,9</sup>.

The HIV-1 Tat derived cell penetrating peptide (<sup>48</sup>GRKKRRQRRRAPQN<sup>61</sup>;CPP) has been shown to effectively carry a large range of cargoes, from nanoparticles, peptides, nucleic acids and even proteins into cells and the nucleus<sup>10-14</sup>. *In vitro* studies have shown that Tat is able to bind nuclear import receptors which mediate nuclear localisation<sup>5,15</sup>, however, a structural basis for this interaction remains to be elucidated. There has also been some

Charles Sturt University, School of Biomedical Sciences, Wagga Wagga, 2678, Australia. K. M. Smith and Z. Himiari contributed equally to this work. Correspondence and requests for materials should be addressed to J.K.F. (email: [jforwood@csu.edu.au](mailto:jforwood@csu.edu.au))



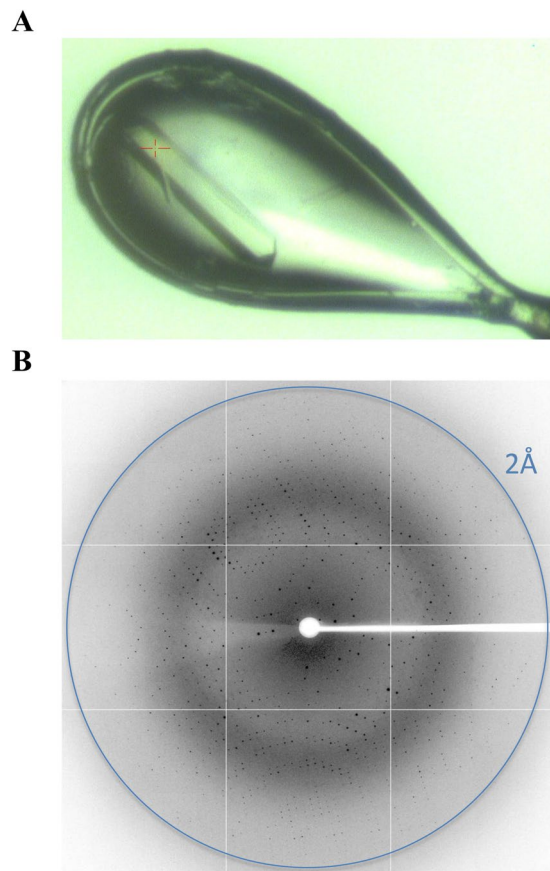
**Figure 1.** Binding of Tat:NLS/CPP to importin- $\alpha$  and importin- $\beta$ . **(A)** SDS-PAGE visualization of complex formation between Tat:NLS/CPP and importin- $\alpha$ . **(B)** SDS-PAGE revealing a lack of complex formation between Tat:NLS/CPP and importin- $\beta$ . Both gels were cropped at the right to remove samples from additional purification steps and other experiments. The full gels are presented in the Supplementary Figure 1.

debate in the literature about whether Tat can bind directly to importin- $\alpha$ <sup>16</sup> or importin- $\beta$ <sup>15</sup>. To determine the precise binding determinants that mediate interaction between the nuclear import receptor and Tat, the entire cell penetrating region of HIV-1 Tat, <sup>48</sup>GRKKRRQRRRAPQN<sup>61</sup>, was recombinantly expressed as a GST-fusion and tested for binding to both importin- $\alpha$  and importin- $\beta$ <sup>6,16</sup>. We found a strong and direct interaction between Tat:NLS/CPP and importin- $\alpha$ , and no direct interaction with importin- $\beta$ . Together with structural elucidation of the interface by x-ray crystallography, this study provides new insights into the interface between these two proteins which mediate localisation of Tat to the nucleus.

## Materials and Methods

**Plasmid preparation.** Tat residues (<sup>48</sup>GRKKRRQRRRAPQN<sup>61</sup>) were codon optimised for expression in *E. coli* and cloned into the PGEX4T-1 vector at BamHI/EcoRI sites with an additionally engineered N-terminal TEV site for GST-tag cleavage. An isolate of mouse importin- $\alpha$  (homologue of human importin- $\alpha$ ; 95% sequence identity) that lacks the auto-inhibitory N-terminal importin- $\beta$  binding (IBB) domain (residues 70–529) and cloned into the pET30 expression vector has been described previously<sup>17</sup>. An isolate of mouse importin- $\beta$  (KPNB1, homologue of human importin- $\alpha$ ; 99% sequence identity) was cloned into the pMCSG21 vector using protocols described previously<sup>18,19</sup>.

**Recombinant Expression and Purification.** Overexpression of importin- $\alpha$  and importin- $\beta$  was performed using the autoinduction method according to Studier<sup>20</sup> and purified as outlined previously<sup>21</sup>. Briefly, cells were resuspended in His buffer A (50 mM phosphate buffer, 300 mM NaCl, 20 mM Imidazole, pH 8), and lysed by two freeze-thaw cycles. The soluble cell extract was injected onto a five mL HisTrap HP column (GE Healthcare) and washed with twenty column volumes of His buffer A on an AKTApurifier FPLC. The sample was eluted using an increasing concentration gradient of imidazole, and eluent fractions were pooled and loaded onto a HiLoad 26/60 Superdex 200 column, pre-equilibrated in buffer A (50 mM Tris pH 8, 125 mM NaCl). Fractions corresponding to the correct molecular weight were collected, and assessed for purity by SDS-PAGE.



**Figure 2.** Tat:NLS/CPP importin- $\alpha$  crystal diffraction. **(A)** Rod-shaped crystal of Tat:NLS/CPP importin- $\alpha$  shown used for data collection **(B)** X-ray diffraction image obtained from the Tat:NLS/CPP importin- $\alpha$  crystal at the Australian Synchrotron.

Tat-NLS was over-expressed as a GST-fusion protein using IPTG to induce the cells at an  $OD_{600}$  of 0.6. Following growth for 24 h, cells were harvested by centrifugation, at 6,000 rpm for 30 min and the cell pellet resuspended in buffer A and stored at  $-20^{\circ}\text{C}$ . The soluble cell extract containing the GST:Tat-NLS fusion was injected onto a GST 5 mL HP column pre-equilibrated in GST buffer A and washed with GST buffer A until the UV reading stabilised to baseline. The purified importin- $\alpha$  or importin- $\beta$  was then added to the GST column, washed for a further ten column volumes, and eluted in GST buffer A containing 10 mM glutathione. To cleave the GST affinity tag from the Tat:NLS/CPP importin- $\alpha$ , TEV protease cleavage was undertaken overnight, and continued until cleavage was complete. The protein mixture was further purification by size-exclusion chromatography on a HiLoad 26/60 Superdex 200 column (GE Healthcare), and the complex concentrated to 7 mg/mL using an Amicon MWCO 10 kDa filter, aliquoted, and stored at  $-80^{\circ}\text{C}$ .

**Crystallization of Tat-NLS with Importin- $\alpha$ .** The final crystallisation condition contained 1.25 M sodium citrate pH 7 and 10 mM DTT. The Tat:NLS/CPP importin- $\alpha$  complex was screened using hanging drop vapour diffusion method where the protein was mixed in a 1:1 ratio with crystallization condition containing reservoir solution and incubated at 296 K. Large rod shaped crystals grew after four days, and the crystals were cryoprotected in a reservoir solution 20% glycerol prior to being flash cooled to 100 K in liquid nitrogen.

**Data Collection and Structure Determination.** A single crystal was used to collect x-ray diffraction data on the MX2 crystallography beamline at the Australian Synchrotron. The diffraction data was processed using iMosflm<sup>22</sup> and scaled and merged using Aimless<sup>23,24</sup>. The data was phased by molecular replacement using Phaser<sup>25</sup> and Protein Data Bank (PDB) structure 5FC8 as a search model. Model rebuilding and refinement were undertaken in Coot<sup>26–28</sup> and Phenix<sup>29</sup>, respectively.

**GST-pulldown Assay for Affinity Determination.** For determination of the dissociation constant,  $K_D$ , glutathione agarose beads washed in GST buffer A were saturated with purified GST-Tat:NLS/CPP. Beads were washed three times in GST buffer A and distributed equally in 10  $\mu\text{L}$  aliquots. To each aliquot, 100  $\mu\text{L}$  of two-fold serially diluted importin- $\alpha$  was added, with the initial concentration of 30  $\mu\text{M}$ . One additional tube containing 0  $\mu\text{M}$  importin- $\alpha$  was used as a control. The samples were incubated for one hour at  $4^{\circ}\text{C}$  before being washed twice with 1 mL of GST buffer A. Binding was assessed by adding 20  $\mu\text{L}$  of tris-glycine sample loading buffer to the beads, boiled for 10 min, and analysed by SDS-PAGE. Images were recorded using BioRad Gel Doc system and

	Tat:NLS/ CPP importin- $\alpha$ (PDB code 5SVZ)
Wavelength (Å)	0.9537
Data-collection temperature (K)	100
Detector Type	CCD
Detector	ADSC QUANTUM 315r
Mosaicity (°)	0.29
Resolution range (Å)	29.66–2.00 (2.05–2.00)
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell (Å)	79.00, 89.83, 99.46
Total reflections	212,694 (15,721)
Unique reflections	46,564 (3,439)
Multiplicity	4.6 (4.6)
Completeness (%)	96.5 (98.2)
Mean I/ $\sigma$ (I)	13.4 (2.4)
Wilson B-factor Å <sup>2</sup>	27.43
Matthew's Coefficient (A <sup>3</sup> /Da)	3.09
Solvent content (%)	60.22
R <sub>pim</sub>	0.047 (0.52)
R <sub>work</sub>	0.17 (0.25)
R <sub>free</sub>	0.19 (0.27)
No. of non-hydrogen atoms	3633
Macromolecules	3319
Solvent	314
Protein residues	434
r.m.s.d.	
bond length (Å)	0.003
bond angle (°)	0.65
Ramachandran favored (%)	98
Ramachandran allowed (%)	1.6
Ramachandran outliers (%)	0

**Table 1.** Data collection and refinement statistics for structure of importin- $\alpha$  in complex with HIV-1 Tat:NLS/ CPP domain. Values in brackets describe the highest resolution shell.

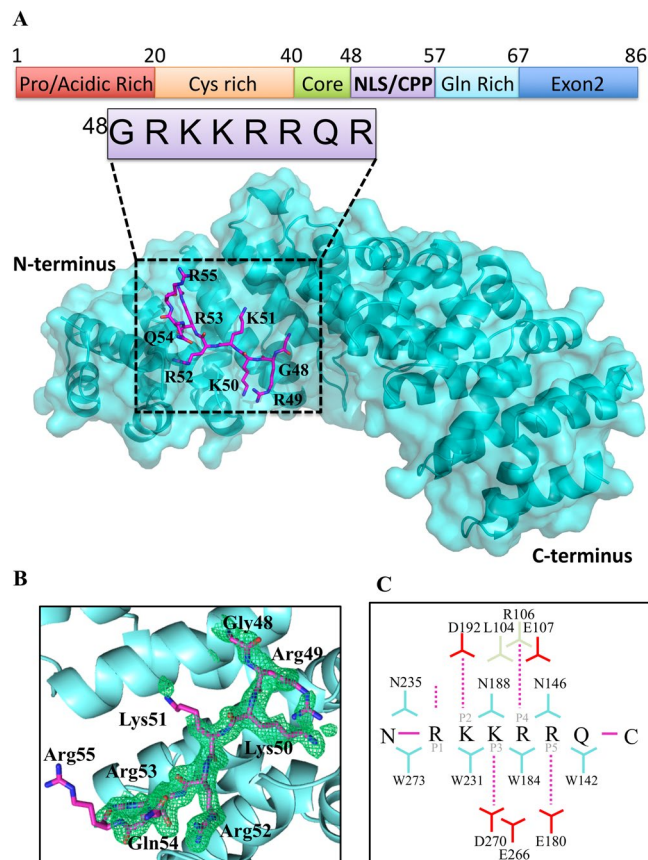
processed in ImageJ<sup>30</sup>. The data was normalised across each replicate experiment and data analysed using one-site specific binding analysis performed in Prism version 7.0b for Mac, GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com).

## Results

**The Tat:NLS/ CPP region forms a direct interaction with importin- $\alpha$ .** The NLS/ CPP region of Tat, spanning residues 49–61, have been shown to contain a functional NLS, however, there has been recent debate as to whether the highly basic cell penetrating peptide region is bound using the importin- $\alpha$  adapter, or can bind directly to importin- $\beta$ . Since this region contains a large stretch of positively charged residues, many of which could fit the definition of a classical NLS binding to importin- $\alpha$ , or an Arg rich importin- $\beta$  interaction, we tested binding against both types of receptors. Here, we immobilised the GST-Tat:NLS/ CPP fusion protein onto a glutathione column, washed the column, then passed each respective importin over the immobilised proteins to assess binding. We observed that most of the importin- $\alpha$  was retained on the column (Fig. 1A), whilst little, if any importin- $\beta$  remained bound (Fig. 1B). These results indicate a direct binding between the Tat:NLS/ CPP and the classical nuclear import receptor importin- $\alpha$ .

**Protein purification and complex formation.** To determine the structural basis for the interaction between the nuclear import receptor importin- $\alpha$  and Tat NLS/ CPP, both proteins were purified to homogeneity and isolated as an equimolar complex using the following series of purifications. The nuclear import receptor importin- $\alpha$  was first purified by 6-His affinity and size exclusion chromatography, then loaded on a column containing purified GST-Tat:NLS/ CPP. The excess importin- $\alpha$  was removed by washing the column extensively and following elution, the GST affinity tag was removed by proteolytic cleavage with the TEV protease. The mixture was then purified by size exclusion chromatography, where the importin- $\alpha$ :Tat NLS/ CPP complex (>58 kDa) was successfully separated from excess Tat NLS/ CPP (<5 kDa), resulting in a homogenous equimolar complex for crystallisation.

**Protein crystallisation and data collection.** The hanging-drop vapour diffusion method was used to obtain large rod-shaped crystals after four days (Fig. 2A). The crystal diffracted to 2.0 Å (Fig. 2B) resolution on the MX2 beam line at the Australian Synchrotron, and a total of 110° of data, collected at 0.5° oscillations, were

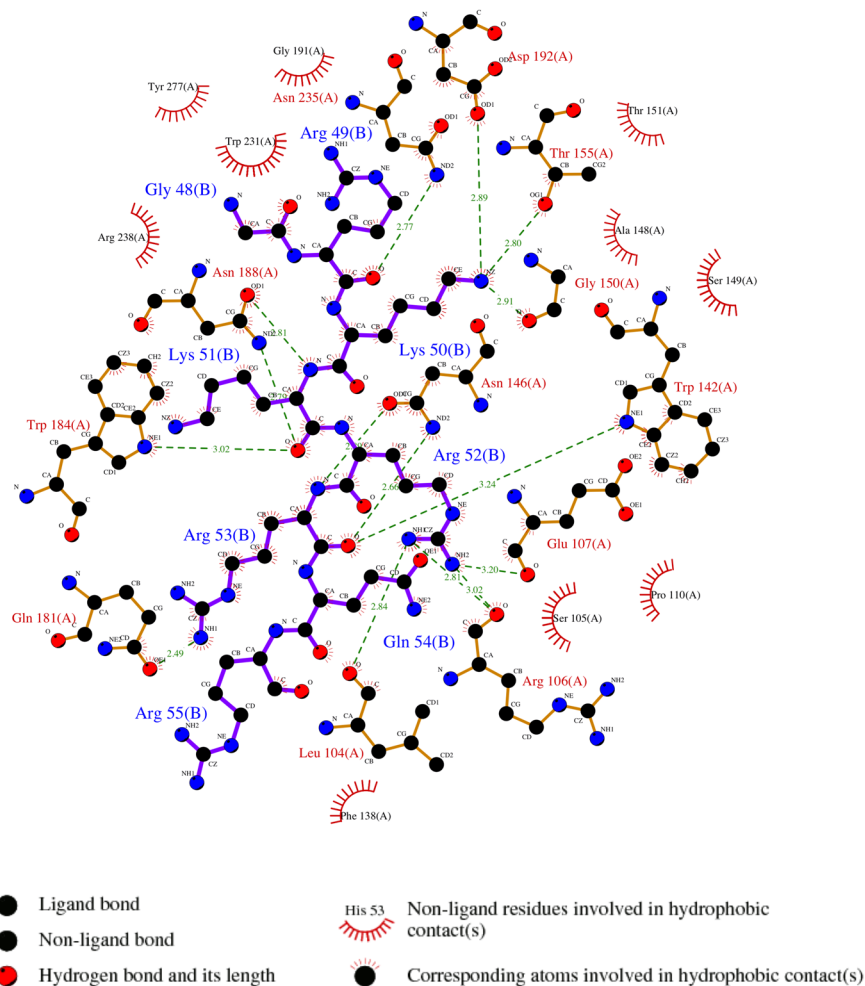


**Figure 3.** Crystal structure of Tat:NLS/CPP importin- $\alpha$ . **(A)** Full structure of Tat-NLS/CPP (purple sticks) and importin- $\alpha$  (cyan ribbons/transparent surface) complex. **(B)** Simulated annealing omit map (green mesh) of Tat-NLS/CPP shown at  $3\sigma$ . **(C)** Schematic representation of importin- $\alpha$  Tat:NLS/CPP interactions. The NLS backbone is indicated as a horizontal magenta line, from the N- to the C-terminus. NLS side chains are represented as vertical dotted magenta lines. Selected importin- $\alpha$  Trp and Asn residues are shown in blue. Selected importin- $\alpha$  Asp and Glu residues are shown in red. Single amino acid code is used. Structure has been deposited to the PDB and issued the code 5SVZ.

Tat:NLS/CPP				Importin- $\alpha$		
Atom	Residue	Number	Distance (Å)	Atom	Residue	Number
NH1	Arg	52	2.84	O	Leu	104
NH1	Arg	52	2.81	O	Arg	106
NH2	Arg	52	3.02	O	Arg	106
NH2	Arg	52	3.20	O	Glu	107
O	Arg	53	3.24	NE1	Trp	142
N	Arg	53	2.89	OD1	Asn	146
O	Arg	53	2.66	ND2	Asn	146
NZ	Lys	50	2.91	O	Gly	150
NZ	Lys	50	2.80	OG1	Thr	155
NH1	Arg	53	2.49	OE1	Gln	181
O	Lys	51	3.02	NE1	Trp	184
N	Lys	51	2.81	OD1	Asn	188
O	Lys	51	2.79	ND2	Asn	188
*NZ	Lys	50	2.89	OD1	Asp	192
O	Arg	49	2.77	ND2	Asn	235

**Table 2.** Tat:NLS/CPP hydrogen bond interactions with importin- $\alpha$  generated by PDBSum<sup>35–38</sup> \*Salt bridge interaction.

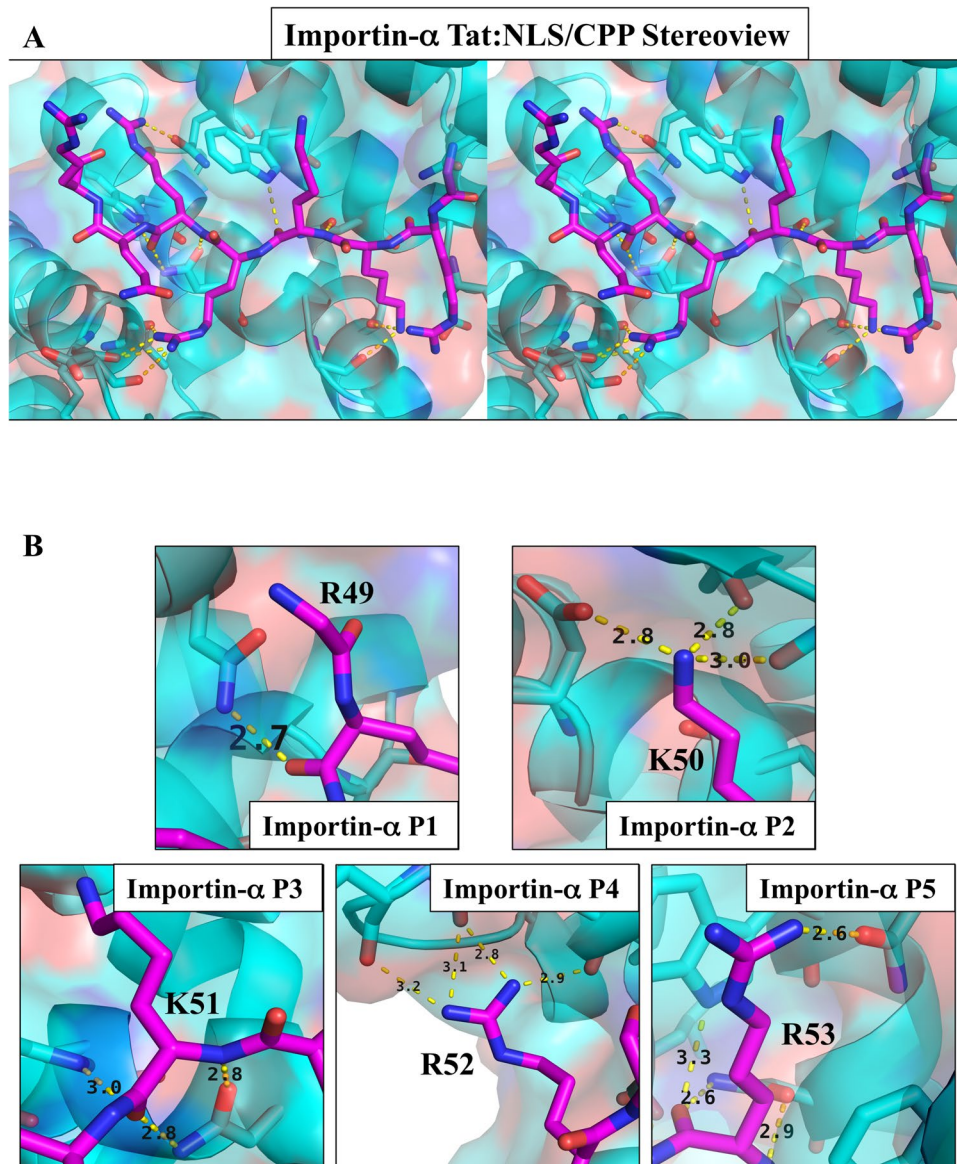




**Figure 4.** LigPlot interaction schematic for Tat:NLS/CPP importin- $\alpha$  structure. Tat:NLS/CPP intramolecular bonds are shown as solid purple lines, importin- $\alpha$  intramolecular bonds are shown as solid orange lines. Importin- $\alpha$  residues involved in hydrophobic contacts are shown as multiple red fanning lines. Corresponding atoms involved in hydrophobic contacts are shown with small orange fanning lines. Hydrogen bonds and their length are shown as green dotted lines. Figure was made using LigPlot<sup>34</sup>.

NLS	Importin- $\alpha$ major binding site					PDB ID
	P1	P2	P3	P4	P5	
HIV1 Tat	R	K	K	R	R	This study
SV40T <sup>7, 39, 40</sup>	K	K	K	R	K	1EJL, 1BK6, 1Q1S, 1Q1T
$\alpha$ IBB <sup>6</sup>	L	K	K	R	N	1IAL, 1IQ1
Venezuelan Equine Encephalitis Capsid	A	K	K	P	K	3VE6
Ku70 <sup>41</sup>	S	K	R	P	K	3RZX
Ku80 <sup>41</sup>	A	K	K	L	K	3RZ9
CLIC4 <sup>42</sup>	A	K	K	Y	R	3OQS
Dengue 2 NS5 C-terminus	M	K	R	F	R	5FC8
Dengue 3 NS5 C-terminus	M	K	R	F	R	5HHG
XPG <sup>43</sup>	S	K	R	K	R	5EKF, 5EKG

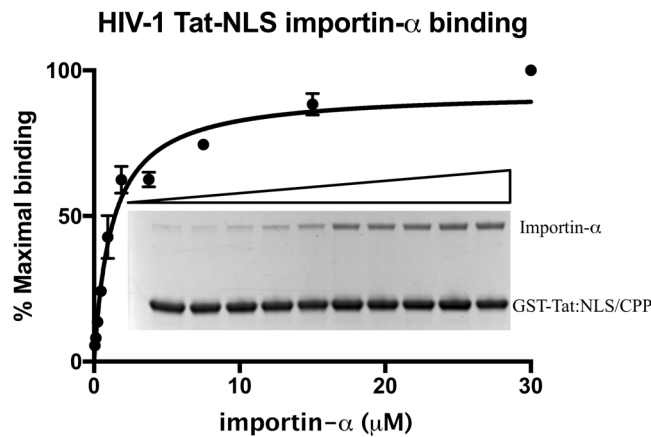
**Table 3.** PDB deposited cNLSs binding to the major site of importin- $\alpha$ .



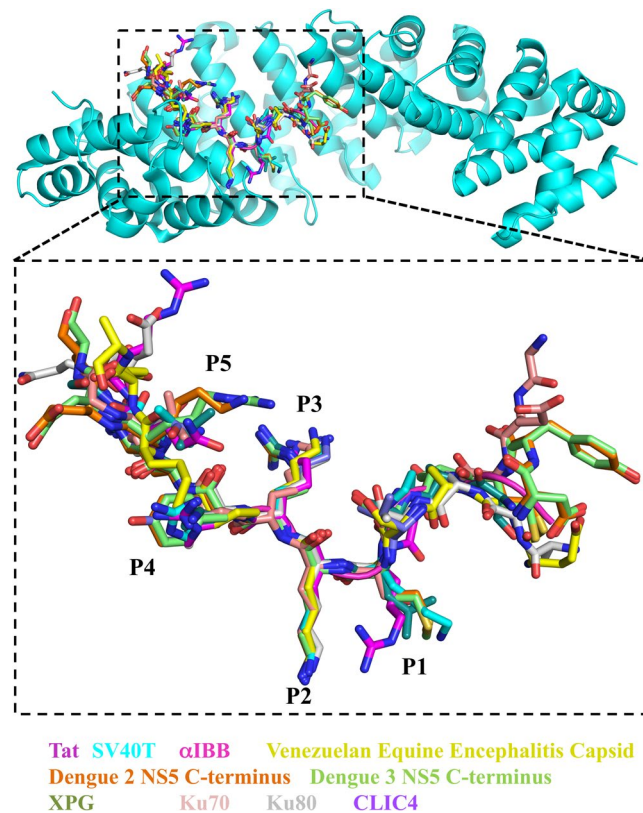
**Figure 5.** Importin- $\alpha$  interacting residues from Tat NLS/CPP region. (A) Importin- $\alpha$  shown in cyan (ribbons and surface view), Tat:NLS/CPP shown in magenta (sticks). NLS residues visible within crystal structure, and bound to the N-terminal concave ARM domains of importin- $\alpha$ , are shown in stereoview for 3D visualisation (B) Each P-site that interacts with importin- $\alpha$  is displayed with each interaction highlighted with dashed lines.

indexed, merged and scaled using *iMosflm* v.1.0.7<sup>22</sup> and AIMLESS<sup>23,24</sup> (Table 1). The structure was solved by molecular replacement using Phaser<sup>25,31</sup> with PDB 5FC8 as a model. Following rebuilding and refinement, the 2.0 Å resolution structure was refined to a  $R_{\text{work}}/R_{\text{free}}$  of 0.16/0.19 using iterative cycles of refinement and modelling in Phenix<sup>29</sup> and COOT<sup>28</sup> respectively. The final model consisted of 426 residues of importin- $\alpha$ , 8 residues of the Tat:NLS/CPP, and 319 water molecules. The N- and C-terminal amino acid residues of the Tat:NLS/CPP peptide displayed poor density due to flexibility, allowing only accurate placement of the protein peptide chain. Stereochemistry and other refinement statistics are presented in Table 1.

**Binding determinants of the HIV-1 Tat:NLS/CPP in complex with importin- $\alpha$ .** The overall structure of importin- $\alpha$  exhibited an all  $\alpha$ -helical structure arranged as ten sequential armadillo (ARM) motifs as described previously<sup>6</sup>. The Tat:NLS/CPP binds to the major binding site of importin- $\alpha$  within ARM domains 2–4 (Fig. 3A,B). There are no NLS residues in the minor binding site, indicating that the NLS region is monopartite. The main chain of Tat residue Arg49 interacts with the side chain of importin- $\alpha$  residue Asn235 in the P1 site (Figs 4 and 5). Tat residue Lys50 binds importin- $\alpha$  at the P2 site where it forms a salt bridge with importin- $\alpha$  residue Asp192 as well as additional side chain interactions with importin- $\alpha$  residues Gly150 and Thr155. The Tat peptide backbone of residue Lys51 hydrogen bonds the side chains of importin- $\alpha$  residues Asn188 and Trp184 in the importin- $\alpha$  P3 binding site. The side-chain of Tat residue Arg52 in the P4 position hydrogen bonds with the importin- $\alpha$  main



**Figure 6.** Quantitative GST-pull down for binding affinity determination. Glutathione agarose containing the GST-Tat:NLS/CPP incubated and washed with two-fold serially diluted importin- $\alpha$  (initial concentration of 30  $\mu$ M). Samples analysed by SDS-PAGE and images recorded using BioRad Gel Doc system were processed from triplicate gels and processed in ImageJ and analysed using one-site specific binding in Prism 7.0. A representative gel showing binding is included, and the original, uncropped gel is provided in Supplementary Figure 2.



**Figure 7.** Overlay of cNLSs bound at the major site of importin- $\alpha$ . Conservation of NLS structures are coloured accordingly: Tat (magenta), SV40T (cyan),  $\alpha$ IBB (pink), Venezuelan equine encephalitis virus (gold), Dengue 2 NS5 C-terminus (orange), Dengue 3 NS5 C-terminus (green), XPG (dark green), Ku70 (nude), Ku80 (grey), CLIC4 (purple), all overlaid onto Tat:NLS/CPP bound importin- $\alpha$  (cyan). Overlay of NLSs are enlarged within the P1-P5 positions.

chain of residues Leu104, Arg106, and Glu107. The P5 binding site is occupied by Tat residue Arg53 which makes main chain interactions with the side chains of importin- $\alpha$  residues Trp142 and Asn146. The overall binding buries 717  $\text{\AA}^2$  of surface area, and is mediated by 15 hydrogen bonds and 1 salt bridge interaction (Figs 3 and 4). Further details on the NLS binding determinants are shown in Figs 4B and 5 and summarised in Table 2.



NLS	PDB	RMSD for C $\alpha$ residues to Tat (Å)
SV40T <sup>7,39,40</sup>	1EJL	0.217
$\alpha$ IBB <sup>6</sup>	1IAL	0.206
Venezuelan Equine Encephalitis Capsid	3VE6	0.315
Ku70 <sup>41</sup>	3RZX	0.274
Ku80 <sup>41</sup>	3RZ9	0.431
CLIC4 <sup>42</sup>	3OQS	0.154
Dengue 2 NS5 C-terminus	5FC8	0.219
Dengue 3 NS5 C-terminus	5HHG	0.350
XPG <sup>43</sup>	5EKF	0.150

**Table 4.** RMSD differences of PDB deposited cNLSs binding to the major site of importin- $\alpha$ .

**Binding affinity of the HIV-1 Tat:NLS/CPP in complex with importin- $\alpha$ .** To estimate the binding affinity, importin- $\alpha$  was serially titrated against equal concentrations of HIV-1 Tat:NLS/CPP, and binding captured using a GST-pulldown. The binding affinity was determined to be  $1.2 \pm 0.2 \mu\text{M}$  from three replicates (Fig. 6). The binding affinity measured for HIV-1 Tat:NLS/CPP is in the low micromolar range and similar to previously reported values of other NLSs including Dengue 2 C-terminal NS5,  $0.27 \pm 0.1 \mu\text{M}$ ; and Dengue 3 C-terminal NS5  $0.37 \pm 0.11 \mu\text{M}$ <sup>32</sup>.

## Discussion

There has been contention as to which nuclear import receptor is responsible for the nuclear translocation of Tat. One study suggests Tat is importin- $\alpha$  mediated<sup>15</sup>, whereas another study has shown that it is dependent on importin- $\beta$ <sup>16</sup>. Here, we show that the C-terminal<sup>55</sup>RRR is not providing additional binding to importin- $\alpha$ , and of the residues visible in the crystal structure<sup>48</sup>GRKKRRQR, only residues<sup>48</sup>GRKKRR mediate binding with importin- $\alpha$ . Our results support the findings of Ruben *et al.*, who have shown nuclear import can be mediated by Tat-NLS/CPP residues GRKKR<sup>16</sup>. This binding motif is consistent with the previously defined monopartite class 2 NLSs<sup>33</sup>. The P1-P5 binding determinants in Tat:NLS/CPP are resolved within the presented 2.0 Å structure, and are consistent with previously solved structures (Table 3 and Fig. 7).

The major binding site of importin- $\alpha$  interact with five principal binding determinants in NLSs, known as P1-P5 (Table 3). These binding determinants P1-P5 bind between ARM domains 2–4 in importin- $\alpha$  and are conserved across previously determined NLS structures (Fig. 7). Position 1 (P1) on an NLS is the most divergent as only a main chain interaction is made with importin- $\alpha$ . Position 2 (P2) is the most highly conserved with a strict requirement of a lysine residue at this position to form the critical salt bridge interaction with importin- $\alpha$  residue D192 as well as side chain interactions with the main chains of importin- $\alpha$  residues T155 and G150. Position 3 (P3), position 4 (P4), and position 5 (P5), prefer long basic amino acids, however, for all characterised NLSs, there is a great variability in the amino acids at these positions (Fig. 7). The P3 site which is less conserved than the P2 site, can bind to Lys, Arg, and Leu residues, whilst the P5 site is similarly conserved and can bind Lys, Arg, and Asn residues. The P4 site is less conserved than P4 and P5, and can bind Pro, Phe, Arg, Lys, Leu, Tyr, and P1 is the least conserved binding to Met, Asn, Lys, Arg, Leu, Ala, and Ser residues<sup>32</sup> (Table 3). The Tat:NLS/CPP (Fig. 5A,B) residues<sup>49</sup>RKKRR fit within the requirements of the P1-P5 NLSs and bind importin- $\alpha$  in a similar fashion to other importin- $\alpha$  bound NLS structures, whilst other positively charged residues<sup>55</sup>RRR were found not to bind in our structure. The similarities between HIV-1 Tat:NLS/CPP and previously reported NLS importin- $\alpha$  structures can be seen in sequence conservation (Table 2), structural overlay comparisons (Fig. 7) and from calculated RMSD values that varied from 0.150–0.431 Å (Table 4). Overall, this is the first time the NLS region within HIV-1 Tat has been determined structurally by X-ray crystallography. Through interaction with the N-terminal arginine rich motifs in importin- $\alpha$ , eight amino acids in Tat have been revealed and binding determinants identified that mediate interaction with the classical nuclear import receptor.

## References

- Mansueto, P. *et al.* Vitamin D Deficiency in HIV Infection: Not Only a Bone Disorder. *BioMed research international* **2015** (2015).
- Strebel, K. HIV accessory proteins versus host restriction factors. *Current opinion in Virology* **3**, 692–699, doi:10.1016/j.coviro.2013.08.004 (2013).
- Southgate, C. D. & Green, M. R. The HIV-1 Tat protein activates transcription from an upstream DNA-binding site: implications for Tat function. *Genes & Development* **5**, 2496–2507 (1991).
- Frankel, A. D. & Pabo, C. O. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* **55**, 1189–1193, doi:10.1016/0092-8674(88)90263-2 (1988).
- Vives, E., Brodin, P. & Lebleu, B. A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *Journal of Biological Chemistry* **272**, 16010–16017, doi:10.1074/jbc.272.25.16010 (1997).
- Kobe, B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin  $\alpha$ . *Nature Structural & Molecular Biology* **6**, 388–397, doi:10.1038/7625 (1999).
- Fontes, M. R., Teh, T. & Kobe, B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin- $\alpha$ . *Journal of molecular biology* **297**, 1183–1194, doi:10.1006/jmbi.2000.3642 (2000).
- Marfori, M., Lonhienne, T. G., Forwood, J. K. & Kobe, B. Structural Basis of High-Affinity Nuclear Localization Signal Interactions with Importin- $\alpha$ . *Traffic* **13**, 532–548, doi:10.1111/j.1600-0854.2012.01329.x (2012).
- Cingolani, G., Petosa, C., Weis, K. & Müller, C. W. Structure of importin- $\beta$  bound to the IBB domain of importin- $\alpha$ . *Nature* **399**, 221–229, doi:10.1038/20367 (1999).

10. Brooks, H., Lebleu, B. & Vivès, E. Tat peptide-mediated cellular delivery: back to basics. *Advanced drug delivery reviews* **57**, 559–577, doi:[10.1016/j.addr.2004.12.001](https://doi.org/10.1016/j.addr.2004.12.001) (2005).
11. Berry, C. *Intracellular delivery of nanoparticles via the HIV-1 tat peptide* (2008).
12. Fawell, S. *et al.* Tat-mediated delivery of heterologous proteins into cells. *Proceedings of the National Academy of Sciences* **91**, 664–668, doi:[10.1073/pnas.91.2.664](https://doi.org/10.1073/pnas.91.2.664) (1994).
13. Eguchi, A. *et al.* Protein transduction domain of HIV-1 Tat protein promotes efficient delivery of DNA into mammalian cells. *Journal of Biological Chemistry* **276**, 26204–26210, doi:[10.1074/jbc.M010625200](https://doi.org/10.1074/jbc.M010625200) (2001).
14. Rudolph, C. *et al.* Oligomers of the arginine-rich motif of the HIV-1 TAT protein are capable of transferring plasmid DNA into cells. *Journal of Biological Chemistry* **278**, 11411–11418, doi:[10.1074/jbc.M211891200](https://doi.org/10.1074/jbc.M211891200) (2003).
15. Truant, R. & Cullen, B. R. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin  $\beta$ -dependent nuclear localization signals. *Molecular and cellular biology* **19**, 1210–1217, doi:[10.1128/MCB.19.2.1210](https://doi.org/10.1128/MCB.19.2.1210) (1999).
16. Ruben, S. *et al.* Structural and functional characterization of human immunodeficiency virus tat protein. *Journal of virology* **63**, 1–8 (1989).
17. Teh, T., Tiganis, T. & Kobe, B. Crystallization of importin  $\alpha$ , the nuclear-import receptor. *Acta Crystallographica Section D: Biological Crystallography* **55**, 561–563, doi:[10.1107/S0907444998012943](https://doi.org/10.1107/S0907444998012943) (1999).
18. Eschenfeldt, W. H., Stols, L., Millard, C. S., Joachimiak, A. & Donnelly, M. I. A Family of LIC Vectors for High-Throughput Cloning and Purification of Proteins. *Methods in molecular biology (Clifton, N.J.)* **498**, 105–115, doi:[10.1007/978-1-59745-196-3\\_7](https://doi.org/10.1007/978-1-59745-196-3_7) (2009).
19. Nanson, J. D. & Forwood, J. K. Crystallization and preliminary X-ray diffraction analysis of FabG from *Yersinia pestis*. *Acta Crystallographica. Section F, Structural Biology. Communications* **70**, 101–104, doi:[10.1107/S2053230X13033402](https://doi.org/10.1107/S2053230X13033402) (2014).
20. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification* **41**, 207–234, doi:[10.1016/j.pep.2005.01.016](https://doi.org/10.1016/j.pep.2005.01.016) (2005).
21. Patterson, E. L., Dombrowski, A. K., Swarbrick, C. M., Raidal, S. R. & Forwood, J. K. Structural determination of importin alpha in complex with beak and feather disease virus capsid nuclear localization signal. *Biochemical and biophysical research communications* **438**, 680–685, doi:[10.1016/j.bbrc.2013.07.122](https://doi.org/10.1016/j.bbrc.2013.07.122) (2013).
22. Batty, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography* **67**, 271–281, doi:[10.1107/S0907444910048675](https://doi.org/10.1107/S0907444910048675) (2011).
23. Evans, P. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* **62**, 72–82, doi:[10.1107/S0907444905036693](https://doi.org/10.1107/S0907444905036693) (2006).
24. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography* **67**, 282–292, doi:[10.1107/S090744491003982X](https://doi.org/10.1107/S090744491003982X) (2011).
25. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658–674, doi:[10.1107/S0021889807021206](https://doi.org/10.1107/S0021889807021206) (2007).
26. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235–242, doi:[10.1107/S0907444910045749](https://doi.org/10.1107/S0907444910045749) (2011).
27. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501, doi:[10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493) (2010).
28. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126–2132, doi:[10.1107/S0907444904019158](https://doi.org/10.1107/S0907444904019158) (2004).
29. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213–221, doi:[10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925) (2010).
30. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671–675, doi:[10.1038/nmeth.2089](https://doi.org/10.1038/nmeth.2089) (2012).
31. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallographica Section D: Biological Crystallography* **61**, 850–855, doi:[10.1107/S0907444905007894](https://doi.org/10.1107/S0907444905007894) (2005).
32. Tay, M. Y. *et al.* The C-terminal 18 Amino Acid Region of Dengue Virus NS5 Regulates its Subcellular Localization and Contains a Conserved Arginine Residue Essential for Infectious Virus Production. *PLoS Pathog* **12**, e1005886, doi:[10.1371/journal.ppat.1005886](https://doi.org/10.1371/journal.ppat.1005886) (2016).
33. Kosugi, S. *et al.* Six classes of nuclear localization signals specific to different binding grooves of importin  $\alpha$ . *Journal of Biological Chemistry* **284**, 478–485, doi:[10.1074/jbc.M807017200](https://doi.org/10.1074/jbc.M807017200) (2009).
34. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering* **8**, 127–134, doi:[10.1093/protein/8.2.127](https://doi.org/10.1093/protein/8.2.127) (1995).
35. de Beer, T. A., Berka, K., Thornton, J. M. & Laskowski, R. A. PDBsum additions. *Nucleic acids research* **42**, D292–D296, doi:[10.1093/nar/gkt940](https://doi.org/10.1093/nar/gkt940) (2014).
36. Laskowski, R. A. PDBsum: summaries and analyses of PDB structures. *Nucleic acids research* **29**, 221–222, doi:[10.1093/nar/29.1.221](https://doi.org/10.1093/nar/29.1.221) (2001).
37. Laskowski, R. A. PDBsum new things. *Nucleic acids research* **37**, D355–D359, doi:[10.1093/nar/gkn860](https://doi.org/10.1093/nar/gkn860) (2009).
38. Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Research* **33**, D266–D268, doi:[10.1093/nar/gki001](https://doi.org/10.1093/nar/gki001) (2005).
39. 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  $\alpha$ . *Cell* **94**, 193–204, doi:[10.1016/S0092-8674\(00\)81419-1](https://doi.org/10.1016/S0092-8674(00)81419-1) (1998).
40. Fontes, M. R. *et al.* Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40 large T-antigen nuclear localization sequences by importin- $\alpha$ . *Biochemical Journal* **375**, 339–349, doi:[10.1042/bj20030510](https://doi.org/10.1042/bj20030510) (2003).
41. Takeda, A. A., de Barros, A. C., Chang, C.-W., Kobe, B. & Fontes, M. R. Structural basis of importin- $\alpha$ -mediated nuclear transport for Ku70 and Ku80. *Journal of molecular biology* **412**, 226–234, doi:[10.1016/j.jmb.2011.07.038](https://doi.org/10.1016/j.jmb.2011.07.038) (2011).
42. Mynott, A. V. *et al.* Crystal structure of importin- $\alpha$  bound to a peptide bearing the nuclear localisation signal from chloride intracellular channel protein 4. *FEBS journal* **278**, 1662–1675, doi:[10.1111/j.1742-4658.2011.08086.x](https://doi.org/10.1111/j.1742-4658.2011.08086.x) (2011).
43. de Barros, A. C. *et al.* Structural and calorimetric studies demonstrate that Xeroderma pigmentosum type G (XPG) can be imported to the nucleus by a classical nuclear import pathway via a monopartite NLS sequence. *Journal of molecular biology* **428**, 2120–2131, doi:[10.1016/j.jmb.2016.01.019](https://doi.org/10.1016/j.jmb.2016.01.019) (2016).

## Acknowledgements

We would like to thank the staff and beamline scientists at the Australian Synchrotron for their valuable assistance during data collection.

### Author Contributions

Experiments performed by Z.H., data collection by Z.H., S.T., J.K.F., and K.M.S., structure solved by J.K.F. and K.M.S., structure analysis by K.M.S., manuscript written by K.M.S. and J.K.F., all authors contributed to final manuscript.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-01853-7](https://doi.org/10.1038/s41598-017-01853-7)

**Competing Interests:** The authors declare that they have no competing interests.

**Accession code:** HIV-1 Tat accession code is AAF35362.1, importin- $\alpha$  is NP\_034785.1, importin- $\beta$  is NP\_032405.3. The PDB code for the structure is 5SVZ.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017