Structural and calorimetric studies reveal specific determinants for the binding of a high-affinity NLS to mammalian importin-alpha

Hamine C. de Oliveira¹; Tainá D. da Silva¹; Guilherme H. M. Salvador¹; Ivan R. Moraes¹; Cíntia A. Fukuda¹; Andrea C. de Barros¹; Marcos R. M. Fontes¹

¹Departamento de Biofísica e Farmacologia, Instituto de Biociências, Universidade Estadual Paulista, Botucatu-SP, Brazil.

Correspondence author: Marcos R. M. Fontes (marcos.fontes@unesp.br)
ABSTRACT

The classical nuclear import pathway is mediated by importin (Impα and Impβ), which recognizes the cargo protein by its Nuclear Localization Sequence (NLS). NLSs have been extensively studied resulting in different proposed consensus; however, recent studies showed that exceptions may occur. This mechanism may be also dependent on specific characteristics of different Impα. Aiming to better understand the importance of specific residues from consensus and adjacent regions of NLSs, we studied different mutations of a high affinity NLS complexed to Impα by crystallography and calorimetry. We showed that although the consensus sequence allows Lys or Arg residues at the second residue of a monopartite sequence, the presence of Arg is very important to its binding in major and minor sites of Impα. Mutations in the N or C-terminus (position P1 or P6) of the NLS drastically reduces their affinity to the receptor, which is corroborated by the loss of hydrogen bonds and hydrophobic interactions. Surprisingly, a mutation in the far N-terminus of the NLS led to an increase in the affinity for both binding sites, corroborated by the structure with an additional hydrogen bond. The binding of NLSs to the human variant Impα1 revealed that these are similar to those found in structures presented here. For human variant Impα3 the bindings are only relevant for the major site. This study increases understanding of specific issues sparsely addressed in previous studies that are important to the task of predicting NLSs, which will be relevant in the eventual design of synthetic NLSs.
INTRODUCTION

The nuclear envelope presented in the eukaryotic cells establishes an essential barrier to controlling and regulating cellular processes, such as gene expression and cell cycle evolution. The transport of the most well-characterized components through the nuclear membrane occurs after the recognition of classic nuclear localization sequences (NLSs) in proteins that will need to be imported [1-3]. These proteins in the cytoplasm are transported by a heterodimeric receptor comprised of importin α and β (Impα and Impβ, respectively). Impα recognizes and binds to NLS, and impβ transports the molecular complex through the nuclear pore [4-6].

Impα has two important regions, the impβ binding domain (IBB) present in the N-terminal and the region containing the NLS binding sites, which are responsible for interacting with other macromolecules awaiting transport [3, 7-9]. Classic NLSs (cNLS) are classified as monopartite when the basic amino acid cluster interact with the major site region of Impα and as bipartite when two basic clusters interact simultaneously with the major and minor site regions of Impα [1, 10]. The major binding site is associated with the main binding region for monopartite cNLSs, and an eventual binding of monopartite cNLSs to the minor site is related to the high concentration used for the NLS peptides in the crystallization assays [8].

The consensus of classic NLS was initially proposed based in the sequences of SV40 TAg NLS (\textsuperscript{126}PKKKRRV\textsuperscript{132}) for monopartite NLSs and by nucleoplasmin NLS (\textsuperscript{155}KRPAATKKAGQAKKKK\textsuperscript{170}) for bipartite NLSs [10, 12]. The monopartite cNLS motif was defined as K(K/R)X(K/R), and the bipartite NLS consensus sequence corresponds to KRX\textsubscript{10-12}K(K/R)X(K/R)(where X corresponds to any residue) [13]. Structural studies with
SV40 TAg and nucleoplasmin NLSs complexed to yeast Impα, and mammalian Impα confirmed these binding modes [7, 8].

In order to gain further understanding of the specificities for monopartite NLSs’ interactions with the protein receptor, Hodel and colleagues [14] performed an alanine scan to study the energetic contribution of each NLS residue using the sequences of SV40 T-Ag and Myc NLSs. They used an experimental approach to measure the $K_d$ for the binding between NLS-GFP fusion proteins and Impα using fluorescence assays. Interestingly, all tested mutations (P0 to P6) resulted in an affinity diminution which the higher impact was for P2 position (100-fold) and P3 and P5 (10-fold) for SV40 TAg fusion protein. In an effort to better categorize experimentally observed NLSs and to describe atypical NLSs, Kosugi and colleagues [15] used various types of NLSs by random peptide libraries, leading to the suggestion of six classes of NLSs, which includes two noncanonical (or atypical) NLSs [15]. In the case of bipartite NLSs, the study by Marfori et al. [16] (by solid-phase binding assays) using two NLSs with the consensus sequence (Bimax1 and Bimax2) revealed insights about the interaction of bipartite peptides into Impα.

Different structural studies emphasized the importance of particular amino acids in the N and C termini of NLSs, the role of flaking sequences in the NLS N-terminal region [17-19], and the role of P4 position using structural and functional assays [20]. Additionally, a review presented by Marfori and colleagues [5], complemented with data from databases and a predictive tool for NLSs. Furthermore, another review published by Christie and collaborators [21] exhibited the different pathways and the proteins related to the nuclear importation using structural biology data.

Different classes of NLSs were suggested by an analysis of the interactions of NLSs with Impα variants from a random peptide library, resulting in six classes, including two types of atypical NLSs [15]. Additionally, structural studies using Impα from the α1
family present in plants (rice) and fungus (*Neurospora crassa*), identified monopartite NLSs that only binds to minor site or have higher affinity to this site compared to the major site [22-25]. In contrast, a recent study with a fungus-specific NLS demonstrated its preference for the major binding site [26, 27]. All these studies demonstrate the high variability of NLSs with several Impα from different families.

Some Impα variants may exist as a result of cellular duplication events. These variants can be divided into three main families (α1, α2 and α3), and may have preferences for specific NLSs that can be associated with specific roles [28-30]. Mouse genomes present six variants, while the human genome presents seven. A variant can play a fundamental role during the early stages of embryonic development [31] and during preferential interactions with specific proteins [32-34]. The causes for the specificity for NLS binding in different Impα variants has not been thoroughly studied, but the residues present near NLS binding sites have been proposed as one of the factors [23, 25]. Of the seven variants for human Impα [30], HsImpα1 is the most general importer, as it carries many different proteins that contain the classic NLS in their sequence [28, 35]. Furthermore, HsImpα1 has higher homology with the most studied Impα (*Mm*Impα isoform α2) sharing approximately 99% of the amino acid sequence [35]. In contrast, the HsImpα3 has higher specificity in cargo import, such as transcription factors, nuclear factors and regulators of gene expression [36-38], which is probably related to the higher flexibility in the armadillo (ARM) repeats 7 and 8 that are positioned closer to the NLS binding site [20, 35, 38].

Several proteins related to DNA repair are transported to the cell nucleus by the classical NLS-import pathway [39], including proteins that belong to the five main DNA repair processes [40]. Several Impα/NLSs complexes from proteins related to DNA repair have been studied using crystallography and affinity techniques [18, 41, 42]. The heterodimer formed by the MLH1 and PMS2 proteins occurs along the pathway responsible
for repairing errors in DNA replication when there is a mismatch repair (MMR) [39]. It has been shown that both MLH1 and PMS2 are imported via the classical nuclear import pathway, alone or forming a heterodimeric complex [42, 43]. We previously studied the interaction of MLH1 NLS (466SSNPRKRHRED476) and PMS2 NLS (571LATPNTKRFKKE583) complexed to *Mus musculus* Impα (MmImpα) and have demonstrated that Impα/MLH1 NLS has the higher affinity already observed in Impα/NLS complexes [42]. Due to this, MLH1 NLS was selected in the present study to further understanding of the classic NLS consensus.

In the present work different mutations in the nuclear localization sequence (NLS) of the MLH1 protein, a DNA repair protein with a very high affinity to MmImpα [42], were selected in order to study key amino acids responsible for this interaction. Thus, four mutated NLS peptides were co-crystallized with Impα. Their structures were solved by X-ray crystallography, and, subsequently, isothermal titration calorimetry assays were performed to quantify these interactions and calculate their stoichiometry and other thermodynamic values. These assays yield valuable information regarding the structural determinants of little studied residues from the consensus sequences and adjacent positions. The data obtained here is not only useful in understanding the role of binding of MLH1 NLS, but for the general binding of NLSs to Impα since majority of these residues in particular positions are found in other NLSs. Additionally, in order to obtain a more general study, we performed bioinformatic assays aiming to obtain binding characteristics of these four mutated MLH1 NLSs with two different variants of human importin α (variants 1 and 3).

**EXPERIMENTAL**

**Synthesis of NLS peptides.**
The peptides corresponding to MLH1-R472K NLS (\textsuperscript{466}SSNPRKKHRED\textsuperscript{476}), MLH1-E475A NLS (\textsuperscript{466}SSNPRKRHRAD\textsuperscript{476}), MLH1-S467A NLS (\textsuperscript{466}SANPRKRHRED\textsuperscript{476}) and MLH1-R470A NLS (\textsuperscript{466}SSNP\textsuperscript{A}KRHRED\textsuperscript{476}) were synthesized and purchased from GenOne company at purity levels that exceeded 95%.

**Protein expression and purification.**

The gene encoding the protein 6xHis-tagged truncated *Mus musculus* importin-\(\alpha\text{2} (\text{MmImp}\alpha)\) was expressed using a heterologous system with *E. coli* (DH5\(\alpha\)) and isolated by a His-trap nickel affinity column (5 mL, GE Healthcare), as previously used by Teh and colleagues [44]. The protein elution was evaluated using a gradient (0 to 100\%) of 0.15 M Imidazole in 0.5 M NaCl and 0.02 M Hepes pH 7.0, followed by overnight dialysis using a buffer composed of 0.02 M Tris HCl pH 8.0 and 0.1 M NaCl. The protein was concentrated using a Vivaspin 20 (30kDa - GE Healthcare) and stored at -20\(^\circ\)C. The sample purity was evaluated using 12\% SDS-PAGE [45].

**Calorimetric assays.**

Isothermal titration calorimetry (ITC) assays were used to quantify the binding of MLH1-R472K, MLH1-E475A, MLH1-S467A and MLH1-R470A NLS peptides with the *MmImp\alpha*. ITC measurements were performed using a MicroCal iTC200 microcalorimeter (GE Healthcare). The sample cell was loaded with 30 \(\mu\)M of *MmImp\alpha* that was individually titrated with the NLS peptides at a concentration of 300 \(\mu\)M for MLH1-R472K, MLH1-S467A and MLH1-R470A and of 600 \(\mu\)M for MLH1-E475A. Titrations were performed at 20\(^\circ\)C with twenty injections containing 2 \(\mu\)l at an interval of 240 s with an 800-rpm homogenization speed.
The thermogram areas of peaks were determined automatically with *Origin* v.7.0 Add-on, provided by the manufacturer, and curve fitting was performed by binding polynomials as previously described [46].

**Crystallization and X-ray data collection.**

Single crystals of all four complexes of *MmImpα/MLH1 NLS* mutants were obtained using the conventional hanging-drop vapor-diffusion method [47]. Crystals were obtained at 291 K after approximately 14 days, from a mixture of 1 µL protein solution (concentrated up to 18 mg/mL\(^{-1}\)), 0.5 µL peptide solution (1:8 – protein:peptide), 1 µL of reservoir solution containing sodium citrate buffer (pH 6.0), sodium citrate and DTT, as previously described by Fontes *et al.*, 2000 [8].

Crystals were mounted in a nylon loop and flesh-cooled in liquid nitrogen using a cryoprotectant composed of a reservoir solution supplemented with 25% glycerol. X-ray diffraction data were collected from single crystals of *MmImpα/MLH1 NLS* mutants using a synchrotron radiation source (MX2 station, Laboratório Nacional de Luz Sincrotron (LNLS), Campinas, Brazil) with a wavelength of 1.425 Å (at 100 K) and a *PILATUS 2M* detector (*Dectris*).

**Structure determination and refinement.**

Data were processed using an *XDS* v.20180126 program [48], and the crystal structure of the complexes was solved by Fourier synthesis using the *MmImpα/MLH1 NLS* (PDB access code 5U5P) coordinates as the initial model [42]. The modeling process was conducted by alternating cycles of manual rebuilding using the *Coot* program v.0.8.9 [49] and automated refinement was performed using the *PHENIX* package v.1.12 [50]. To check the overall quality of the final model, the *PHENIX* package v.1.12 was used. X-ray
crystallographic structure coordinates were deposited in the Protein Data Bank (RCSB PDB—www.rcsb.org). Data collection and refinement information were described in Table 1.

Comparative analysis.

X-ray crystallographic structures of MmlImpα/MLH1 NLS (Importin alpha from Mus musculus—PDB id: 5U5P) [42], HsImpα1 (Importin subunit alpha-1 from Homo Sapiens—PDB id: 5H43) [51], and HsImpα3 (Importin subunit alpha-3 from Homo Sapiens—PDB id: 6BWB) [20] were used. Comparative structural analyses were performed using the Coot program v.0.8.9 [49] and the PyMOL program v.1.8.6 [52]. The figures were generated using the PyMOL program v.1.8.6 and the LigPlus program v.1.4.5 [53].

Protein-peptide docking.

Molecular docking using the MLH1 and mMHL1 peptides and crystal structures of the HsImpα1 and HsImpα3 was performed by the Autodock Vina tool, present in the PyRx v.0.9.5 program [54]. The best complexes obtained were chosen by two selection criteria: i) the antiparallel conformation of NLS related to the Impα models and ii) Gibbs free energy expressed in kcal/mol. The first criterion was adopted because all Impα/NLS complexes solved to date present this conformation [55].

RESULTS

Crystallography structures of Importin-α bound to mMLH1 NLS.

The crystallographic structures of MmlImpα were solved with the truncated conformation at the residues 1 to 69, which represents the IBB domain responsible for the autoinhibition of Impα [56]. Additionally, all the crystal structures present the conserved ten
armadillo repetitions (ARM) responsible for the maintenance of the two binding sites for the NLS sequences [57]. The truncated protein was co-crystallized with four different mutated MLH1 NLSs (mMLH1): MLH1-R472K NLS (\(\text{SSNP}^4\text{RKKHR}^6\)), MLH1-E475A NLS (\(\text{SSNP}^4\text{KRHRAD}^6\)), MLH1-S467A NLS (\(\text{SSNP}^4\text{ANPRKRHRED}^6\)), and MLH1-R470A NLS (\(\text{SSNP}^4\text{AKRHR}^6\)) peptides (Fig. 1). All the complexes have been deposited in PDB under the respective access codes 6WBA, 6WBB, 7M60, and 6WBC.

The crystal structure of the complex \(\text{MmImp}\alpha/\text{MLH1-R472K NLS}\) only presented the NLS interacting at the major binding site (Fig. 1 A), featuring a monopartite structure as observed in other crystal structures of complexes \(\text{MmImp}\alpha/\text{Ku70 NLS}\) and \(\text{MmImp}\alpha/\text{Ku80 NLS}\) [18]. It was possible to insert 10 residues in the electron density (1.2\(\sigma\)) that correspond to the NLS (\(\text{SNPRKKHRED}^6\)), P1-P5 residues highlighted) that interact with Imp\(\alpha\) via 14 hydrogen bonds (Fig. 2 C).

The crystal structures of other three complexes present interactions of mMLH1 at both major and minor bind sites (Fig. 1 B, C and D), as previously observed in the crystal structure of \(\text{MmImp}\alpha/\text{MLH1 NLS}\). The complex \(\text{MmImp}\alpha/\text{MLH1-E475A NLS}\) present interactions of 10 residues of NLS at the major site (\(\text{SNPRKRHRAD}^6\)) via 14 hydrogen bonds with the Imp\(\alpha\) (Fig. 2 D). At the minor site, 8 NLS residues interact with the Imp\(\alpha\) via 11 hydrogen bonds, including the P1’ to P4’positions (\(\text{KRHR}^4\)) (Fig. 2 E). Similarly, the complex \(\text{MmImp}\alpha/\text{MLH1-R470A NLS}\) also presents interaction of the NLS at the major and minor binding sites of Imp\(\alpha\). The numbers of hydrogen bonds at the major (14) and minor (11) binding sites are similar (Fig. 2 H and I). Finally, the complex \(\text{MmImp}\alpha/\text{MLH1-S467A NLS}\) shows the interaction of P1 to P5 NLS residues at the major binding site through 17 hydrogen bonds and at the minor binding site through 12 hydrogen bonds (Fig. 2 F and G).

It was observed lower B-factors for the regions P1–P5 in the major site and P1’–P4’ in the minor site compared to the entire peptides for all \(\text{MmImp}\alpha/m\text{MLH1 complexes}\) (Table 1).
Calorimetric assays of mMLH1 NLS and Importin α.

Isothermal titration calorimetry (ITC) assays were performed using the MmlImpα and the four mMLH1 NLSs (MLH1-R472K, MLH1-E475A, MLH1-S467A and MLH1-R470A). The isotherms obtained were fitted at two non-identical and independent binding sites. The dissociation constant (K_d) and the enthalpy (ΔH) were obtained from the thermograms shown in Figure 3 and described in Table 2.

The MmlImpα/MLH1-E475A, MmlImpα/MLH1-R470A and MmlImpα/MLH1-S467A NLSs complexes display a stoichiometry of 2, indicating that the peptides interact at the major and minor binding sites. The first two complexes present lower affinity than MmlImpα/MLH1 NLS in both sites (Table 2); however, the third complex (MmlImpα/MLH1-S467A) present higher affinity than MmlImpα/MLH1 NLS in both sites. The MmlImpα/MLH1-R472K NLS complex display a stoichiometry of 1, suggestion the interaction of the peptide only with the major binding site. In addition, it can be observed for the Table 2 that the interaction between MmlImpα and MLH1-S467A NLS present the higher K_d at the major site compared to other three complexes.

Molecular docking.

Molecular docking using the MLH1 and mMLH1 NLSs with human variants HsImpα1 and HsImpα3 were performed in order to study the binding behavior of NLS peptides at the minor and major binding sites. As demonstrated in the previous sections, the crystal structures of MmlImpα/MLH1 NLS [42] and MmlImpα/MLH1 mutants NLS complexes revealed the presence of monopartite peptides at similar binding sites, except the MmlImpα/MLH1-R472K complex whose peptide was absent in its minor site.
Molecular docking results involving HsImpα1 complexed to MLH1 and mMLH1 (MLH1-E475A, MLH1-S467A and MLH1-R470A) NLSs present similar interactions at the minor and major sites, as found in the crystal structures presented in the present study. The HsImpα1/MLH1-R472K complex only interacts with the major site (Table 3), as also found in MmImpα1/MLH1-R472K crystal structure.

In contrast, docking results using HsImpα3 with MLH1 and mMLH1 (MLH1-E475A, MLH1-R470A and MLH1-R472K) reveal the presence of the peptides interacting at the major site with no significant differences as found in HsImpα1, but no NLS were found at the minor site. The MLH1-S467A NLS are found at both the major and minor sites (Table 3).

DISCUSSION

MLH1-R472K NLS—Mutation in the P3 position and abolishment of minor site binding.

According to the consensus sequence for monopartite NLSs [K(K/R)X(K/R)], the position P3 can be occupied by Lys or Arg residues. Thus, we decided to test the mutation Arg→Lys (R472K) in a monopartite NLS with a high affinity (MLH1 NLS) for the residue that binds at the P3 position of the major site. Additionally, this mutation also has an impact in the binding of this residue at the P2' position of the minor site.

Calorimetric data of the MmImpα/MLH1-R472K NLS complex showed a decreasing affinity in one order of magnitude compared to the non-mutated peptide (Table 2) and a stoichiometry n=1, indicating a singular binding site. The crystal structure of this complex showed that the mutated residue (R472K) is still bound to the protein in the P3 position; however, the hydrogen bond of the Lys residue to Asp228 (Impα) present in the MmImpα/MLH1 structure is absent in the mutated complex (Fig. 2 C). Indeed, according to
a previous study [14], an Ala mutation in the residue binding to the P3 position had the second highest impact for SV40 TAg and Myc NLSs and only the mutation in the invariant P2 (Lys) had a higher contribution.

In Ku70 and Ku80 proteins, the formation of a heterodimeric complex involved in DNA repair occurs, but they also may have a separate role. The NLSs from both proteins have the same residues in the consensus region (P2–P5, Table 4), with exception of a presence of a Lys instead of an Arg in the P3 position (the same sequence of R472K NLS). Reinforcing the present study, Ku80 NLS displayed a lower affinity (Kd four times higher) than Ku70 NLS [18]. Other protein related to DNA repair, XPG (Xeroderma pigmentosum type G), presents its NLSs (XPG1 and XPG2) with Kd in the same order of magnitude of MLH1 and also displays an Arg residue in the P3 position [41]. The optimized bipartite NLS sequences, Bimax1 and Bimax2 also present an Arg residue in the P3 position [5]. Finally, SV40 TAg NLS presents a Lys in the P3 position and a Kd of 1.8 μM, that is eighteen times higher than MLH1 NLS [42]. Thus, considering all these data, it is possible to affirm that an Arg residue is more favorable than a Lys residue in the P3 position because it frequently occurs at the NLSs with high affinity, e.g., MLH1 and XPG1.

Importantly, the mutation studied here (R472K) prevents the binding of the peptide to the minor site, highlighting the importance of Arg residue in the P2’ position, as observed in several monopartite NLS sequences (Table 4). The exceptions for Arg in the P2’ position are only pUL15 and SV40 TAg NLS peptides, i.e., those that have the KK sequence (Table 4). Although their crystal structures have the NLS bound at the minor site, both have a high B-factor, reaffirming the importance of arginine in the P2’ position [3, 8, 55]. Additionally, it has been shown the SV40 TAg NLS presents multiple binding modes at the minor site in the structure with Mmlmapa [8] and a lower affinity compared to similar NLS peptides [42]. The crystal structure of Mmlmapa/CN-SV40 TAg NLS (N-terminally extended
SV40Tag NLS, \(1^{10}GPGSDDEAAADAQHAAPPKKRKG\) \(^{132}\) [17] presents Arg bound to the P2' position (Table 4). Thus, although an arginine residue at the position is not mandatory in this position, it can be observed that the large majority of NLS peptides have KR residues in the P1' and P2' positions (Table 4) [41].

In summary, although the consensus sequence allows Lys or Arg residues at the second amino acid of a monopartite basic cluster, the presence of an Arg in monopartite NLSs seems to be very important for their binding at both major (P3 position) and minor (P2' position) sites of Imp\(\alpha\), particularly for the variant \(MmImp\alpha\) isoform \(a2\) as studied here [21].

**MLH1-R470A and E475A NLSs — Mutations in the P1 and P6 positions.**

The importance of the P1 and P6 positions is not well established, and their affinity with importin has not been studied in depth. According to the study of Hodel and colleagues [14], Ala mutation in the residue binding at the P6 position impacted SV40 Tag (\(K_d\) is ~5 higher) and Myc NLSs affinity (\(K_d\) is ~14 higher), while an Ala mutation in the P1 position impacted SV40 TAg (\(K_d\) is ~2 higher). Few additional studies have considered the relevance of these positions [5, 8, 16, 42]. Thus, we tested the mutations Arg→Ala and Glu→Ala in a monopartite NLS with high affinity (MLH1 NLS), which are responsible for the interactions at the P1 and P6 positions of the major site.

Calorimetric data of \(MmImp\alpha/R470A\) and \(MmImp\alpha/MLH1-E475A\) NLSs show that both mutations lead to an important decrease in the affinities of these peptides for \(MmImp\alpha\) at both binding sites (major and minor sites). In particular, the mutation in the N-terminus (R470A) reduced the affinity 30-fold and 5-fold in the major and minor binding sites, respectively. Similarly, the mutation at the C-terminus (E475A) reduced the affinity of the
mutated peptide to Impα, but with less intensity. It decreased the affinity by 7-fold and 2.5-fold for the major and minor sites, respectively (Table 2).

The crystal structures corroborate the calorimetric data, revealing that the mutations in the P1 (R470A) and P6(E475A) positions lead to the loss of hydrogen bonds and hydrophobic interactions at both binding sites (major and minor). Two hydrogen bonds are not observed at the major binding site for MmlImpα/MLH1-R470A compared to the MmlImpα/MLH1 NLS structure [42]: Trp231 (Impα)—Pro469 (NLS) and Asn239 (Impα)—Arg470 (NLS), but a new hydrogen bond Thr277 (Impα)—Asn468 (NLS) is observed (Fig. 2 H). In the case of MmlImpα/MLH1-E475A NLS, one hydrogen bond (Trp231 (Impα)—Pro469 (NLS) is lost in comparison to MmlImpα/MLH1 NLS for the major binding site (Fig. 2 D).

Regarding the minor binding site, both R470A and E475A NLS structures maintain eleven hydrogen bonds when compared to non-mutated NLS; however, the Asn361 (Impα)—Lys471 (NLS) is replaced by the interaction Asn361 (Impα)—Ala470 (NLS) (Fig. 2 I). Thus, these NLS mutations have less impact in the minor site than in major binding site, as observed in the calorimetric study.

MLH1-S467A NLS — Mutations in N-terminal sequence.

The importance of flanking sequences in the N-terminal region of the monopartite NLS sequences have been emphasized in some articles. The first structural study that used the N-terminally extended SV40TA g NLS (CN-SV40TA g NLS) [17] provided important information about the binding profile of this region, which was also observed in the Ku70 NLS structure [18]. This study [17] aimed to analyze the role of the phosphorylation of specific residues in the N-terminal region, which was further investigated by other authors [19, 58]. Kosugi and colleagues [15] classified proposed classes with residues in the N-
terminal region (P1 position). More recently, a possible role for Ser residue in the N-termini of MLH1 NLS was evidenced. However, no study has related structural and affinity data in the residues of the N-terminal region to date. Thus, we have tested the mutation Ser→Ala in MLH1 NLS for the residue that binds at the P-2 of the major site.

Surprisingly, contrasting to the other three complexes presented here, calorimetric studies with MmImpα/MLH1-S467A NLS showed a higher affinity for both binding sites, despite the values being in the same order of magnitude of non-mutated complexes (Table 4). The crystal structure of the complex MmImpα/MLH1-S467A NLS reveals the presence of an additional hydrogen bond Tyr277 (Impα) - Asn468 (NLS) in the major site, which can be related by the increase of NLS accessibility led by the mutation S467A (Fig. 2 F). At the minor site, the hydrogen bonds found between the Impα and NLS from both structures (MmImpα/MLH1-S467A NLS and MmImpα/MLH1) remain similar.

Indeed, bipartite NLSs with higher affinities had prolines and acidic amino acids in the linker region, and in the regions closer to the minor and major binding sites [15]. Another study showed that the monopartite conformation of the XPG1 peptide (^{1057}KRGITNTLEESSLKRKL^{1074}) is more favorable than the bipartite version of this peptide due, among other factors, to the presence of uncharged polar residues in the linker region (Ser, for example) [41]. Finally, serine residues in the flaking region of the NLSs have also been related to the phosphorylation process, which is able to further modulate the nucleocytoplasmatic trafficking [17, 19, 59, 60]. Thus, although, this regulation process was not yet observed for this specific residue of the MLH1, our study shed some light on future studies related to this topic [61].

Human Impα variants display important binding differences with wild and mutants MLH1 NLSs.
Molecular docking was performed in order to understand how the MLH1 NLSs (wild and mutant) bind to both sites of two human variants of importin-α (1 and 3). Due to the higher sequence identity (94%) between MmImpα and HsImpα1, the docking solutions found in HsImpα1 were similar to those found in the crystal structures of MmImpα/MLH1 NLS and MmImpα/mMLH1 NLSs. In the structures of both MmImpα and HsImpα1 receptors, MLH1, MLH1-E475A, MLH1-S467A and MLH1-R470A interact with the major and minor sites, while MLH1-R472K was only found at the major site for both protein receptors (MmImpα and HsImpα1). These data highlight the importance of several studies in the literature that dealt with human NLSs and MmImpα [18, 19, 24, 41, 62-64]. These studies were usually conducted with MmImpα due to the well-established expressing system within this recombinant protein.

In contrast, docking solutions for HsImpα3 and MLH1, MLH1-R472K, MLH1-E475A, and MLH1-R470A NLSs were only found at the major site. A structural comparison between HsImpα3 and HsImpα1 minor sites reveals that three natural substitutions in key residues near the binding site: N350K, K392G and S406I are able to prevent the interaction of the NLSs (Fig. 4). Other structural feature observed is the presence of the natural substitution K435Q (HsImpα3 compared to HsImpα1), which changes the positive charge to neutral and also can interfere with the interaction of NLSs at the minor site (Fig. 5). Additionally, an extensive search in the PDB server did not present any crystal structures of the HsImpα3 in the presence of monopartite peptides at the minor site, suggesting that this site does not participate in the transport of proteins with monopartite NLS. However, docking assays using HsImpα3 with MLH1-S467A NLS produced a favorable solution with -6.4 kcal/mol of Gibbs free energy (Table 3) (Fig. 7A and B). This fact can be related to the very high affinity of MLH1-S467A NLS for MmImpα and the previously cited mutations found in the HsImpα3 sequence that can change the protein surface charges (Fig. 6A).
Interesting, the mutation S467A (NLS) changes the peptide accessibility, causing the interaction of the next residue of the peptide (Asn468) with the Ser407 of the HsImpα3 (Fig. 6 B), which in turn causes the shift of an amino acid position of the NLS peptide (when compared to the MmImpα crystal structure), changing Lys→Arg at the P1’, Arg→Lys at the P2’ and His→Arg at the P3’ positions (Fig. 7 C, D and E).

This study corroborates previously published results [23, 25-27, 32-34, 65] that observed high specificity for the major or minor sites of Impα from particular organisms or variants from the same organism, while other Impα [27] display affinity for both minor and major sites depending on each NLS. All this complexity of different Impα and variety of NLSs is responsible for the selectivity of the importation process of different cargo proteins and is also dependent of other regulation mechanisms, including phosphorylation of particular residues near the NLSs [17, 19, 57, 59].

CONCLUSION

The transport of important components through the nuclear membrane is essential for cellular processes, including gene expression and DNA repair and transcription. The most general importer is the importin α (Impα), which recognizes the proteins that can be transported by the Nuclear Localization Sequence presented in the cargo sequence. This work presents crystallographic, calorimetric and bioinformatic studies, including the Impα from Mus musculus and four MLH1 NLSs with mutations in key residues responsible for interaction with Impα. Based in the data presented here, we were able to move forward on some specific issues that had not been addressed in previous studies. Finally, bioinformatics studies with two different variants of human Impα and their comparison with the well-studied Mus musculus Impα also revealed how specific the nuclear transport mediated by Impα can be.
DATA AVAILABILITY

All data are included within the main manuscript. The four crystal structures of this work are deposited in the Protein Data Bank by access codes: 6WBA (MmlImpa/MLH1-R472K NLS), 6WBB (MmlImpa/MLH1-E475A NLS), 7M60 (MmlImpa/ MLH1-S467A NLS) and 6WBC (MmlImpa/ MLH1-R470A NLS).

COMPETING INTERESTS

The authors declare that there are no competing interests associated with the manuscript.

FUNDING

This work was financially supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil).

ACKNOWLEDGMENTS

We thank the use of the Laboratório Nacional de Luz Síncrotron (LNLS, Brazil) for X-ray diffractometry experiments.

CREDIT AUTHOR CONTRIBUTION

HCO performed crystallographic, calorimetric and molecular docking assays under supervision MRMF. TDS performed protein purification and crystallographic studies under supervision ACB and MRMF. GHMS performed bioinformatic studies. CAF assist with
crystallographic and calorimetric assays. IRM assist crystallographic and bioinformatics assays. HCO, GHMS and MRMF wrote the manuscript with inputs of all others authors. All authors further revised the manuscript.

REFERENCES

1. Gorlich, D. and Laskey, R. A. (1995) Roles of importin in nuclear protein import. Cold Spring Harb Symp Quant Biol. 60, 695-699
2. Gorlich, D. and Kutay, U. (1999) Transport between the cell nucleus and the cytoplasm. Annu Rev Cell Dev Biol. 15, 607-660
3. Conti, E. and Kuriyan, J. (2000) Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure. 8, 329-338
4. Lange, A., Mills, R. E., Lange, C. J., Stewart, M., Devine, S. E. and Corbett, A. H. (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. J Biol Chem. 282, 5101-5105
5. Marfori, M., Mynott, A., Ellis, J. J., Mehdi, A. M., Saunders, N. F., Curmi, P. M., Forwood, J. K., Boden, M. and Kobe, B. (2011) Molecular basis for specificity of nuclear import and prediction of nuclear localization. Biochim Biophys Acta. 1813, 1562-1577
6. Chang, C. W., Coumago, R. M., Williams, S. J., Boden, M. and Kobe, B. (2013) 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-alpha. Plant Signal Behav. 8
7. Conti, E., Uy, M., Leighton, L., Blobel, G. and Kuriyan, J. (1998) Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell. 94, 193-204
8. Fontes, M. R., Teh, T. and Kobe, B. (2000) Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. J Mol Biol. 297, 1183-1194
9. Fontes, M. R., Teh, T., Jans, D., Brinkworth, R. I. and Kobe, B. (2003) Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha. J Biol Chem. 278, 27981-27987
10. Dingwall, C. and Laskey, R. A. (1991) Nuclear targeting sequences--a consensus? Trends Biochem Sci. 16, 478-481
11. Gorlich, D., Prehn, S., Laskey, R. A. and Hartmann, E. (1994) Isolation of a protein that is essential for the first step of nuclear protein import. Cell. 79, 767-778
12. Kalderon, D., Richardson, W. D., Markham, A. F. and Smith, A. E. (1984) Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature. 311, 33-38
13. Chelsky, D., Ralph, R. and Jonak, G. (1989) Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol Cell Biol. 9, 2487-2492
14. Hodel, M. R., Corbett, A. H. and Hodel, A. E. (2001) Dissection of a nuclear localization signal. J Biol Chem. 276, 1317-1325
15. Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M. and Yanagawa, H. (2009) Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J Biol Chem. 284, 478-485
16 Marfori, M., Lonhienne, T. G., Forwood, J. K. and Kobe, B. (2012) Structural basis of high-affinity nuclear localization signal interactions with importin-alpha. Traffic. 13, 532-548
17 Fontes, M. R., Teh, T., Toth, G., John, A., Pavo, I., Jans, D. A. and Kobe, B. (2003) Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40 large T-antigen nuclear localization sequences by importin-alpha. Biochem J. 375, 339-349
18 Takeda, A. A., de Barros, A. C., Chang, C. W., Kobe, B. and Fontes, M. R. (2011) Structural basis of importin-alpha-mediated nuclear transport for Ku70 and Ku80. J Mol Biol. 412, 226-234
19 Rona, G., Marfori, M., Borsos, M., Scheer, I., Takacs, E., Toth, J., Babos, F., Magyar, A., Erdei, A., Bozoky, Z., Buday, L., Kobe, B. and Vertessy, B. G. (2013) Phosphorylation adjacent to the nuclear localization signal of human dUTPase abolishes nuclear import: structural and mechanistic insights. Acta Crystallogr D Biol Crystallogr. 69, 2495-2505
20 Smith, K. M., Tsimbalyuk, S., Edwards, M. R., Cross, E. M., Batra, J., Soares da Costa, T. P., Aragao, D., Basler, C. F. and Forwood, J. K. (2018) Structural basis for importin alpha 3 specificity of W proteins in Hendra and Nipah viruses. Nat Commun. 9, 3703
21 Christie, M., Chang, C. W., Rona, G., Smith, K. M., Stewart, A. G., Takeda, A. A., Fontes, M. R., Stewart, M., Vertessy, B. G., Forwood, J. K. and Kobe, B. (2016) Structural Biology and Regulation of Protein Import into the Nucleus. J Mol Biol. 428, 2060-2090
22 Giesecke, A. and Stewart, M. (2010) Novel binding of the mitotic regulator TPX2 (target protein for Xenopus kinesin-like protein 2) to importin-alpha. J Biol Chem. 285, 17628-17635
23 Chang, C. W., Counago, R. L., Williams, S. J., Boden, M. and Kobe, B. (2012) Crystal structure of rice importin-alpha and structural basis of its interaction with plant-specific nuclear localization signals. Plant Cell. 24, 5074-5088
24 Lott, K. and Cingolani, G. (2011) The importin beta binding domain as a master regulator of nucleocytoplasmic transport. Biochim Biophys Acta. 1813, 1578-1592
25 Bernardes, N. E., Takeda, A. A., Dreyer, T. R., Freitas, F. Z., Bertolini, M. C. and Fontes, M. R. (2015) Structure of Importin-alpha from a Filamentous Fungus in Complex with a Classical Nuclear Localization Signal. PLoS One. 10, e0128687
26 Bernardes, N. E., Takeda, A. A. S., Dreyer, T. R., Cupertino, F. B., Virgilio, S., Pante, N., Bertolini, M. C. and Fontes, M. R. M. (2017) Nuclear transport of the Neurospora crassa NIT-2 transcription factor is mediated by importin-alpha. Biochem J. 474, 4091-4104
27 Bernardes, N. E., Fukuda, C. A., da Silva, T. D., de Oliveira, H. C., de Barros, A. C., Dreyer, T. R., Bertolini, M. C. and Fontes, M. R. M. (2020) Comparative study of the interactions between fungal transcription factor nuclear localization sequences with mammalian and fungal importin-alpha. Sci Rep. 10, 1458
28 Kohler, M., Haller, H. and Hartmann, E. (1999) Nuclear protein transport pathways. Exp Nephrol. 7, 290-294
29 Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T. and Adam, S. A. (2004) Importin alpha: a multipurpose nuclear-transport receptor. Trends Cell Biol. 14, 505-514
30 Mason, D. A., Stage, D. E. and Goldfarb, D. S. (2009) Evolution of the metazoan-specific importin alpha gene family. J Mol Evol. 68, 351-365
31 Rother, F., Shmidt, T., Popova, E., Krivokharchenko, A., Hugel, S., Vilianovich, L., Ridders, M., Tenner, K., Alenina, N., Kohler, M., Hartmann, E. and Bader, M. (2011) Importin alpha7 is essential for zygotic genome activation and early mouse development. PLoS One. 6, e18310
32 Sekimoto, T., Imamoto, N., Nakajima, K., Hirano, T. and Yoneda, Y. (1997) Extracellular signal-dependent nuclear import of Stat1 is mediated by nuclear pore-targeting complex formation with NPI-1, but not Rch1. EMBO J. 16, 7067-7077
33 Talcott, B. and Moore, M. S. (2000) The nuclear import of RCC1 requires a specific nuclear localization sequence receptor, karyopherin alpha3/Qip. J Biol Chem. 275, 10099-10104
34 Ma, J. and Cao, X. (2006) Regulation of Stat3 nuclear import by importin alpha5 and importin alpha7 via two different functional sequence elements. Cell Signal. 18, 1117-1126
35 Pumroy, R. A. and Cingolani, G. (2014) Jamming up the "beta-staple": regulation of SIRT1 activity by its C-terminal regulatory segment (CTR). J Mol Biol. 426, 507-509
36 Aratani, S., Oishi, T., Fujita, H., Nakazawa, M., Fujii, R., Imamoto, N., Yoneda, Y., Fukamizu, A. and Nakajima, T. (2006) The nuclear import of RNA helicase A is mediated by importin-alpha3. Biochem Biophys Res Commun. 340, 125-133
37 Pumroy, R. A. and Cingolani, G. (2015) Diversification of importin-alpha isoforms in cellular trafficking and disease states. Biochem J. 466, 13-28
38 Pumroy, R. A., Ke, S., Hart, D. J., Zachariae, U. and Cingolani, G. (2015) Molecular determinants for nuclear import of influenza A PB2 by importin alpha isoforms 3 and 7. Structure. 23, 374-384
39 Fukui, K. (2010) DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids. 2010
40 Chatterjee, N. and Walker, G. C. (2017) Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen. 58, 235-263
41 Barros, A. C., Takeda, A. A., Dreyer, T. R., Velazquez-Campoy, A., Kobe, B. and Fontes, M. R. (2016) 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. J Mol Biol. 428, 2120-2131
42 de Barros, A. C., Takeda, A. A. S., Dreyer, T. R., Velazquez-Campoy, A., Kobe, B. and Fontes, M. R. M. (2018) DNA mismatch repair proteins MLH1 and PMS2 can be imported to the nucleus by a classical nuclear import pathway. Biochimie. 146, 87-96
43 Wu, X., Platt, J. L. and Cascalho, M. (2003) Dimerization of MLH1 and PMS2 limits nuclear localization of MutLalpha. Mol Cell Biol. 23, 3320-3328
44 Teh, T., Tiganis, T. and Kobe, B. (1999) Crystallization of importin alpha, the nuclear-import receptor. Acta Crystallogr D Biol Crystallogr. 55, 561-563
45 Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227, 680-685
46 Vega, S., Abian, O. and Velazquez-Campoy, A. (2015) A unified framework based on the binding polynomial for characterizing biological systems by isothermal titration calorimetry. Methods. 76, 99-115
47 Watts, A. (1992) Crystallization of Nucleic Acids and Proteins. A Practical Approach: Edited by A. Ducruix and R. Giegé. FEBS J. 319
48 Kabsch, W. (2010) Xds. Acta Crystallogr D Biol Crystallogr. 66, 125-132
49 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 60, 2126-2132
50 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 66, 213-221
Zheng, W., Wang, R., Liu, X., Tian, S., Yao, B., Chen, A., Jin, S. and Li, Y. (2018) Structural insights into the nuclear import of the histone acetyltransferase males-absent-on-the-first by importin alpha1. Traffic. 19, 19-28

Schrodinger, L. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.

Laskowski, R. A. and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model. 51, 2778-2786

Dallakyan, S. and Olson, A. J. (2015) Small-molecule library screening by docking with PyRx. Methods Mol Biol. 1263, 243-250

Sankhala, R. S., Lokareddy, R. K. and Cingolani, G. (2016) Divergent Evolution of Nuclear Localization Signal Sequences in Herpesvirus Terminase Subunits. J Biol Chem. 291, 11420-11433

Kobe, B. (1999) Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. Nat Struct Biol. 6, 388-397

Peifer, M., Pai, L. M. and Casey, M. (1994) Phosphorylation of the Drosophila adherens junction protein Armadillo: roles for wingless signal and zeste-white 3 kinase. Dev Biol. 166, 543-556

Rona, G., Palinkas, H. L., Borsos, M., Horvath, A., Scheer, I., Benedek, A., Nagy, G. N., Zagyva, I. and Vertessy, B. G. (2014) NLS copy-number variation governs efficiency of nuclear import--case study on dUTPases. FEBS J. 281, 5463-5478

Jans, D. A. (1995) The regulation of protein transport to the nucleus by phosphorylation. Biochem J. 311 (Pt 3), 705-716

Xiao, C. Y., Hubner, S. and Jans, D. A. (1997) SV40 large tumor antigen nuclear import is regulated by the double-stranded DNA-dependent protein kinase site (serine 120) flanking the nuclear localization sequence. J Biol Chem. 272, 22191-22198

Wessbecher, I. M., Hinrichsen, I., Funke, S., Oellerich, T., Plotz, G., Zeuzem, S., Grus, F. H., Biondi, R. M. and Brieger, A. (2018) DNA mismatch repair activity of MutLalpha is regulated by CK2-dependent phosphorylation of MLH1 (S477). Mol Carcinog. 57, 1723-1734

Mynott, A. V., Harrop, S. J., Brown, L. J., Breit, S. N., Kobe, B. and Curmi, P. M. (2011) Crystal structure of importin-alpha bound to a peptide bearing the nuclear localisation signal from chloride intracellular channel protein 4. FEBS J. 278, 1662-1675

Chang, C. C., Chen, C. J., Grauffel, C., Pien, Y. C., Lim, C., Tsai, S. Y. and Hsia, K. C. (2019) Ran pathway-independent regulation of mitotic Golgi disassembly by Importin-alpha. Nat Commun. 10, 4307

Jung, H., Takeshima, T., Nakagawa, T., MacMillan, K. S., Wynn, R. M., Wang, H., Sakiyama, H., Wei, S., Li, Y., Bruick, R. K., Posner, B. A., De Brabander, J. K. and Uyeda, K. (2020) The structure of importin alpha and the nuclear localization peptide of ChREBP, and small compound inhibitors of ChREBP-importin alpha interactions. Biochem J. 477, 3253-3269

Chang, C. W., Williams, S. J., Counago, R. M. and Kobe, B. (2014) Structural basis of interaction of bipartite nuclear localization signal from Agrobacterium VirD2 with rice importin-alpha. Mol Plant. 7, 1061-1064

FIGURE LEGENDS
**Figure 1:** Crystal structures of the complexes MmImpα/mMLH1 NLSs and the omit electron density map corresponding to the NLS at the binding sites. The peptides are represented as sticks in the electron density map. (A) MLH1-R472K NLS (orange) at the major site. (B) MLH1-E475A NLS (cyan) at the major and minor sites. (C) MLH1-S467A NLS (green) at both sites. (D) MLH1-R470A NLS (salmon) at the major and minor sites. All the maps are depicted in coefficients $2|F_{\text{obs}}|−|F_{\text{calc}}|$ corresponding to the NLSs molecules, contoured at 2.5σ.

**Figure 2:** Schematic diagram of the interactions between the MLH1 and mMLH1 NLSs and the MmImpα. (A) MLH1 NLS in the major and (B) minor sites; (C) MLH1-R472K NLS in the major site; (D) MLH1-E475A NLS in the major and (E) minor sites; (F) MLH1-S467A NLS in the major and (G) minor sites; (H) MLH1-R470A NLS in the major and (I) minor sites. The peptide backbones are depicted in purple (MLH1 NLS), orange (MLH1-R472K NLS), cyan (MLH1-E475A NLS), green (MLH1-S467A NLS) and salmon (MLH1-R470A NLS) with the residues identified by the one-letter code. The MmImpα side chain residues which are interacting with the peptide are represented with their names and different colors. Hydrogen bonds are depicted as broken lines and the hydrophobic contacts are shown as arcs with radiating spokes.

**Figure 3:** Isothermal calorimetric titration of the NLSs into MmImpα. The upper panel depicts the raw data thermogram (thermal power/time) and the lower panel shows the binding isotherm (ligand-normalized integrated heat as a function of the molar ratio). The ITC experiments were performed using MmImpα and the mMLH1 NLSs: (A) MLH1-R472K NLS; (B) MLH1-E475A NLS; (C) MLH1-S467A NLS; and (D) MLH1-R470A NLS.
**Figure 4:** Minor site amino acid sequence alignment of Impα from *Mus musculus* (*MmImpα*) and Impα1 and Impα3 from *Homo sapiens* (*HsImpα1* and *HsImpα3*). The numbers above of the line correspond to mutated residues found in the *HsImpα3*. The superposition of *MmImpα* (yellow), *HsImpα1* (cyan) and *HsImpα3* (orange) amino acids highlighted are represented as sticks.

**Figure 5:** Electrostatic surface charged distribution of the Impα from *Mus musculus* and Impα1 and Impα3 from *Homo sapiens*. The structures presented in the left column shows the major site, pointed using the yellow arrows. The structures in the right column are rotated 120° and the green arrows are pointing the different charges in the minor sites. (A) *MmImpα*, (B) *HsImpα1* and (C) *HsImpα3*.

**Figure 6:** Docking simulations representing the interaction between *HsImpα3* and MLH1-S467A NLS and. (A) Electrostatic surface charged distribution of the *HsImpα3* and the molecular docking solution of the MLH1-S467A NLS (orange sticks). (B) Hydrogen bond (2.8Å) interaction between Asn468 from NLS and Ser407 from *HsImpα3*.

**Figure 7:** Comparison of the MLH1-S467A NLS binding at the minor site from different Impα. (A) MLH1-S467A bound to *MmImpα* crystal structure. (B) Docking solution with the MLH1-S467A NLS and *HsImpα3*. C. MLH1-S467A NLS with the P1’–P4’ amino acids highlighted as determined by *MmImpα*. (D) MLH1-S467A NLS position in *HsImpα3*, highlighting shifting position of the P1’–P4’ amino acids, as determined by *HsImpα3*. (E) Superposition of the MLH1-S467A NLS peptide at the P1’–P4’ positions bound to *MmImpα* (yellow) and *HsImpα3* (orange) structures.
Table 1: X-ray data collection and refinement statistics of the complexes MmImpα/MLH1 mutated NLS peptides.

|                          | MmImpα/MLH1-R472K | MmImpα/MLHA1-E475A | MmImpα/MLHA1-S467A | MmImpα/MLHA1-R470A |
|--------------------------|-------------------|-------------------|-------------------|-------------------|
| Unit cell parameters (Å) | a= 78.70; b= 90.23; c= 100.08 | a= 78.27; b= 90.56; c= 99.60 | a= 78.39; b= 90.35; c= 99.33 |
| Space group              | P2₁2₁2₁           | P2₁2₁2₁           | P2₁2₁2₁           | P2₁2₁2₁           |
| Resolution (Å)           | 43.76-2.15 (2.23-2.15) | 43.91-2.66 (2.76-2.66) | 42.04-2.30 (2.38-2.30) | 41.95-2.15 (2.23-2.15) |
| Unique reflections       | 39385 (3862)      | 21004 (2042)      | 31428 (3024)      | 38621 (3719)      |
| Completeness (%)         | 99.76 (99.36)     | 99.80 (99.36)     | 98.12 (96.31)     | 99.95 (97.15)     |
| R_merge                  | 0.086 (1.75)      | 0.151 (1.12)      | 0.088 (0.97)      | 0.099 (1.33)      |
| I/σ(I)                   | 24.40 (1.53)      | 24.0 (2.49)       | 24.27 (2.72)      | 18.29 (0.97)      |
| Multiplicity             | 13.1 (13.2)       | 13.1 (13.4)       | 11.0 (11.3)       | 13.2 (13.1)       |
| CC ½                     | 1.0 (0.84)        | 0.99 (0.80)       | 0.99 (0.80)       | 0.99 (0.80)       |
| Total reflections        | 514128            | 274757            | 346941            | 508565            |
| R_work (%)               | 17.97             | 18.34             | 17.37             | 17.20             |
| R_free (%)               | 21.15             | 22.77             | 20.68             | 19.73             |
| Number of molecules:     |                   |                   |                   |                   |
| Protein                  | 1                 | 1                 | 1                 | 1                 |
| Peptide                  | 1                 | 2                 | 2                 | 2                 |
| Solvent                  | 244               | 70                | 230               | 249               |
| Average B factor (Å²):   |                   |                   |                   |                   |
| Protein                  | 51.34             | 53.69             | 48.98             | 51.50             |
| NLS in Major site        | 52.30             | 49.45             | 50.59             | 57.32             |
| P1 - P5                  | 43.20             | 41.40             | 41.94             | 43.08             |
| NLS in Minor Site        | 68.67             | 67.28             | 68.40             | 68.40             |
| P1'' - P4''              | 61.30             | 60.95             | 62.25             | 62.25             |
| RMS (bonds)              | 0.006             | 0.005             | 0.005             | 0.005             |
| RMS (angles)             | 1.08              | 1.05              | 1.08              | 1.08              |
| Clashcore                | 3.74              | 5.09              | 5.69              | 2.92              |
| Ramachandran plot (%)    |                   |                   |                   |                   |
| Favored                  | 97.22             | 98.17             | 98.41             | 98.18             |
| Allowed                  | 2.55              | 1.60              | 1.37              | 1.82              |
| Outliers                 | 0.23              | 0.23              | 0.23              | 0.00              |

Numbers in parenthesis correspond to the highest resolution data.

\[ R_{\text{merge}} = \frac{\sum_h \sum_k \sum_l |I_{h,k,l} - I_{h,k,l}|}{\sum_h \sum_k \sum_l I_{h,k,l}} \] at where \( I_{h,k,l} \) is the intensity of each individual measure of the reflection with Miller indices \( h, k, \) and \( l \), and \( < I_{h,k,l} > \) is the average intensity of that reflection. Calculated for \( I > 3\sigma (I) \).
Table 2: Thermodynamic-binding data of MmImpα and MLH1 NLSs. The estimated specific dissociation constants ($k_d$) and enthalpy changes ($\Delta H$) were determined for the binding of MLH1 NLSs to MmImpα at 20°C, by binding polynomials.

| Peptide   | Kd (μM)   | $\Delta H$ (kcal/mol) |
|-----------|-----------|------------------------|
| MLH1*     | Major site | 0.10 ± 0.02            | -6.5 ± 0.70       |
|           | Minor Site | 1.5 ± 0.20             | -2.9 ± 0.70       |
| MLH1-R472K| 2.22 ± 0.41|                        | -5.21 ± 0.18      |
|           | Major Site | 0.69 ± 0.07            | -11.08 ± 0.33     |
| MLH1-E475A| Minor Site | 3.76 ± 0.04            | -1.42 ± 0.36      |
| MLH1-S467A| Major Site | 0.04 ± 0.01            | -9.19 ± 0.26      |
|           | Minor Site | 0.97 ± 0.04            | -4.13 ± 0.33      |
| MLH1-R470A| Major Site | 3.54 ± 0.59            | -10.05 ± 0.94     |
|           | Minor Site | 7.7 ± 0.54             | -3.26 ± 1.04      |

*De Barros, 2018
**Table 3:** Molecular docking solutions for the binding of MLH1 and mMLH1 in the major site and minor site of homologous human importin α.

|                | Binding site | ΔG (kcal/mol) | Theoretical Kd (µM)* |
|----------------|--------------|---------------|----------------------|
| MLH1           | Major site   | -6.5          | 17.2                 |
| MLH1           | Minor Site   | -6.3          | 24.1                 |
| MLH1-R472K     | Major site   | -6.4          | 20.3                 |
| MLH1-E475A     | Major site   | -7.1          | 6.2                  |
| MLH1-S467A     | Minor Site   | -7.0          | 7.4                  |
| MLH1-R470A     | Minor Site   | -6.9          | 8.7                  |
| MLH1-R470A     | Major site   | -6.5          | 17.2                 |
| MLH1-R470A     | Minor Site   | -6.5          | 17.2                 |
| MLH1-R472K     | Major site   | -6.4          | 20.3                 |
| MLH1-E475A     | Major site   | -7.8          | 1.9                  |
| MLH1-S467A     | Minor Site   | -6.4          | 20.3                 |
| MLH1-R470A     | Major site   | -6.6          | 14.5                 |

Theoretical Kd value was calculated using: \( \Delta G = R \cdot T \cdot \ln(K_d) \), where \( \Delta G \) is Gibbs free Energy obtained in the docking simulations, \( R \) is the ideal gas constant, \( T \) is the temperature (298.15 K).
Table 4: MLH1 and mutant MLH1 NLS amino acid sequences and monopartite consensus for the binding NLSs to Importin α proteins.

| NLS       | Minor Site | Major Site | Consensus               |
|-----------|------------|------------|-------------------------|
|           | P1' P2' P3' P4' | P1' P2' P3' | Residues highlighted in red correspond to the consensus positions. |
| MLH1      | S N P R K R H R E | S N P R K R H R E D |                                      |
| MLH1-R472K| S N P R K R H R E | S N P R K K H R E D |                                      |
| MLH1-E475A| A N P R K R H R E | A N P R K R H R A D |                                      |
| MLH1-S467A| S N P A K R H R E | S N P A K R H R E D |                                      |
| MLH1-R470A| S N P R K K R K V | A A P P K K K R K R V E |                                      |
| SV40      | P K K K K R V | A A P P K K K R K R V |                                      |
| CN-SV40   | K R K K V | P A A K R V K L D |                                      |
| c-Myc     | K R V K L | P A A K R V K L D |                                      |
| pUL15     | P P K K R A | G P P K K R A K V D |                                      |
| pUL56     | T R K R P R | A T R K R P R R R A |                                      |
| XPG1      | S L K R K R | S L K R K R |                                      |
| XPG2      | R K R K T R | G K K R R K L R |                                      |
| Ku70      | N E G S G S K R P K |                                      |
| Ku80      | D G P T A K K L K T E |                                      |
| PLSCR1    | G K I S K H W |                                      |
| PLSCR4    | S I I R K W N |                                      |
| mPetTM    | K K R R E A | F K K K R R E A |                                      |
| PMS2      | T P N T K R F K K E E |                                      |
| TPX2      | K R K H | P V K M I K |                                      |
| Consensus | K K/R X K/R |                                      |
