UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

Instituto de Biociências
de Botucatu - Unesp

MARCOS TADEU GERALDO

Caracterização *in silico* dos mecanismos de interação entre sequências de localização nuclear e Importina-α

Botucatu-SP

2016

MARCOS TADEU GERALDO

Caracterização *in silico* dos mecanismos de interação entre sequências de localização nuclear e Importina-α

Tese apresentada ao Instituto de Biociências, Campus de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Ciências Biológicas (Genética).

Universidade Estadual Paulista – UNESP

Instituto de Biociências de Botucatu

Programa de Pós-Graduação em Ciências Biológicas (Genética)

Orientador: Ney Lemke

Coorientadora: Agnes Alessandra Sekijima Takeda

Botucatu-SP

2016

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Geraldo, Marcos Tadeu.
Caracterização in silico dos mecanismos de interação entre sequências de localização nuclear e Importina -α / Marcos Tadeu Geraldo. - Botucatu, 2016

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Ney Lemke
Coorientador: Agnes Alessandra Sekijima Takeda
Capes: 20205007

1. Dinâmica molecular. 2. Transporte ativo do núcleo celular. 3. Carioferinas. 4. Cristalografia de raio X.

Palavras-chave: Dinâmica molecular; Importação nuclear; Importina -α; Modos normais; Sequência de localização nuclear.

Este trabalho é dedicado aos meus queridos pais
José Eduardo e Maria Helena e à minha amada noiva Natália Totti,
por tudo o que fizeram e ainda fazem por mim,
sempre com muito amor.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelas bolsas de doutorado e de estágio no exterior (processos: 2012/19447-2 e 2014/21976-9). Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de doutorado (processo: 142110/2012-4) no início do curso.

Ao meu orientador Ney Lemke, por participar diretamente nas discussões e análises dos meus resultados, sempre contribuindo ativamente, além de ser um orientador exemplar por nos oferecer algo muito importante para a nossa carreira científica: autonomia e liberdade para tomadas de decisão. Estas qualidades pretendo levar durante toda a minha vida profissional.

À minha coorientadora Agnes Takeda por ter sido a pessoa que acompanhou de perto o desenvolvimento do meu trabalho, além dos seus ensinamentos e conselhos que, com certeza, foram fundamentais para a produção desta tese.

Ao professor Antônio Sérgio (UFABC) pela colaboração, me ensinando sobre a técnica de análise de modos normais que utilizei em meu trabalho.

Ao professor David Perahia (ENS – Cachan, França) por me receber e ser o supervisor do meu estágio em seu laboratório por 5 meses, me auxiliando diretamente na preparação dos *scripts* para as análises dos dados de modos normais.

A todos do Laboratório de Bioinformática e Biofísica Computacional pelo companheirismo no dia a dia.

Ao Programa de Pós-Graduação em Ciências Biológicas (Genética) do Instituto de Biociências de Botucatu.

Ao Departamento de Física e Biofísica pelo apoio institucional e pelos espaços físicos.

A todos da Seção de Pós-Graduação por toda ajuda referente ao controle da documentação e solicitude para atender a todas as minhas dúvidas.

À minha família, em especial aos meus pais José Eduardo e Maria Helena por todo o carinho, apoio e dedicação em todos os passos de minha vida. A eles sou eternamente grato! Agradeço também aos meus irmãos Cássia e Eduardo pelo carinho, risadas e apoio durante minha vida.

Aos meus sogros Wanda e Silvio por me receberem muito bem, além de terem sempre carinho e torcerem muito por mim.

À minha noiva Natália Totti por seu amor, carinho e atenção em todos os momentos

da minha vida desde que nos conhecemos. Não tenho palavras para expressar o quanto sou abençoado por estar ao seu lado! Todo o seu companheirismo e amor dedicados certamente foram essenciais durante estes 4 anos de doutorado, e sei que para as próximas etapas, tanto da minha vida científica e pessoal, você estará comigo. Muito obrigado, meu amor! Amo você imensamente! E, é claro, um agradecimento à nossa filhinha felina, Greta, que nos ensina sempre a meditar, com aquela sua carinha de paz e tranquilidade.

Por fim, desejo concluir meus agradecimentos lembrando que apesar dos problemas e entraves que possam surgir durante qualquer atividade, sempre precisamos da humildade e atenção necessária para perceber tudo de bom que também nos acontece. E por isso, tenho um grande sentimento de gratidão por tudo que vivenciei durante o meu curso de doutoramento e estou muito satisfeito pelo o que foi produzido!

When you complain,
you make yourself a victim.
Leave the situation,
change the situation or accept it.
All else is madness.
Eckhart Tolle

Resumo

Os sistemas de importação nuclear são responsáveis pelo intercâmbio entre o citoplasma e o núcleo da célula, permitindo que proteínas com função nuclear migrem através da membrana que separa essas duas regiões. A via de importação mais estudada é a via clássica de importação nuclear mediada pela Importina-α (Impα). A Impα é uma proteína solenóide, composta por repetições em *tandem* do motivo *Armadillo* (ARM) que formam uma estrutura longa e contorcida, com pequenos arcabouços ao longo do eixo da proteína. As sequências de localização nuclear clássicas (cNLSs) presentes nas proteínas-alvo de importação são compostas por resíduos carregados positivamente e estabelecem pontes salinas, ligações de hidrogênio e contatos hidrofóbicos com esses arcabouços da Impα. Esse reconhecimento pode ocorrer em um ou em dois sítios da Impα, caracterizando a cNLS como monopartida ou bipartida, respectivamente. A maioria das informações estruturais do complexo cNLS-Impα provém de dados de cristalografia e pouco se sabe sobre a dinâmica conformacional deste sistema. Uma abordagem para tratar da dinâmica de um sistema é o uso de técnicas de simulação de biomoléculas, tais como dinâmica molecular e análise de modos normais. Com base nessas técnicas de simulação, o presente estudo teve como objetivo compreender os mecanismos de interação e dinâmica conformacional envolvidos no reconhecimento de cNLSs pela Impα. Particularmente, este trabalho focou nas cNLSs das proteínas Nucleoplasmina e Ku70 complexadas com a Impα. O estudo com a Nucleoplasmina determinou dois movimentos principais da Impα que podem estar associados na função de reconhecimento de cNLSs: dobramento e torção. Os movimentos de dobramento podem estar envolvidos na entrada da cNLS e na acomodação da proteína que a contém, dependendo do seu tamanho, enquanto que os movimentos de torção podem estar envolvidos no reconhecimento da cNLS e na sua acomodação aos sítios de ligação da Impα. Além disso, resíduos correspondentes à região *linker*, situada entre os grupos de resíduos básicos da cNLS bipartida, também podem auxiliar no ajuste da cNLS na Impα. Por fim, o estudo com a Ku70 verificou, com base na análise de contatos e correlações na interface peptídeo-proteína e nos perfis geométricos da Impα, que esta não se ligaria à Impα como uma cNLS bipartida. Em conclusão, os dados aqui obtidos podem auxiliar no entendimento das afinidades entre as cNLSs já descritas, como também na análise de outras potenciais cNLSs.

Palavras-chave: importação nuclear. importina-α. sequência de localização nuclear. dinâmica molecular. modos normais.

Abstract

Nuclear import systems are responsible for the exchange between the cytoplasm and the nucleus of a cell, allowing nuclear proteins to migrate through the membrane that separates these two regions. The most studied import pathway is the classical nuclear import mediated by Importin-α (Impα). Impα is a solenoid protein consisting of tandem repeats of the Armadillo (ARM) motif, forming an extended and twisted structure with small grooves along the protein axis. The classical nuclear localization sequences (cNLSs) of cargo proteins are composed of positively charged residues and establish salt bridges, hydrogen bonds and hydrophobic contacts with the grooves of Impα. Such recognition can occur at one or two sites of Impα, thus characterizing the cNLS as monopartite or bipartite, respectively. Most structural information of the cNLS-Impα complex is from crystallographic data and little is known about the conformational dynamics of this system. One approach to address the dynamics of a system is the use of biomolecular simulation techniques such as molecular dynamics and normal modes analysis. Based on these techniques, this study aimed to understand the mechanisms of interaction and conformational dynamics involved in the recognition of cNLSs by Impα. In particular, this work focused on the cNLSs of Nucleoplasmin and Ku70 proteins complexed with Impα. The study of Nucleoplasmin determined two main motions of Impα that may be associated to the cNLS recognition: bending and twisting. The bending motion may be involved in the cNLS entrance and the accommodation of cargo protein depending on its size, whereas the twisting motions may be involved in the cNLS recognition and accommodation into the binding sites of Impα. Furthermore, the residues corresponding to the linker region, situated between the groups of basic residues from the bipartite cNLS, may also assist in setting the cNLS into Impα. Finally, the study of Ku70 verified, based on the contact and correlation analyses of the peptide-protein interface and the geometric profiling of Impα, that its peptide would not bind to Impα as a bipartite cNLS. In conclusion, the data obtained here may help in understanding the affinities between the cNLSs already described, as well as the analysis of other potential cNLSs.

Keywords: nuclear import. importin-α. nuclear localization sequence. molecular dynamics. normal modes.

Sumário

1 Introdução

1.1 Transporte núcleo-citoplasma

Um dos principais passos evolutivos que ocorreram no nível celular foi a aquisição de compartimentos internos, delimitados por membranas lipídicas. Internalização da membrana e a formação de organelas forneceram uma grande vantagem evolutiva pela realização simultânea de diferentes funções dentro destes compartimentos distintos, permitindo que as células aumentassem a sua robustez e complexidade (FLOCH; PALANCADE; DOYE, 2013).

O núcleo celular é a organela que distingue células eucarióticas de procarióticas e o seu limite, a membrana nuclear, permite que as células eucarióticas regulem espacial e temporalmente as distintas fases de expressão do genoma (transcrição do mRNA, maturação e tradução), pois ela se torna uma barreira importante para proteínas e RNA, já que ambos precisam ser transportados de forma regulada (FLOCH; PALANCADE; DOYE, 2013). Além da transferência de grandes quantidades de proteínas nucleares constitutivas, tais como as histonas, alterações na expressão gênica em geral requerem a entrada controlada no núcleo de moléculas de sinalização (STEWART, 2007). No entanto, o conteúdo nuclear não é totalmente isolado a partir do citoplasma, graças a um sistema chamado complexo poro nuclear (NPC). Macromoléculas maiores que aproximadamente 40 kilo-Daltons (kDa) são transportadas ativamente através do envelope nuclear por meio dos NPCs, utilizando fatores de transporte solúveis ou moléculas transportadoras que circulam entre o citoplasma e o núcleo (GÖRLICH; KUTAY, 1999; FAHRENKROG; AEBI, 2003; PEMBERTON; PASCHAL, 2005). Os NPCs são grandes complexos de macromoléculas que formam um canal através da membrana nuclear e possibilitam a troca de moléculas de maneira bidirecional entre o citoplasma e o núcleo com um diâmetro de canal limitante variando de 25 a 30 nanômetros (nm) (FELDHERR; AKIN; COHEN, 2001). NPCs são construídos a partir de múltiplas cópias de aproximadamente 30 diferentes proteínas denominadas nucleoporinas (ROUT; WENTE, 1994; ROUT et al., 2000; CRONSHAW et al., 2002).

A translocação de proteínas através de NPCs requer o suporte adicional de proteínas ou fatores de transporte. Muitas dessas transportadoras pertencem à superfamília das β-carioferinas (β-Kap) (ou Importina-β, Impβ). Todos os membros da família β-Kap são construídos de repetições em *tandem* HEAT (ANDRADE; BORK, 1995), cada um dos quais contém 40-45 aminoácidos que formam duas α-hélices antiparalelas ligadas por um *loop*. Esta estrutura repetitiva aloca as β-Kaps na categoria das proteínas em solenóide (KOBE; KAJAVA, 2000), que aparece com destaque entre proteínas envolvidas no transporte nucleocitoplasmático.

Um componente-chave adicional das vias de transporte nuclear é a Ran-GTPase. Os ciclos da Ran entre os estados ligados a GDP e GTP (GÖRLICH et al., 1996), e o estado do nucleotídeo ligado são determinados por proteínas Ran reguladoras, incluindo a Ran-GEF (*Fator de troca da guanosina*: induz a Ran-GDP a trocar o GDP por GTP) (BISCHOFF; PONSTINGL, 1991) e Ran-GAP (*Proteína ativadora da GTPase*: induz a proteína Ran a hidrolisar o GTP) (BISCHOFF et al., 1994). O gradiente da Ran-GDP/GTP, gerado por Ran-GEF e Ran-GAP que estão localizados no núcleo e no citoplasma, respectivamente, estabelece a direção das vias de transporte nucleocitoplasmático. Como resultado, os receptores de importação que se ligam no citoplasma a alguma molécula a ser importada, podem liberá-la no núcleo por meio da ligação à Ran-GTP, enquanto que os receptores de exportação que se ligam no núcleo a alguma molécula a ser exportada, podem se ligar simultaneamente à Ran-GTP e, posteriormente, podem liberá-la no citoplasma após a hidrólise do GTP (GÖRLICH et al., 1996; MOROIANU; BLOBEL; RADU, 1996; REXACH; BLOBEL, 1995).

A localização nuclear de uma proteína está geralmente associada à presença de sequências de localização nuclear (NLS). O primeiro sinal de direcionamento nuclear descoberto, e o melhor caracterizado até o momento, é a sequência de localização nuclear clássica (cNLS) reconhecida pela proteína Importina-α (Impα) (ou Carioferina-α) (GÖRLICH et al., 1995). Impα é também uma proteína do tipo solenóide, mas é composta a partir de 10 repetições ou motivos chamados de *Armadillo* (ARM) (KOBE, 1999; CONTI; KURIYAN, 2000) (Figura 1). Cada motivo ARM é formado por três α-hélices (H1, H2 e H3) com uma rotação em seu eixo que confere à molécula uma estrutura alongada e contorcida (CONTI et al., 1998). A porção côncava da Impα é o local de ligação da cNLS e apresenta uma alta conservação de resíduos justamente na interface de contato com a cNLS (MARFORI et al., 2012) (Figura 1). A Impα também apresenta uma pequena estrutura auto-inibitória na região N-terminal, que ocupa a região de ligação da cNLS. Esta estrutura é chamada de domínio de ligação à Impβ (IBB), pois se liga à Impβ (WEIS; RYDER; LAMOND, 1996) a fim de liberar o sítio de ligação à cNLS.

Além dos sinais de importação, algumas proteínas também têm sinais de exportação nuclear (NESs) (KUTAY; GÜTTINGER, 2005; COOK et al., 2007) e podem ser transportadas para dentro e fora do núcleo. Os membros da família β-Kap transportam proteínas que contêm NLS, ligando-se diretamente a ela ou por meio de uma molécula adaptadora tal como a Impα ou a Snurportina-1 (STEWART, 2007).

1.2 Importação nuclear mediada por Importina-α

A via de importação nuclear clássica (Figura 2) é a melhor caracterizada destes ciclos de transporte, e todos os seus componentes foram identificados (GÖRLICH; KUTAY,

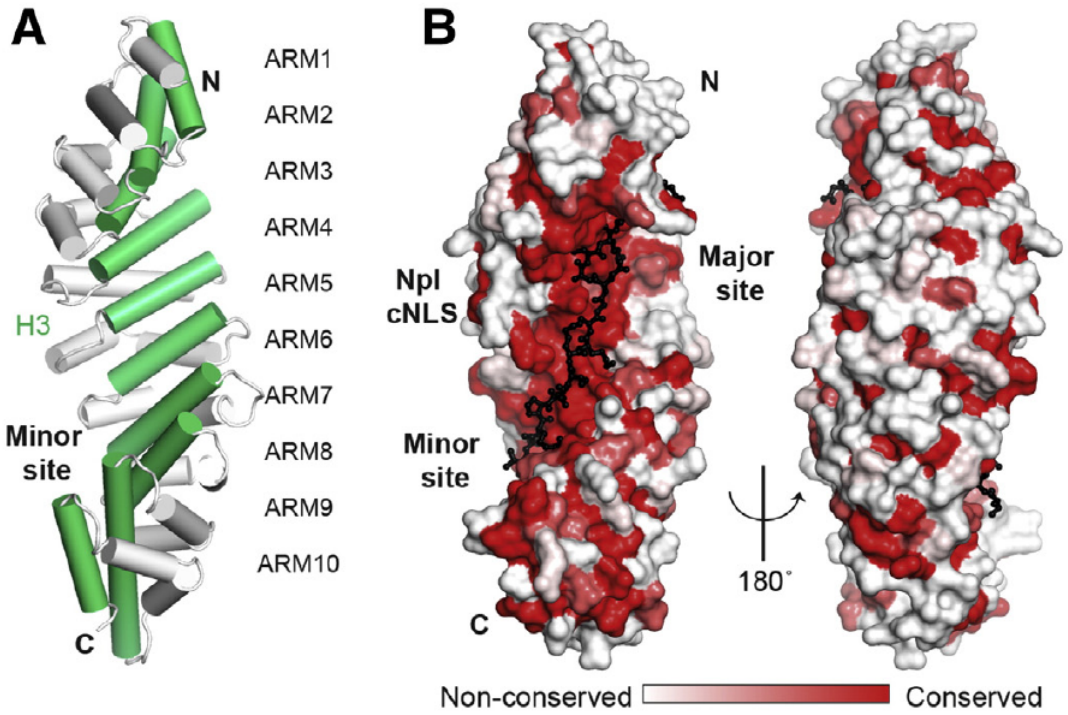


Figura 1 – Estrutura da Impα [adaptado de Christie et al. (2015)]. (A) Estrutura da Impα de arroz (PDB ID 4BQK) com α-hélices H3 na cor verde. (B) Estrutura da Impα de camundongo complexada com a cNLS da Nucleoplasmina (PDB ID 3UL1) (mostrado em *stick*). A Impα está colorida com base na conservação das sequências de estruturas de Impα conhecidas de diferentes espécies (*Homo sapiens*, *Mus musculus*, *Saccaromyces cerevisiae*, *Oryza sativa*, *Arabidopsis thaliana* e *Neurospora crassa*).

1999; CHOOK; BLOBEL, 2001; FAHRENKROG; AEBI, 2003; PEMBERTON; PASCHAL, 2005). Embora um número considerável de componentes e interações sejam necessárias para gerar o transporte nuclear, este sistema é menos complexo do que os sistemas que funcionam em outros processos celulares centrais, tais como sinalização, motilidade celular, ou o transporte de vesícula (STEWART, 2007). Isto, combinado com a disponibilidade de informação estrutural sobre a maioria dos componentes e muitos dos seus complexos, faz com que o ciclo da importação nuclear de proteínas se torne um sistema atrativo para desenvolver conceitos que podem servir como paradigmas da funcionalidade em sistemas biológicos mais complexos.

Existem diversas vias de importação nuclear de proteínas que usam diferentes carregadores, mas que compartilham muitas características comuns e são baseadas em uma série de proteínas que atuam em concerto: interações proteicas em que as proteínas transportadas são reconhecidas no citoplasma, translocadas através das NPCs e liberadas no núcleo (GÖRLICH; KUTAY, 1999; CHOOK; BLOBEL, 2001; FAHRENKROG; AEBI, 2003; PEMBERTON; PASCHAL, 2005). Em cada via, as proteínas a serem transportadas são reconhecidas por meio de um sinal denominado sequência de localização nuclear (NLS). As NLSs foram subdivididas em diversas classes (KOSUGI et al., 2009), podendo inclusive

ser reconhecidas por componentes de vias diferentes. A via clássica de importação nuclear, a qual utiliza Impα e Impβ como transportadoras, importa uma ampla variedade de proteínas e tem sido estudada com mais detalhes. Proteínas com uma NLS clássica são importadas pela Impβ, que se liga a essas proteínas por meio da proteína adaptadora chamada Impα (GÖRLICH; KUTAY, 1999; CHOOK; BLOBEL, 2001; FAHRENKROG; AEBI, 2003; PEMBERTON; PASCHAL, 2005). Ambas as importinas são moléculas alongadas formadas por uma série de sequências repetidas em *tandem*.

De forma simplificada, o ciclo de importação nuclear pode ser dividido em quatro etapas: i) montagem no citoplasma do complexo importina e proteína a ser transportada; ii) translocação através das NPCs; iii) desmontagem do complexo no núcleo e iv) reciclagem do complexo de importação (Figura 2).

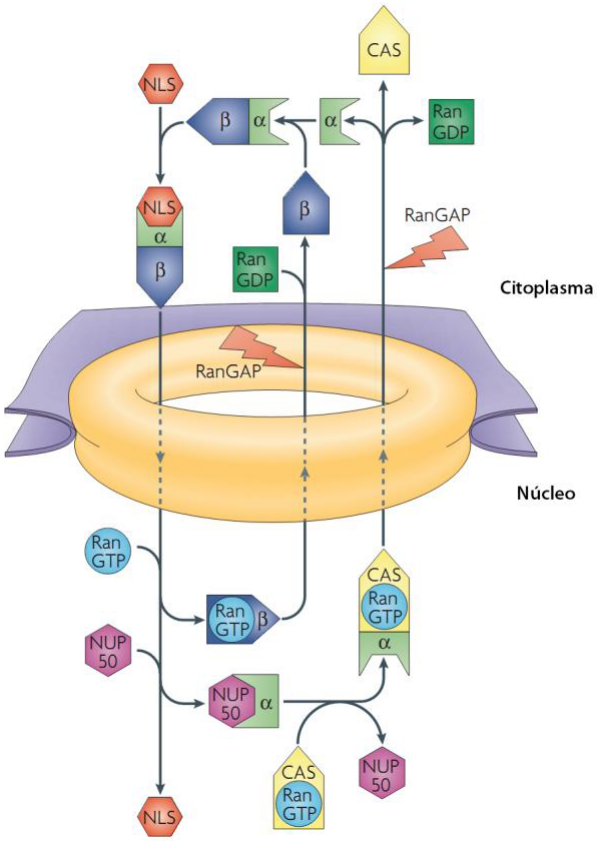


Figura 2 – Via clássica de importação nuclear [adaptado de Stewart (2007)]. O complexo de importação é formado no citoplasma entre proteínas a serem importadas que apresentam sinais de localização nuclear (NLSs), Impα e Impβ. Depois de passar através do complexo poro-nuclear (NPCs), a ligação de RanGTP na Impβ dissocia esta da Impα. A proteína contendo NLS é então deslocada da Impα e esta é transportada para o citoplasma por seu fator de exportação nuclear (i.e. CAS) complexado com RanGTP. No citoplasma, RanGAP estimula a hidrólise de GTP, liberando as importinas para um outro ciclo de importação.

Os estudos de biologia estrutural tem desempenhado um papel importante em decifrar os eventos moleculares necessários para a importação nuclear. A técnica de

cristalografia de raios-X tem sido utilizada extensivamente para elucidar os detalhes moleculares da ligação de cNLSs à Impα.

O primeiro motivo de direcionamento nuclear foi identificado no antígeno-T (TAg) do vírus símio SV40 (SV40TAg). O SV40TAg é composto por uma pequena extensão de aminoácidos básicos (126-P**KKKRR**V-132: os resíduos básicos estão em negrito) fundamentais para a importação nuclear, já que substituições de resíduos neste motivo inibiram o transporte da proteína (KALDERON et al., 1984a). Posteriormente, outro sinal de localização nuclear foi identificado para a proteína Nucleoplasmina de *Xenopus laevis*. Esse sinal compreende dois grupos de resíduos básicos separados por uma região intermediária chamada de *linker*, composta de 10 a 12 resíduos (155-**KR**PAATKKAGQA**KKKK**-170: os resíduos básicos estão em negrito e a região *linker* está sublinhada) (DINGWALL et al., 1988). De maneira semelhante ao SV40TAg, substituições de resíduos em qualquer um destes grupos afetou o transporte nuclear da proteína, sugerindo que ambos os motivos foram necessários para o direcionamento nuclear (ROBBINS et al., 1991).

Muitas proteínas com cNLS já foram identificadas com base na similaridade com estas duas sequências, consequentemente, as cNLSs são definidas pela presença de um ou dois grupos de seqüência ricas em aminoácidos básicos, principalmente, arginina e lisina, que são necessários e suficientes para a importação nuclear pelo complexo Impα:Impβ (FONTES; TEH; KOBE, 2000). Estes dois grupos são os sítios de ligação da cNLS à Impα e são frequentemente denominados de sítio principal (repetições ARM 2–4) e secundário (repetições ARM 6–8). Além disso, as cNLS que se ligam aos dois ou a apenas um sítio da Impα são denominadas de cNLSs bipartidas (e.g. Nucleoplasmina) e monopartidas (e.g. SV40TAg), respectivamente (FONTES; TEH; KOBE, 2000).

As análises estruturais por cristalografia de raios-X demonstraram que as cNLSs monopartidas ligam-se preferencialmente ao sítio de ligação principal (CONTI et al., 1998; FONTES; TEH; KOBE, 2000; YANG et al., 2010). Estes dados também foram corroborados por um estudo de mutagênese que mostrou que a mutação em resíduos do sítio principal inibiram a importação nuclear da cNLS monopartida, no entanto, as mutações no sítio secundário praticamente não afetaram o processo de importação (LEUNG et al., 2003). Por contraste, como já foi mencionado anteriormente para o estudo da Nucleoplasmina, a mutação em resíduos de qualquer um dos sítios de ligação afetaram substancialmente a importação das cNLSs bipartidas (LEUNG et al., 2003; ROBBINS et al., 1991).

Para cada sítio de ligação existem resíduos em posições específicas da cNLS que se ligam à Impα durante o processo de formação deste complexo. Na região da cNLS que se liga ao sítio principal da Impα, as cadeias laterais dos resíduos nas posições P2–P5 formam contatos em quatro cavidades principais da proteína, destes, o mais fundamental é o contato em P2 que é representado por uma lisina, cuja cadeia lateral forma uma ponte salina com a cadeia lateral do ácido aspártico da Impα (HODEL; CORBETT; HODEL,

2001; COLLEDGE et al., 1986; FONTES; TEH; KOBE, 2000; FONTES et al., 2003). Para as demais posições (P3–P5), há também uma preferência por resíduos com cadeia lateral básica, no entanto, este pode não ser um requerimento crítico. Por exemplo, um estudo mostrou que as cadeias laterais básicas não são estritamente necessárias para estas posições, contanto que a afinidade geral da cNLS seja suficiente para estabelecer e manter contatos mínimos com a Impα a fim de manter a cNLS funcional (HODEL; CORBETT; HODEL, 2001). A manutenção desta afinidade pode ocorrer por meio da realização de contatos em regiões diretamente flanqueadoras do sítio principal, ou então, da ligação ao sítio secundário no caso das cNLSs bipartidas (HODEL; CORBETT; HODEL, 2001). Isto foi mostrado em um estudo da SV40TAg que teve seu sítio principal mutado e, consequentemente, teve sua distribuição nuclear afetada (MAKKERH; DINGWALL; LASKEY, 1996). Neste estudo, foram adicionados resíduos na região N-terminal que correspondem aos resíduos de ligação ao sítio secundário, e esta alteração foi suficiente para direcionar o acúmulo nuclear desta proteína. As posições fundamentais de ligação ao sítio secundário são compostas também por resíduos básicos, normalmente, de lisina (P1') e arginina (P2'), sendo que a cadeia lateral da arginina forma uma ponte salina com o resíduo de ácido glutâmico da Impα (FONTES; TEH; KOBE, 2000; MARFORI et al., 2012). Uma comparação de valores de energia livre de ligação entre as posições de ambos os sítios demonstra que P1' e P2' são semelhantes às posições P3 e P5 de ligação ao sítio principal (HODEL; CORBETT; HODEL, 2001).

Para as cNLS bipartidas, as posições P2' e P2 são separadas por uma região intermediária, chamada de *linker*, de no mínimo 10 resíduos que possibilita que a cNLS se ligue simultaneamente nos dois sítios da Impα (LAI et al., 2000). Os dados estruturais tem mostrado a ocorrência de algumas interações a partir da região *linker* com os resíduos $_{\text{Imp}\alpha}\text{R}^{315}$, $_{\text{Imp}\alpha}\text{Y}^{277}$ e $_{\text{Imp}\alpha}\text{R}^{238}$ de algumas cNLSs bipartidas (FONTES; TEH; KOBE, 2000; FONTES et al., 2003; CONTI; KURIYAN, 2000; CHEN et al., 2005; CUTRESS et al., 2008; GIESECKE; STEWART, 2010; YANG et al., 2010; BARROS et al., 2012), porém estas interações não são conservadas e aparentam ser específicas de algumas cNLSs (MARFORI et al., 2012).

Com base nos estudos estruturais foi possível determinar os principais contatos envolvidos na ligação das cNLSs à Impα (um esquema geral para os contatos entre a Impα e a cNLS é mostrado na Figura 3), e assim, propor sequências-consenso para ambos os tipos de cNLSs. As cNLSs monopartidas são definidas como **K**(K/R)X(K/R), enquanto que as bipartidas correspondem a $\underline{\text{KR}}\text{X}_{10\text{-}12}\mathbf{K}\text{RRK}$, $\underline{\text{KR}}\text{X}_{10\text{-}12}\mathbf{K}\text{(KR)(KR)}$ e $\underline{\text{KR}}\text{X}_{10\text{-}12}\mathbf{K}\text{(K/R)X(K/R)}$ (X corresponde a qualquer resíduo, resíduos de lisina em negrito indicam a posição crítica em P2 e os motivos KR do sítio secundário estão sublinhados) (MARFORI et al., 2011; MARFORI et al., 2012).

Apesar de todo o avanço na elucidação dos mecanismos de importação nuclear para

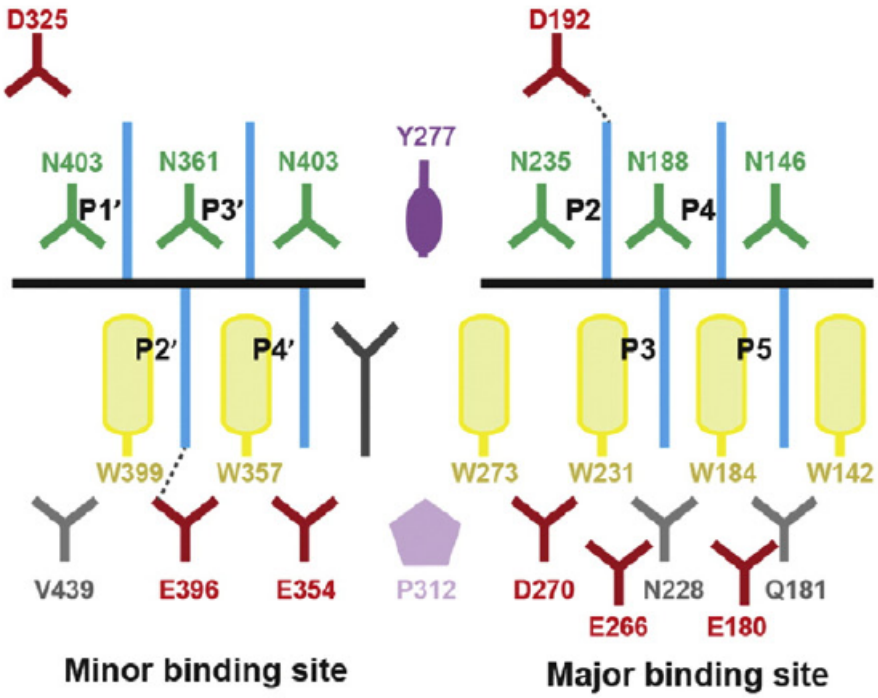


Figura 3 – Representação esquemática de uma NLS monopartida ligada à Impα nos sítios principal (*major binding site*) e secundário (*minor binding site*) [adaptado de Christie et al. (2015)]. Resíduos de asparagina e triptofano conservados da Impα são mostrados em verde e amarelo, respectivamente. As cadeias principal e lateral da cNLS são mostradas como linhas pretas e azuis, respectivamente. Linhas tracejadas indicam interações de ponte salina comuns nas cavidades de ligação P2 e P2'.

as vias clássicas, questões importantes ainda permanecem. Os movimentos funcionalmente importantes que a Impα pode adotar durante o reconhecimento da cNLS ainda não foram explorados, bem como uma análise sistemática das interações na região *linker* a fim de avaliar a ocupação destes contatos durante a sua ligação à Impα. A maior parte dos estudos estruturais com Impα tiveram um enfoque estático do sistema estudado. Uma abordagem interessante para tratar questões dinâmicas de um sistema é o uso de técnicas computacionais envolvendo a simulação de biomoléculas. Elas permitem analisar a evolução de componentes interativos da molécula e acessar comportamentos dinâmicos globais do sistema, que seriam de alto custo ou de difíceis (até mesmo limitados) procedimentos para serem obtidos experimentalmente.

1.3 Simulação computacional de biomoléculas

1.3.1 Dinâmica molecular

Ao longo do último meio século, os avanços em biologia estrutural forneceram a resolução atomística de muitas das moléculas que são essenciais à vida, incluindo

proteínas e ácidos nucleicos (DROR et al., 2012). Embora modelos estruturais estáticos determinados por meio de técnicas como a cristalografia de raios-X sejam extremamente úteis, as moléculas que estes modelos representam são, na realidade, altamente dinâmicas, e seus movimentos são muitas vezes fundamentais para a sua função celular (DROR et al., 2012).

As proteínas estão envolvidas em processos biológicos como metabolismo, transmissão de sinal, armazenamento de energia, defesa contra invasores e formação de tecido muscular. A capacidade para a realização dessas funções depende das possíveis alterações conformacionais da proteína, da dinâmica dessas conformações e das consequentes interações geradas (HALPERIN et al., 2002). Um completo entendimento da função proteica, portanto, requer uma compreensão tanto do comportamento da dinâmica como das características estruturais estáticas da proteína (GIPSON et al., 2012).

Simulações computacionais de sistemas biomoleculares têm crescido rapidamente ao longo das últimas décadas, simulando desde proteínas muito pequenas no vácuo até grandes complexos de proteínas na presença de solvente (KARPLUS; MCCAMMON, 2002; OROZCO, 2014). Neste contexto, a simulação de dinâmica molecular é uma ferramenta computacional amplamente utilizada para simular as propriedades dos líquidos e sólidos a nível atomístico em áreas de investigação, tais como a química, termodinâmica e várias outras áreas (YANG; WANG; CHEN, 2007). Simulações de dinâmica molecular *all atom*, empregando mecânica clássica, permitiram o estudo de complexos muito grandes de proteínas, tais como o ribossomo (SOTHISELVAM et al., 2014) e capsídeos de vírus (SUN et al., 2013; ZHAO et al., 2013). Métodos híbridos QM/MM (*Quantum Mechanics/Molecular Mechanics*) permitiram o estudo da atividade enzimática (SENN; THIEL, 2007) e moléculas polarizáveis em membranas biológicas (BERNARDI; PASCUTTI, 2012).

Métodos computacionais de simulação biomolecular oferecem uma clara vantagem no entendimento da flexibilidade da proteína, podendo caracterizar a dinâmica deste sistema. Neste contexto, simulações de dinâmica molecular modelam as interações físicas entre átomos resolvendo as equações de movimento de Newton, Lagrange ou Langenvin (ADCOCK; MCCAMMON, 2006). No entanto, a solução da dinâmica desses sistemas é complicada (SCHLICK et al., 1999; HENZLER-WILDMAN; KERN, 2007). Simulações de dinâmica molecular são ainda limitadas em dois aspectos: precisão dos campos de força e alto custo computacional. Essas limitações podem gerar uma amostragem inadequada dos estados conformacionais, limitando a capacidade de analisar e revelar propriedades funcionais dos sistemas examinados. Moléculas biológicas são conhecidas por terem perfis rugosos de energia, com muitos mínimos locais frequentemente separados por barreiras de alta energia (ONUCHIC; LUTHEY-SCHULTEN; WOLYNES, 1997), o que pode ocasionar num aprisionamento em algum estado não-funcional, de difícil deslocamento pela maioria dos métodos de simulações convencionais. Estudos recentes têm demonstrado que, de fato,

em simulações longas, as proteínas podem ficar "presas" em conformações não relevantes sem conseguir voltar à conformação original funcionalmente relevante (BERGONZO et al., 2013; MARSILI et al., 2010). Na tentativa de sobrepor essas limitações, vários métodos têm sido desenvolvidos para esta finalidade, como dinâmica molecular com troca de réplica (*replica-exchange molecular dynamics*) (REMD) (SUGITA; OKAMOTO, 1999; GARCÍA; SANBONMATSU, 2001), metadinâmica (LAIO; PARRINELLO, 2002; BARDUCCI; BUSSI; PARRINELLO, 2008) e simulação por termalização generalizada (*generalized simulated annealing*) (TSALLIS; STARIOLO, 1996). Outra alternativa para expandir a exploração conformacional de proteínas tem sido a aplicação do método analítico de modos normais de vibração. Independentemente dos métodos aplicados, todos estão associados a vantagens e desvantagens para a simulação efetiva de sistemas biológicos e a escolha adequada de qual método utilizar está atrelada à particularidade e ao tipo de informação que se deseja obter a partir de cada sistema biológico de interesse.

1.3.2 Análise de modos normais

Nas últimas décadas, houve um grande aumento no número de estudos baseados na análise de componentes principais de estruturas biomoleculares (JOLLIFFE, 2002). Estes estudos têm se mostrado útil em desvendar os modos coletivos, e em particular aqueles modos de baixa frequência, que fundamentam a dinâmica em equilíbrio de proteínas (KITAO; GO, 1999). Análise de modos normais de estruturas em equilíbrio (BAHAR; RADER, 2005; CUI; BAHAR, 2010), análise da dinâmica essencial de matrizes de covariância obtidas de simulações de dinâmica molecular (AMADEI; LINSSEN; BERENDSEN, 1994) e decomposição em valores singulares de trajetórias de dinâmica molecular ou Monte Carlo (BAHAR et al., 1997; GARCÍA; HARMAN, 1996; ROMO et al., 1995), todos estes compõem a categoria de métodos baseados em análise de componentes principais.

Análise de modo normal é uma ferramenta poderosa para predizer os possíveis movimentos de uma determinada macromolécula. Modos normais de vibração são oscilações harmônicas simples a partir de um mínimo local de energia, característico de um sistema de estrutura $\vec{R}$ e sua função de energia $V(\vec{R})$ (CUI; BAHAR, 2010). Para uma função $V(\vec{R})$ puramente harmônica, qualquer movimento pode ser exatamente expresso como uma combinação de modos normais (CUI; BAHAR, 2010). Para uma função $V(\vec{R})$ não-harmônica, o potencial próximo ao ponto de mínimo de energia ainda pode ser aproximado apropriadamente por um potencial harmônico e qualquer movimento de pequena amplitude ainda pode ser bem descrito por uma soma de modos normais (CUI; BAHAR, 2010).

Uma aplicação importante dos modos normais é a identificação de potenciais mudanças de conformação, por exemplo, de enzimas após a associação do ligante (TAMA et al., 2000; TAMA; SANEJOUAND, 2001; DELARUE; SANEJOUAND, 2002). O método também tem sido utilizado recentemente no estudo da abertura dos canais de membrana

(VALADIE et al., 2003), na análise de movimentos estruturais do ribossomo (TAMA et al., 2003), na maturação do capsídeo de vírus (KIM; JERNIGAN; CHIRIKJIAN, 2003) e na análise de movimentos de domínio em grandes proteínas em geral (HINSEN, 1998; HINSEN; THOMAS; FIELD, 1999).

Análise de modos normais é mais frequentemente usada na tentativa de prever que tipo de mudança conformacional uma proteína sofre, a fim de cumprir a sua função, por meio da análise de seus modos de menor frequência e, consequentemente, maior amplitude. Normalmente, 50% dos movimentos observados na proteína podem ser descritos com precisão apenas por um ou dois modos de baixa frequência (SUHRE; SANEJOUAND, 2004).

Cálculos de modos normais de moléculas de interesse biológico foram introduzidos vários anos após a primeira simulação de dinâmica molecular (CUI; BAHAR, 2010). É reconhecido que os modos normais têm vários atributos importantes, que os tornam de interesse como um complemento para simulações de dinâmica molecular. A principal desvantagem dos modos normais é justamente a sua aproximação do potencial total por uma função harmônica em torno de uma estrutura de mínima energia (CUI; BAHAR, 2010). Apesar de ser uma aproximação harmônica, há vantagens importantes que a tornam uma técnica que vem sendo utilizada em estudos de dinâmica de macromoléculas: (a) ao contrário de simulações de dinâmica molecular, os resultados são essencialmente precisos e sem erros estatísticos, (b) é necessária apenas a diagonalização de uma matriz, (c) a quantificação, que é de particular importância para os cálculos de calor e cálculos específicos de entropia, pode ser introduzida de uma forma simples, e (d) o fornecimento de informações sobre as alterações de conformação é muitas vezes mais fácil de visualizar do que simulações de dinâmica molecular, principalmente por proporcionar a observação de movimentos de maior amplitude (CUI; BAHAR, 2010).

Diferentes exemplos demonstram movimentos de transição funcionalmente importantes, frequentemente seguindo as trajetórias de um ou mais modos normais (MA, 2005). Uma conclusão importante nestes estudos é que as estruturas das proteínas evoluíram de uma maneira em que as flexibilidades estruturais intrínsecas, observadas nos modos normais, facilitam a ocorrência de variações conformacionais funcionalmente importantes.

A combinação do cálculo de modos normais com simulações de dinâmica molecular curtas tem favorecido a maior exploração conformacional atrelada a um refinamento das estruturas geradas. Recentemente, foi desenvolvido um protocolo de *dinâmica molecular com modos normais excitados* (COSTA et al., 2015), em que é possível explorar os perfis de energia livre de mudanças conformacionais amplas.

2 Objetivo

O objetivo deste estudo foi compreender os mecanismos de interação e dinâmica conformacional envolvidos no reconhecimento de sequências de localização nuclear clássicas pela Importina-α.

3 Manuscrito I: Nucleoplasmina

O capítulo trata do estudo estrutural envolvendo a proteína Impα complexada com a NLS da proteína Nucleoplasmina. Nosso trabalho teve como enfoque determinar os principais movimentos da Impα possivelmente associados ao reconhecimento da NLS e também avaliar dinamicamente o comportamento das interações na interface do complexo. Para isso, adotamos a combinação das técnicas de dinâmica molecular e de modos normais. Nossos dados mostram que os movimentos de dobramento e torção da Impα podem estar associados ao reconhecimento de NLSs. Além disso, determinamos os contatos críticos durante este reconhecimento, sugerindo que os resíduos da região conhecida como *linker* podem ter um papel no aumento da afinidade com a NLS e, portanto, estarem envolvidos diretamente no processo de importação nuclear.

O trabalho foi publicado na PLOS ONE (doi: **10.1371/journal.pone.0157162**).

Observação:

Conforme estabelecido pelo Conselho do Programa de Pós-Graduação em Ciências Biológicas (Genética), expresso na Instrução Normativa Nº01/2012, os resultados apresentados neste capítulo foram redigidos na forma de manuscrito.

Bending-Twisting Motions and Main Interactions in Nucleoplasmin Nuclear Import

Marcos Tadeu Geraldo[1,*], Agnes Alessandra Sekijima Takeda[1,2], Antônio Sérgio Kimus Braz[3], Ney Lemke[1]

1 Laboratório de Bioinformática e Biofísica Computacional, Departamento de Física e Biofísica, Instituto de Biociências de Botucatu, UNESP – Universidade Estadual Paulista, Botucatu, SP, 18618-970, Brazil
2 Instituto de Biotecnologia (IBTEC), UNESP – Universidade Estadual Paulista, Botucatu, SP, 18607-440, Brazil
3 Laboratório de Biologia Computacional e Bioinformática, Centro de Ciências Naturais e Humanas, UFABC – Universidade Federal do ABC, Santo André, SP, 09210-170, Brazil

* mtgeraldo@gmail.com

3.1 Abstract

Alpha solenoid proteins play a key role in regulating the classical nuclear import pathway, recognizing a target protein and transporting it into the nucleus. Importin-α (Impα) is the solenoid responsible for cargo protein recognition, and it has been extensively studied by X-ray crystallography to understand the binding specificity. To comprehend the main motions of Impα and to extend the information about the critical interactions during carrier-cargo recognition, we surveyed different conformational states based on molecular dynamics (MD) and normal mode (NM) analyses. Our model of study was a crystallographic structure of Impα complexed with the classical nuclear localization sequence (cNLS) from nucleoplasmin (Npl), which was submitted to multiple 100 ns of MD simulations. Representative conformations were selected for calculating the 87 lowest frequencies NMs of vibration, and a displacement approach was applied along each NM. Based on geometric criteria, using the radius of curvature and inter-repeat angles as the reference metrics, the main motions of Impα were described. Moreover, we determined the salt bridges, hydrogen bonds and hydrophobic interactions in the Impα-NplNLS interface. Our results show the bending and twisting motions participating in the recognition of nuclear proteins, allowing the accommodation and adjustment of a classical bipartite NLS sequence. The essential contacts for the nuclear import were also described and were mostly in agreement with previous studies, suggesting that the residues in the cNLS linker region establish important contacts with Impα adjusting the cNLS backbone. The MD simulations combined with NM analysis can be applied to the Impα-NLS system to help

understand the interactions between Impα and cNLSs and the analysis of non-classic NLSs.

3.2 Introduction

Solenoid proteins are molecules composed of structural motifs that are arranged in tandem, creating a superhelical architecture. This modular characteristic provides the establishment of folding and binding contacts that contrasts the globular proteins, allowing higher flexibility and the arrangement of cooperative protein-protein interactions due to the formation of diversified interfaces (KOBE; KAJAVA, 2000; KAPPEL et al., 2010; DOYLE et al., 2015; SETTANNI et al., 2013).

A remarkable characteristic in nuclear protein import regulation is the contribution of solenoid proteins to recognizing a target protein and transporting it into the nucleus (CHRISTIE et al., 2015). One of the most studied pathways is the classical nuclear import pathway, which requires the interaction between the solenoid Importin-α (Impα) and Importin-β (Impβ) proteins, followed by the assembly of the cargo protein to Impα (CONTI et al., 1998; RADU; BLOBEL; MOORE, 1995; CINGOLANI et al., 1999; GÖRLICH et al., 1995; GÖRLICH et al., 1996). This trimeric complex is translocated through the nuclear pore complex (NPC), and the cargo protein is delivered into the cell nucleus mediated by Ran-GTP-dependent steps of protein-protein interactions (MOROIANU; BLOBEL; RADU, 1996; LEE et al., 2005). The dissociation of the cargo protein from Impα is catalyzed by nucleoporins (e.g., NUP50) (MATSUURA; STEWART, 2005), and Impα binds to its export factor CAS complexed with RanGTP (MATSUURA; STEWART, 2004). A final step is the recycling of both Impα and Impβ back to the cytoplasm.

Impα is composed of ten tandem repeats of armadillo motifs (ARM) oriented in an elongated and curved-twisted shape (CONTI et al., 1998; RIGGLEMAN; WIESCHAUS; SCHEDL, 1989; GÖRLICH et al., 1996). Each motif is formed by a superhelical architecture of three α-helices (CONTI et al., 1998; RIGGLEMAN; WIESCHAUS; SCHEDL, 1989; KOBE, 1999). From the curved orientation, convex and concave surfaces can be identified; in particular, the inner concave surface harbors conserved residues that mediate the cargo protein binding (CONTI et al., 1998; MARFORI et al., 2012).

Cargo protein transport depends on the recognition of a specific sequence signal by Impα called the nuclear localization sequence (NLS). For classical NLSs (cNLSs), the binding pattern to Impα is primarily mediated by one or two clusters of positively charged residues, the monopartite or bipartite cNLSs, respectively (DINGWALL et al., 1988; KALDERON et al., 1984b; ROBBINS et al., 1991; FONTES et al., 2003; KOSUGI et al., 2009). The clusters of bipartite cNLSs are separated by 10-12 variant residues,

denominated as the linker region.

The inner concave surface of Impα is adapted to receive either monopartite or bipartite cNLSs, and its specific binding sites can be identified as the major and minor sites. Structural and biophysical studies have related important positions of Impα to the NLS binding: positions P2-P5 from the major site (ARMs 2-4; binding to both monopartite and bipartite cNLSs) and P1'-P2' from the minor site (ARMs 6-8; binding preferentially to bipartite cNLSs) (FONTES et al., 2003; HODEL; CORBETT; HODEL, 2001). Moreover, the region between major and minor sites (ARMs 4-6) provides interactions and is also considered fundamental in cNLS recognition (FONTES et al., 2003; FONTES; TEH; KOBE, 2000; CONTI; KURIYAN, 2000; CHEN et al., 2005; CUTRESS et al., 2008; GIESECKE; STEWART, 2010; YANG et al., 2010; BARROS et al., 2012).

Few studies have reported the flexibility and structural integrity of Impβ (KAPPEL et al., 2010; FORWOOD et al., 2010); however, the application of computational simulation approaches is underexplored for understanding the wide structural motions and the interaction basis of nuclear import mediated by Impα binding. Therefore, two main questions arose: (i) Are there motions related to NLS recognition? (ii) What is the dynamical behavior of the interactions on the complex interface? Motivated by these questions, we combined molecular dynamics simulations (MD) and normal mode (NM) analysis. As our cNLS model of study, we used the crystallographic structure of nucleoplasmin NLS (NplNLS) complexed with Impα because experimental works involving Npl (MARFORI et al., 2012; FONTES; TEH; KOBE, 2000; MAKKERH; DINGWALL; LASKEY, 1996) are available and could be used to support and contrast with our simulation results. Npl is the first protein to be described as a molecular chaperone involved in chromatin reprogramming (DINGWALL; LASKEY, 1990), and it is characterized as containing a bipartite cNLS (Npl:^{151}GSAV**KR**PAATKKAGQA**KKKK**LD^{172}; residues in bold indicate the positions in contact with Impα minor and major binding sites, respectively). We determined that bending and twisting-like major movements of Impα may influence the NLS binding. In addition, we confirmed the importance of contacts in the major and minor sites, along with contacts flanking these sites, including the linker region, for the establishment of carrier-cargo recognition.

3.3 Materials and Methods

3.3.1 Model of study: Classical bipartite NLS

All simulation analyses were conducted using the crystallographic structure of NplNLS complexed to mouse Impα isoform 2 (Impα–NplNLS; PDB ID: 3UL1) (MARFORI et al., 2012). The missing atoms from residues G^{152} and S^{342} of NplNLS (^{152}GSAVKRPAATKK

AGQAKKKKLD[172]) and Impα (Imp$\alpha\Delta$IBB, residues 72-497), respectively, were modeled with MODELLER v9.11 software (ESWAR et al., 2006). The N-terminal Impβ-binding (IBB, residues 1 to 71) domain from Impα was kept truncated because it competes to the binding in the NLS region. The choice criteria for determining the best model were guided by the correct stereochemistry and the occurring interactions in the complex interface, using Molprobity (CHEN et al., 2009) and PISA (KRISSINEL; HENRICK, 2007) servers, respectively. An additional modeling for Impα was conducted removing the peptide (Apo Impα) to compare the motions and flexibility of Impα in the presence and absence of the cNLS.

3.3.2 MD simulations

The topology and parameter files were generated on the program GROMACS v4.5.3 (HESS, 2008), employing the force field Charmm36 (HUANG; MACKERELL, 2013) without a protonation requirement because the PROKPA webserver (http://propka.org) analysis indicated no protonation changes. A cubic box of 56,568 explicit TIP3P water molecules (JORGENSEN et al., 1983; MACKERELL et al., 1998) was generated, ensuring at least 10 ångströms (Å) from the protein system to each edge of the box. Counter ions were added for system-charge neutralization by replacing water molecules. The system was submitted to a gradual energy minimization composed of four steps: (i) 500 steps of energy minimization by the steepest descent method, limiting the protein and peptide movement to accommodate the solvent molecules; (ii) 50,000 steps of energy minimization by the steepest descent method, limiting the movement of the protein and peptide's main chains; (iii) 50,000 steps of unrestricted energy minimization by the steepest descent method; and (iv) 50,000 steps of unrestricted energy minimization by the conjugated gradients (CG) method.

Equilibration and unrestrained MDs were performed in periodic boundary conditions. The leapfrog integrator was used for integrating Newton's equations of motion. The linear constraint solver (LINCS) method (HESS et al., 1997; HESS, 2008) was used to freeze bonds involving hydrogen atoms, allowing an integration step of 2 femtoseconds (fs). The cutoff distance for short-range electrostatic and van der Waals interactions was 10 Å. The Particle Mesh Ewald method (PME) (ESSMANN et al., 1995) was used to treat long-range electrostatics.

The system was equilibrated in two steps, both applying a position restraining force on the heavy atoms of the protein. The first stage involved the adoption of NVT conditions (constant number of particles, volume and temperature), heating the system to the target temperature of 300 kelvin (K) and simulating in this condition by 100 picoseconds (ps) with a velocity-rescaling (V-rescale) thermostat (BUSSI; DONADIO; PARRINELLO, 2007). The second stage involved 100 ps of equilibration by the adoption of NPT

conditions (constant number of particles, pressure and temperature), with pressure coupling using the Parrinello-Rahman barostat (PARRINELLO; RAHMAN, 1981) and keeping the pressure relatively constant, close to the value of 1 bar. After the equilibration procedures, the restraints were removed, and the system was submitted to three MD simulations of 100 nanoseconds (ns) each, with structure sampling every 10 ps.

To ensure the conformational sampling of the system, a clustering technique proposed by Lyman and Zuckerman (LYMAN; ZUCKERMAN, 2006) was applied. In this approach, the steps below were applied to generate a collection of reference structures of the simulations $\{S_i\}$: (1) A cutoff root-mean square deviation (RMSD) d was defined. (2) Each simulation was merged into a single trajectory file. (3) One structure from the trajectory file was sampled randomly and denominated as the reference structure. (4) All structures were compared to the sampled reference structure, and the ones with an RMSD less than d were removed from the trajectory file. (5) Steps 1, 2, 3 and 4 were repeated until all structures were removed, thus generating a collection of reference structures $\{S_i\}$. (6) Based on the collection $\{S_i\}$, each frame from the trajectory file was clustered with the nearest reference structure, and the frequency of structures of each cluster was then calculated. An estimation of convergence is assessed when each reference structure is equivalently represented in each simulation. The parameter for the calculation of d was the global RMSD of carbon-α (Cα) atoms. The choice of d (=2.5 Å) was constrained to a feasible number of reference structures for the subsequent analysis of NMs.

3.3.3 NM analysis

Each reference structure obtained from MD simulations was minimized in the GROMACS program with explicit solvent using the same methodology described above. All NM analyses were carried out in the CHARMM v.36b1 program (BROOKS et al., 2009). The topology and parameter files required for CHARMM were generated with the CHARMM-GUI server (www.charmm-gui.org) employing an additional energy minimization. The CG algorithm was applied with harmonic constraints that were progressively decreased from 250 to 5 kcal/mol^{-1}Å^{-2}, with 100 steps of minimization at each decrease. Then, the constraints were removed, and 10,000 steps of CG were carried out. Afterwards, the adopted basis Newton Raphson (ABNR) algorithm was applied with no constraints for 300,000 steps.

The final minimized structure was used for the calculation of the 87 lowest frequency NMs, using the VIBRAN module of CHARMM for each reference structure. An NM-displacement method using the VMOD facility of CHARMM was applied, generating structures along each NM based on short MD simulations at a low temperature (30 K), followed by energy minimization. Based on the values of mass-weighted root mean square (MRMS), the maximum displacement range was set to 3 Å for each direction of the NM

with a 0.1 Å projection step, totaling 61 structures per NM. For each MRMS step, a harmonic force constant over the Cα atoms was applied (increasing from 1,000 until 10,000 kcal/mol^{-1}Å^{-2}), and a short MD simulation was carried out for 1 ps for each constant value, totaling 10 ps of simulation. Keeping the restraints, 1,000 steps of CG energy minimization were employed to generate the final structure. In addition, for each projection step, a value of the total restraint energy, according to the miscellaneous mean field potential (MMFP) facility of CHARMM, was used as a criterion for discarding unfavorable conformations.

3.3.4 Data Analysis

PCA over MD simulations

Principal component analysis (PCA) from MD simulations was performed using the quasi-routine of the module VIBRAN of CHARMM to obtain the covariance matrix of Cα atomic displacements from the trajectories and identify the most relevant structural variations. The calculation was performed according to the description of Floquet et al. (2015), and the resulting principal components (PCs) were used to compare the motions observed in NMs.

Collectivity

The measurement of the involvement of atoms in a particular protein motion (referred to as degree of collectivity) for a given NM was calculated according to Brüschweiler (1995) and Tama and Sanejouand (2001), using an *in-house* CHARMM script. The degree of collectivity is comprised between 0 and 1. Values close to 1 indicate maximum collectivity.

Geometric analysis of Impα

The wide motions from Impα were described in terms of geometrical measurements. For bending characterized motions, three vertices (represented by Impα residues R^{117}, A^{313} and K^{486}) were manually selected: two vertices in distal tips and one vertex at the middle point of the protein. The plane formed by these three vertices corresponded to the plane of the observed bending motion (Fig 1). Then, the radius of curvature (R) was calculated:

$$R = \frac{d^2}{2\sqrt{d^2 - m^2}} \tag{3.1}$$

d is the distance from the distal vertices to the middle vertex, and m is half the distance between the two distal vertices.

For twisting characterized motions, the angle between α-helices 3 (H3) from neighboring ARM repeats were calculated using an available script for PyMol

(http://www.pymolwiki.org/index.php/AngleBetweenHelices). For each pair of helices, vectors were defined along the Cα atoms, and the torsion angles between these vectors were determined.

This geometric analysis was also applied to X-ray-solved structures of Impα complexed with different types of bipartite NLSs. These structures were retrieved by the basic local alignment search tool allocated at Prody software (BAKAN; MEIRELES; BAHAR, 2011) using the Impα–NplNLS model as the query. Only structures with 100% of sequence identity to Impα were selected to be compared to the simulation data.

Maps of cross-correlations

The initial comparison of residue-residue contacts from the MD and NM results was initially performed by the generation of maps of cross-correlation to evaluate the associated movements between NplNLS and Impα. The calculation of correlations corresponded to the ensemble of the trajectories of all MD simulations into one final pseudo-trajectory. Similarly, an ensemble of all structures from NM-displacements into one pseudo-trajectory was conducted. The cross-correlation calculations were performed in the Wordom software (SEEBER et al., 2007).

Interactions evaluation

The occurrence of specific interactions in the Impα-NplNLS interface was evaluated. The determination of salt bridges and hydrogen bonds were performed using the VMD software (HUMPHREY; DALKE; SCHULTEN, 1996). The criteria for considering the occurrence of these interactions were the donor-acceptor distance for salt bridges and hydrogen bonds $\leqslant$3.5 Å and the donor-hydrogen-acceptor angle deviation for hydrogen bonds $\leqslant$60 degrees. The determination of hydrophobic contacts was based initially on the LIGPLOT program, using the crystallographic and *in silico* model of Impα–NplNLS to generate a list of possible interactions. Later, based on this list, the distances of the closest carbon atom from the hydrophobic side chains of each residue-pair were calculated for the MD and NM ensemble pseudo-trajectories to calculate the percentage of the occurrence of each hydrophobic contact. The criterion adopted was distances $\leqslant$4 Å.

Complementary analysis

The backbone RMSD and the Cα root-mean-square fluctuations (RMSF) of MD and NM-displacements were calculated in the Wordom software (SEEBER et al., 2007). The structural alignment of NplNLS peptides was performed in the Crystallographic Object-Oriented Toolkit (Coot) software (EMSLEY et al., 2010). All graphics from the calculations above were performed using the R software (IHAKA; GENTLEMAN, 1996),

and the structural analysis, visualization and generation were performed in the Pymol software (SCHRÖDINGER, 2010).

3.4 Results

3.4.1 Selection of Impα–NplNLS model

A stereochemical analysis allowed for the selection of the best complex model, taking into account the geometry and maintenance of the main interactions in the peptide-protein interface. The best-selected model showed a Ramachandran plot for Impα with 96.7% and 3.3% of the residues in favored and allowed regions, respectively, without residues in the outlier region, whereas NplNLS had 100% of the residues in favored regions.

An overall evaluation of the Impα–NplNLS model indicated the NplNLS harboring in both major and minor binding sites of Impα, with the main residues inside the binding pockets (Fig 2). Furthermore, we were able to evidence the previously stated interactions in the interface of this complex (MARFORI et al., 2012; FONTES; TEH; KOBE, 2000); the only major exception was the absence of $_{\text{Imp}\alpha}\text{D}^{325}$ in contact with $_{\text{Npl}}\text{K}^{155}$ in the minor site (S1 Fig).

3.4.2 Standard MD combined with NM-displacement method

The trajectories obtained from the three Impα-NplNLS MD simulations were clustered into three reference structures (67,730 ps, 207,080 ps and 274,970 ps), each one exhibiting a similar frame-frequency among simulations, indicating a likely convergence of the MDs (S2 Fig). Reference structure 207,080ps was the most representative in the simulations (approximately 80% of the trajectories clustered with this structure) and was more similar to the X-ray-solved NplNLS structure based on the backbone RMSD values (S1 Table) obtained from the structural alignment of the NLSs (S3 Fig). In this alignment, important positions from major and minor sites were occupied by the expected residues, with similar side chain conformations, represented by $_{\text{Npl}}\text{K}^{167}$, $_{\text{Npl}}\text{K}^{168}$ and $_{\text{Npl}}\text{K}^{170}$ in positions P2, P3 and P5, respectively, and $_{\text{Npl}}\text{K}^{155}$ and $_{\text{Npl}}\text{R}^{156}$ in P1' and P2', respectively. Reference structures 67,730 ps and 274,970 ps showed a greater structural variance of side chains in the major site than in the minor site.

The subsequent step was the NM-displacement method applied to the three MD reference structures. We observed flexible and favorable motions in the early NMs (modes 7-20) of Impα-NplNLS, represented by lower values of restriction energy along the whole displacement range, compared to the remaining modes (blue areas in the S4 Fig). Similar patterns were also observed for Apo Impα; however, the favorable motions were extended

up to NM29 (S5 Fig), indicating that more conformations could be acquired favorably compared to the Impα bound to NplNLS.

The distribution of backbone RMSD showed that standard MD in combination with the NM-displacement method increased the conformational exploration, reaching values over 5 Å, particularly to reference structures 207,080 ps and 274,970 ps, whereas MD alone reached only approximately 2.5 Å (S6 Fig). Moreover, greater Cα fluctuations were observed in the NM-displacement technique, based on the RMSF values (S7 Fig and S8 Fig). The comparison between the RMSF values for Impα-NplNLS and Apo Impα showed small differences, primarily limited to the region of the major site (within the range of residue 100 to 200), where Apo Impα was more flexible (S9 Fig).

3.4.3 Collective motions of Impα

The description of the motions obtained from MD and NM calculations combined the results from a qualitative vector-based analysis and a quantitative geometrical analysis. The first NMs showed wide and collective types of motions of the ARM repeats of Impα, primarily along modes 7-17 (S10 Fig). A qualitative analysis of vectors from modes 7 and 9 clearly showed a motion pattern, described by a bend and a twist (Fig 3 and S1 Movie and S3 Movie), respectively.

The bend motions in NM7/PC1 were characterized by the opening and closing of Impα in the concave surface, along the NLS binding pockets. The quantitative analysis using geometric measurements determined the radius of curvature in opened and closed configurations of Impα (Fig 4A, S12 Fig and S1 Movie) and showed different amplitudes for the bending motion, in which Apo Impα had higher amplitudes than the Impα-NplNLS complex.

The twist pattern for the Impα structures along NM9/PC3 (Fig 3, S12 Fig and S3 Movie) oscillated from maximum and minimum values of torsion over the entire protein (Fig 4B). A general observation of the angles for the Impα-NplNLS complex and Apo Impα showed similarities among the structures. A more detailed analysis, considering the inter-repeat angles (Fig 5), detected oscillation between the ARMs in the Impα-NplNLS complex, primarily between ARMs 5-6. Movements with smaller amplitudes were observed for most pairs of ARMs in Apo Impα, except the oscillation observed between ARMs 6-7.

Bending and twisting movements were also evaluated for the crystal structures of Impα in the presence of different types of bipartite cNLSs. A comparison of geometries indicated a small difference among them in the order of tenths of ångströms (S5 Table). The qualitative vector analysis of Impα-NplNLS NM8/PC2 indicated a combination of two motions, characterized by a "lateral" bending tendency (based on a 90º X-axis rotation of Impα in relation to the bending orientation in NM7/PC1) in ARMs 1-6, mixed with a

twisting in ARMs 7-10 (Fig 3, S12 Fig and S2 Movie). The amplitude of the inter-repeat angle variation was slightly distinct between NM8 and PC2; however, in both cases, we could verify motions that almost reached both sites, including their intermediate region (Fig 5). The complexity of these mixed motions increased in the following modes (S11 Fig and S13 Fig). The same analysis for Apo Impα NM8 showed the reduction of movement amplitude for most ARMs.

3.4.4 Main contacts in Impα-NplNLS interface

In general, the cross-correlations results from MD and NM analyses were similar (Fig 6). Well-bounded areas of positive correlations could be identified, highlighting the NplNLS range of residues in contact with Impα major and minor sites and the linker region. However, for standard MD, the correlations were more scattered in the linker.

Subsequent analysis showed that most correlations could be related to specific interactions in the interface of Impα-NplNLS. Compared to our simulation data, the starting structure exhibited more contacts of hydrogen bonds and hydrophobic interactions (S1 Fig). Throughout the simulations, some of those initial contacts were lost, and MD and NM showed mostly common results (Fig 7). In particular, we observed salt bridges ($_{\text{Imp}\alpha}\text{E}^{396}$, $_{\text{Imp}\alpha}\text{D}^{280}$, $_{\text{Imp}\alpha}\text{D}^{192}$ and $_{\text{Imp}\alpha}\text{D}^{270}$) that were established along the NplNLS, specifically in P2', P2 and the linker region. Hydrophobic contacts mediated by tryptophans ($_{\text{Imp}\alpha}\text{W}^{399}$, $_{\text{Imp}\alpha}\text{W}^{357}$, $_{\text{Imp}\alpha}\text{W}^{273}$, $_{\text{Imp}\alpha}\text{W}^{231}$, $_{\text{Imp}\alpha}\text{W}^{184}$ and $_{\text{Imp}\alpha}\text{W}^{142}$) occurred mostly in major and minor sites, specifically in P2', P3 and P5. As expected, a great number of hydrogen bonds were established along the NplNLS, such as $_{\text{Imp}\alpha}\text{S}^{360}$, $_{\text{Imp}\alpha}\text{N}^{361}$, $_{\text{Imp}\alpha}\text{G}^{323}$, $_{\text{Imp}\alpha}\text{V}^{321}$, $_{\text{Imp}\alpha}\text{R}^{315}$, $_{\text{Imp}\alpha}\text{Y}^{277}$, $_{\text{Imp}\alpha}\text{R}^{238}$, $_{\text{Imp}\alpha}\text{A}^{148}$, $_{\text{Imp}\alpha}\text{G}^{150}$ and $_{\text{Imp}\alpha}\text{N}^{188}$.

Most of the aforementioned interactions occurred with side chains of the charged residues of the NplNLS. Moreover, we highlighted residues $_{\text{Imp}\alpha}\text{D}^{192}$, $_{\text{Imp}\alpha}\text{E}^{396}$, $_{\text{Imp}\alpha}\text{W}^{184}$, $_{\text{Imp}\alpha}\text{W}^{231}$, $_{\text{Imp}\alpha}\text{W}^{357}$, $_{\text{Imp}\alpha}\text{W}^{399}$ and $_{\text{Imp}\alpha}\text{Y}^{277}$ that were observed in more than 90% of the analyzed trajectory frames from MD and NM ensembles (S2 Table, S3 Table and S4 Table).

3.5 Discussion

3.5.1 Bending and twisting motions may be directly related to Impα function

Both bending and twisting motions are good candidates to adapt the Impα to the cNLS, because their motion pattern promotes conformational changes in the NLS binding pockets. The combination of these motion patterns appeared to be similarly recurring in other modes, such as NM8. The vectors from other high-collectivity modes, such as NM10-13 and NM17-18, also exhibited these two main motions but in smaller portions of the protein, and they were also associated with undefined types of motions. This finding

was expected because modes with higher frequency are normally associated with localized vibrations (KESKIN et al., 2002). Studies involving the protein Impβ – a solenoid protein similar to Impα (KOBE; KAJAVA, 2000; FORWOOD et al., 2010; KAJAVA, 2001) – suggested the importance of bending and twisting motions to generate the flexibility of Impβ, allowing it to bind to different types of proteins (FORWOOD et al., 2010; LEE et al., 2000). In our Impα computational analysis, the described movements could be equally important to adapt to different sizes of cargo proteins and NLSs, enhancing the contacts over the NLS binding site. Although no significant changes were observed by the geometric analysis from the Impα crystallographic structures bound to different types of bipartite NLSs, we must consider that the presence of only a peptide (approximately 20 amino acid length) submitted to similar crystallization conditions may not be sufficient to induce large conformational changes in this system. Therefore, the role of Impα motions is likely critical for the cNLS accommodation, considering the entire protein that contains it.

The MD/NM approaches complementarily showed the flexibility of Impα by the analysis of NM7-9 and PC1-3. The bending observed in NM7/PC1 may be directly related to the accessibility to the binding sites and the release of NLSs. This movement was observed in both Impα-NplNLS and Apo Impα, with small differences in their amplitude, resembling "open" and "close" movements. Bending motions have been reported in globular proteins to identify movements of domains along with opened and closed states (BROOKS; KARPLUS, 1985; ICHIYE; KARPLUS, 1991). The bending analysis for Impα was possible because its small curvature (KOBE; KAJAVA, 2000) allowed for the establishment of a plane comprising two distal (N- and C-termini) atoms and a central atom. Solenoid Impβ required a more complex analysis, determining angles between vectors projected onto a reference plane in each motif to evaluate the curvature changes (FORWOOD et al., 2010). Specifically, the concerted bending motion of Impα NM7 could operate as an opening-closing gateway, allowing the NLS entrance and adjusting the amplitude of this motion in relation to the cargo protein size. The absence of a ligand may imply in a wider curvature favoring the access to the inner concave surface of Impα, allowing the IBB domain or cargo protein binding. Pumroy et al. (2015) compared the flexibility of three human Impα isoforms (α-1, α-3 and α-7) considering their bound and unbound states and applying MD simulations. Based on the protein end-to-end distance measurements, the authors observed an increase in the flexibility of Apo Impα isoforms, in accordance with our bending characterization that showed a higher radius of curvature for Apo Impα.

According to Kobe and Kajava (2000), the "twist" takes into account the rotations of the neighboring repeats relative to each other along the backbone direction, and ARM repeat proteins have large twist movements, allowing the accommodation of extended and flexible peptides. The twist observed in the NM9/PC3 of Impα-NplNLS showed similarities to each other, confirming the observations of Hayward, Kitao and Berendsen

(1997), which compared NM/PC for lysozyme protein. Although Impα-NplNLS and Apo Impα had similar average values (Fig 4B), the differences were found only in individual analyses of the neighboring ARM repeats. The angle variation between ARM6 and ARM7 in NM9 of Apo Impα (Fig 5) is remarkable, and a similar result was already observed for Apo Impα-3 (PUMROY et al., 2015). The lack of a ligand (IBB domain or NLS) interacting with the Apo Impα concave portion may lead to a higher twist in the middle of Apo Impα due to the lack of interactions that stabilize and provide binding specificity. The increment of the angle amplitude in this region was contrasted by lower values for the remaining protein regions compared to Impα-NplNLS. Therefore, Impα-NplNLS may require conformational changes for the NLS adjustment to the binding pockets of the Impα inner surface, which explains the greater variation in the ARM's angles in contrast to Apo Impα, which would not require such adjustments. This same explanation applies to the amplitude variations also found in NM8 for the bound/unbound states of Impα.

Experiments and MD simulation with *Neurospora crassa* Impα emphasize the instability of an N-terminally truncated Impα in the absence of an NplNLS peptide, indicating the occupancy of the NLS binding sites as a requirement for Impα crystallization (TAKEDA et al., 2013). In addition, Falces et al. (2010) also showed, performing circular dichroism assays with Impα-1ΔIBB from *Xenopus laevis*, the stabilization of the protein upon association with Npl, thus reinforcing the aforementioned observations. The establishment of polar contacts settles the NLS backbone, whereas hydrophobic and electrostatic interactions with positively charged NLS residues allow the specificity (FONTES et al., 2003; KOBE, 1999). The greater structural flexibility of Apo Impα was only observed by the geometrical analysis for the NM7 bending. Despite the non-association of the NM8/NM9 twisting motions to a greater overall flexibility of Impα, the NLS absence allowed for a greater motion amplitude of other NMs, such as modes 21-27, comparing the restriction energies of Impα-NplNLS (S4 Fig) and Apo Impα (S5 Fig), indicating that the higher flexibility of Apo Impα appeared to be distributed over these modes. In addition, considering the Cα fluctuations of Impα in its bound and unbound states (S9 Fig), we observed a higher flexibility in the major site of Apo Impα. This result was also observed for Apo Impα-3 (PUMROY et al., 2015), and the authors predicted a weaker binding for NLSs that relies primarily on the major site. However, this statement does not seem to apply to our case because Impα-2 has been crystallized with different types of NLSs, including those with preference to the major site; e.g., simian virus 40 (SV40) NLS (FONTES; TEH; KOBE, 2000). Moreover, the binding stability of Impα-2 to some of those NLSs was also observed with ligand binding assays (MARFORI et al., 2012; KIRBY et al., 2015; BARROS et al., 2016). In summary, the NLS binding stabilizes Impα, apparently restricting the motion range represented by some higher-frequency NMs; however, it may also allow localized adjustments between the ARM repeats to improve the overall affinity to the bound NLS.

3.5.2 The role of the linker residues in cNLS recognition

The cross-correlation calculations performed for both MD simulations (Fig 6A) and displacement along NMs (Fig 6B) showed positive correlations from residues of both major and minor sites of Impα. Further analysis showed that the majority of positive correlations observed could be explained by the correspondence of salt bridges, hydrogen bonds and hydrophobic interactions. We could determine that positions P2 and P5 from the major site and P1' and P2' of the minor binding site have high levels of contact with the NplNLS, based on the occupancy values of some interactions in those regions. These data confirm the maintenance of the main contacts between Impα and NplNLS, showing that our simulation results are in agreement with the experimental data (MARFORI et al., 2012; FONTES; TEH; KOBE, 2000).

A classical bipartite NLS sequence interacts simultaneously with major and minor binding sites of Impα and depends on a linker region containing a minimum of 10 residues between P2' and P2 positions to act as a cNLS (CHRISTIE et al., 2015; BARROS et al., 2012). The linker region in the NplNLS appears to play a key role in the process of cNLS recognition. The cross-correlation maps clearly indicated an involvement of this region in the promotion of contacts with Impα. A detailed analysis of the interactions showed the occurrence of salt bridges, hydrogen bonds and hydrophobic contacts along the NplNLS. The involvement of $_{\text{Imp}\alpha}\text{R}^{238}$, $_{\text{Imp}\alpha}\text{R}^{315}$, $_{\text{Imp}\alpha}\text{W}^{273}$ and $_{\text{Imp}\alpha}\text{Y}^{277}$ was already observed in Impα-NplNLS crystal structures (MARFORI et al., 2012; FONTES; TEH; KOBE, 2000). However, the computational approach indicated other interactions previously reported in single structures; $_{\text{Imp}\alpha}\text{K}^{353}$ and $_{\text{Imp}\alpha}\text{N}^{350}$ were reported with the NLS from FEN-1 (BARROS et al., 2012) and $_{\text{Imp}\alpha}\text{D}^{280}$ from SV40 NLS bound to the Impα from a filamentous fungus (BERNARDES et al., 2015). Moreover, new possible contacts were detected; e.g., residues $_{\text{Imp}\alpha}\text{N}^{446}$, $_{\text{Imp}\alpha}\text{R}^{101}$ and $_{\text{Imp}\alpha}\text{R}^{102}$ interacted in the N- and C-terminal regions of the NLS.

Our simulation analyses suggest that the linker contacts are important to settle the cNLS, and help to accommodate the cNLS side chains into the grooves of major and minor binding sites. The linker region, in addition to the N- and C-terminals of an NLS, may also compensate for interactions for the establishment of an activation pattern of the other NLS region (KOSUGI et al., 2009), which indicates that the linker contacts may occur in different NLSs and can be maintained after the docking of the bipartite NLS in both major and minor binding sites, depending on the residues composing the linker, such as proline and acidic amino acids (CHRISTIE et al., 2015; KOSUGI et al., 2009). In summary, the recurrent presence of residues outside the major and minor sites strongly reinforces their importance for the proper binding of the NLS, which corroborates other studies involving other NLSs (FONTES et al., 2003; CONTI; KURIYAN, 2000; CHEN et al., 2005; CUTRESS et al., 2008; GIESECKE; STEWART, 2010; YANG et al., 2010) and

further encourages us to understand in greater detail their roles and effects during the nuclear import process.

3.5.3 NM analysis with classic MD simulations

MD simulations and NM analysis have been used for macromolecules as a complementary analysis to the experimental data to describe their main motions and relate to a specific function (ALEXANDROV et al., 2005; BROOKS; KARPLUS, 1985; ICHIYE; KARPLUS, 1991; SKJAERVEN; MARTINEZ; REUTER, 2011). The approach used for the Impα-cNLS complex combines NMs robustness – to evaluate wider protein movements – with the reference structures from the MD and indicates the benefits of this association in protein-peptide analysis.

The data sampling from MD simulations, despite the time constraints and convergence difficulties associated with this technique (GROSSFIELD; ZUCKERMAN, 2009; GENHEDEN; RYDE, 2012), were balanced among the selected reference structures. Considering the time of the simulations, there is a likely stability of the studied system. Although we have not performed classical MD of Apo Impα, Takeda et al. (2013) reinforce the instability of an N-terminally truncated Impα in the absence of an NLS peptide. One way of computationally evaluating the N-terminally truncated Impα would be applying long MDs to observe the effects of high protein flexibility. However, with a simple NM calculation, we could observe in general, based on the restraint energy profile, RMSF values and geometrical analysis, the higher flexibility of Impα in the absence of the NLS.

The MD data with the Impα-NplNLS complex not only supported the maintenance of the protein-peptide complex but also showed some of the major movements and interactions occurring at the complex interface. NMs promoted an analytical method to access the dynamics of the system, allowing the possibility of the recognition of new interactions and dribbling the convergence and conformational restrictions from standard MD. The main motions described here in NMs were also obtained in the PC analysis, in agreement with the comparison between the lowest NMs and the first PCs obtained from MD simulations in the lysozyme model (HAYWARD; KITAO; BERENDSEN, 1997) and a subunit of the GroEl chaperone (SKJAERVEN; MARTINEZ; REUTER, 2011). The highlighted movements are of high occurrence in the protein's lifetime and are likely to be functionally important. Moreover, the low computational cost of NM is once more an attractive feature for application in biological systems.

3.6 Conclusions

Computational approaches of MD and NM analysis were combined to evaluate the main of motions of Impα and its interaction to a cNLS peptide. The bending motion may be

involved in the NLS entrance and the accommodation of cargo protein depending on its size, whereas the twisting motions may be involved in the NLS recognition and accommodation into the Impα binding sites. The combination of these movements could allow local adjustments between the ARM repeats, which could improve the overall affinity to the cNLS. The absence of an NLS was also evaluated and may imply in a wider curvature of Apo Impα, allowing for the IBB domain or cargo protein binding. Moreover, a higher twist in the middle of Apo Impα was detected possibly due to the lack of interactions that stabilize and provide binding specificity, which could explain the challenges in crystallizing N-terminally truncated Apo Impα.

The evaluation of salt bridges, hydrogen bonds and hydrophobic interactions corroborates the fundamental interactions between Impα and NplNLS and gives additional support for interactions outside the classical binding pockets that are important during this process. The linker contacts in cNLS assist the adjustment of the peptide backbone, which helps the interactions between cNLS side chains and residues from major and minor binding grooves. In conclusion, MD simulations combined to NM analysis supported the maintenance of the Impα-NplNLS complex exploring the conformational space and accessing the dynamics of the system with a lower computational cost. This approach may help to understand the affinities between Impα and cNLS peptides and non-classic NLSs.

3.7 Acknowledgments

This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [grant numbers 2012/19447-2; 2014/21976-9] and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [grant number 142110/2012-4]. We are grateful to Prof. Dr. Cesar Martins for providing computational access for running part of the simulations.

3.8 Figures

488

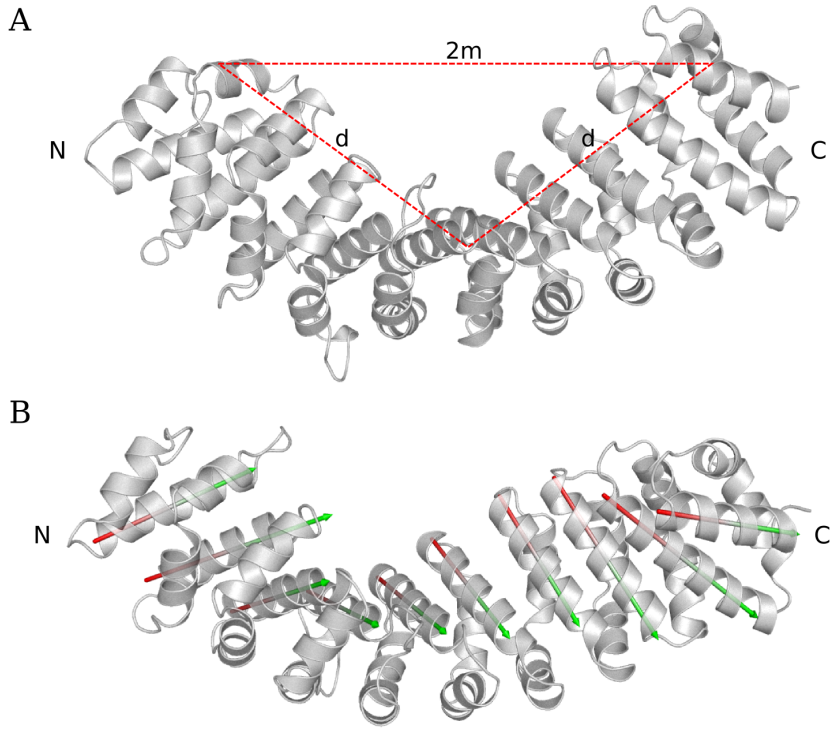


Fig 1 – Scheme of geometrical methods adopted for the description of the Impα motions. (A) Display of the selected vertices and the determined distances (2m and d) for the calculation of the radius of curvature. (B) The vectors generated to H3 for each ARM repeat.

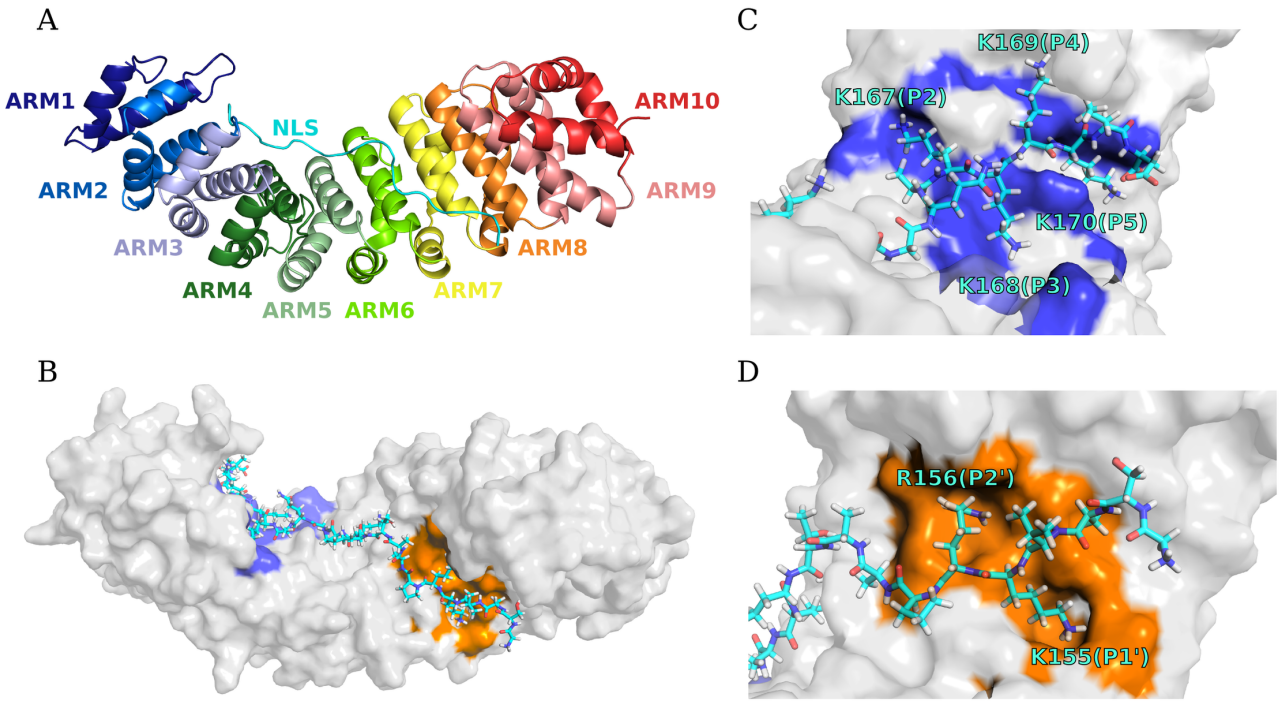


Fig 2 – The starting structure of Impα-NplNLS for MD simulations. (A) The Impα as a cartoon diagram colored based on each ARM repeat as a rainbow spectrum from N-terminal (blue) to C-terminal (red) and the NplNLS as a cyan cartoon diagram positioned in an antiparallel configuration compared to Impα. (B) The surface representation of Impα with the NplNLS as a cyan stick diagram, indicating both major (blue) and minor (orange) binding sites. (C) The major site zoom indicating positions P2-P5 and (D) the minor site zoom in P1' and P2'. In both sites, the positively charged side chains are positioned in the main pockets of the Impα binding core.

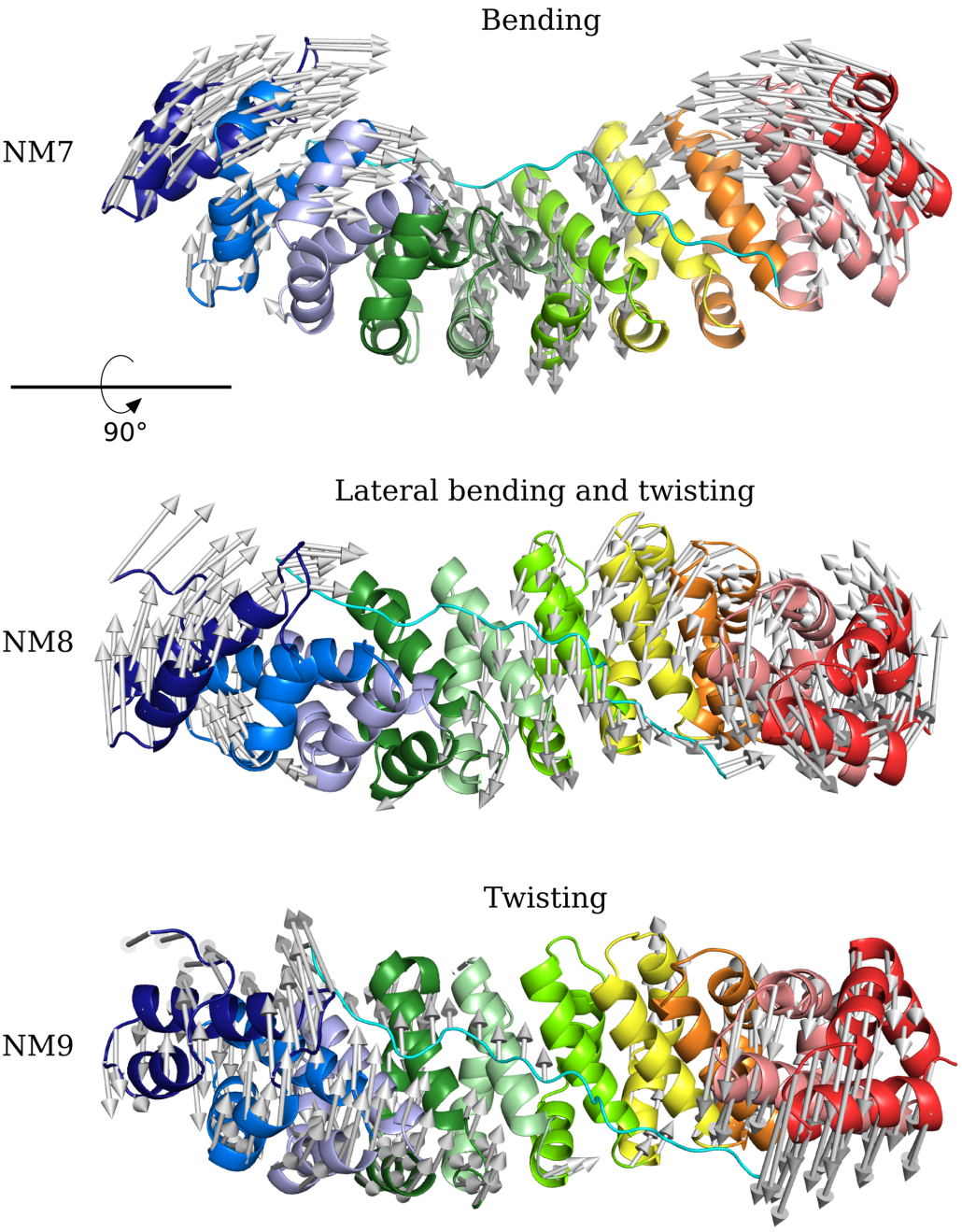


Fig 3 – Main motions observed from NM analysis of Impα–NplNLS. Impα (cartoon model) is shown as a rainbow spectrum from N-terminal (blue) to C-terminal (red), and NplNLS (cyan cartoon model) is positioned in an antiparallel configuration compared to Impα. The vector arrows for NM7-9 are shown with the correspondent description of the motion. NM7 is shown in a front view, whereas NM8 and NM9 are shown in an upper view (90º rotation in the X-axis).

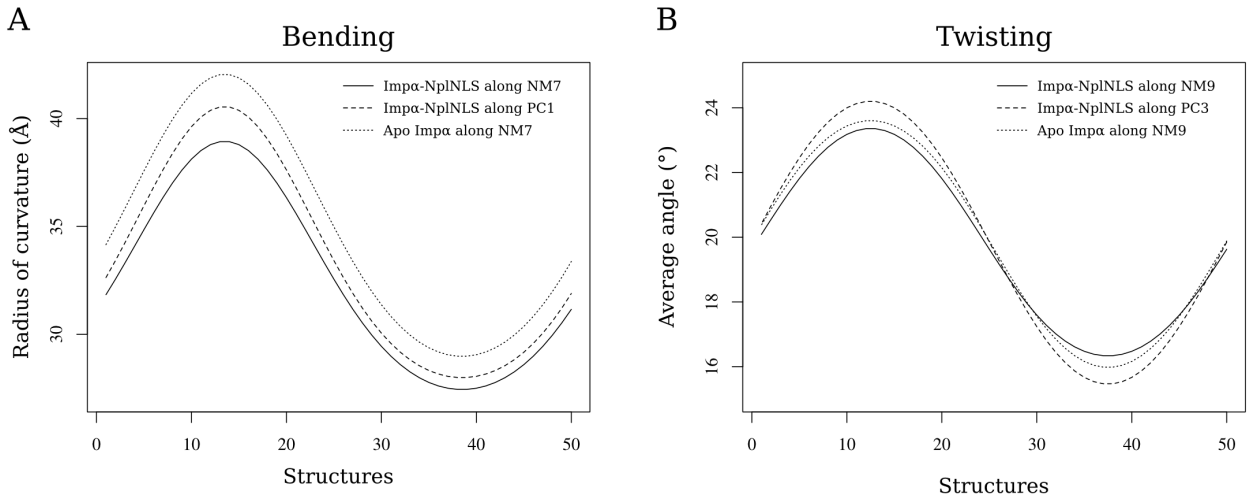


Fig 4 – Geometric analysis of the bending and twisting motions of Impα. (A) The bending motion was quantitatively characterized by the radius of curvature along NM7 (solid line) and PC1 (dashed line) for Impα-NplNLS and along NM7 for Apo Impα (dotted line), whereas (B) the twisting motion was quantitatively characterized by the average values for the angles between helices along NM9 (solid line) and PC3 (dashed line) for Impα-NplNLS and along NM9 for Apo Impα (dotted line).

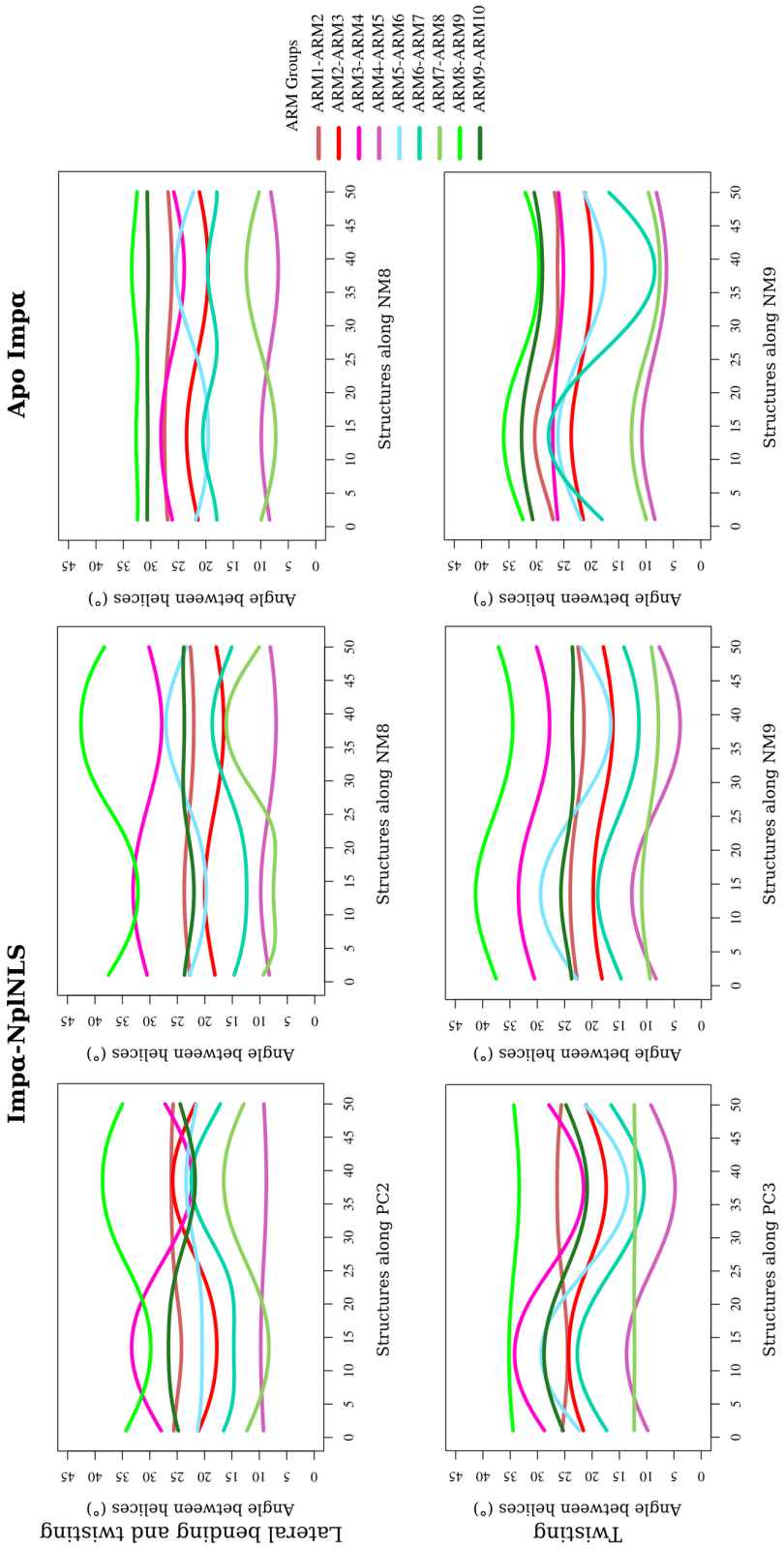


Fig 5 – Angle between helices of Impα. The angles between neighboring H3 pairs from the motions described as lateral-bending/twisting (along PC2 and NM8) and twisting (along PC3 and NM9) for Impα–NplNLS and Apo Impα. The ARM groups considered for each angle calculation are depicted with different color assignments.

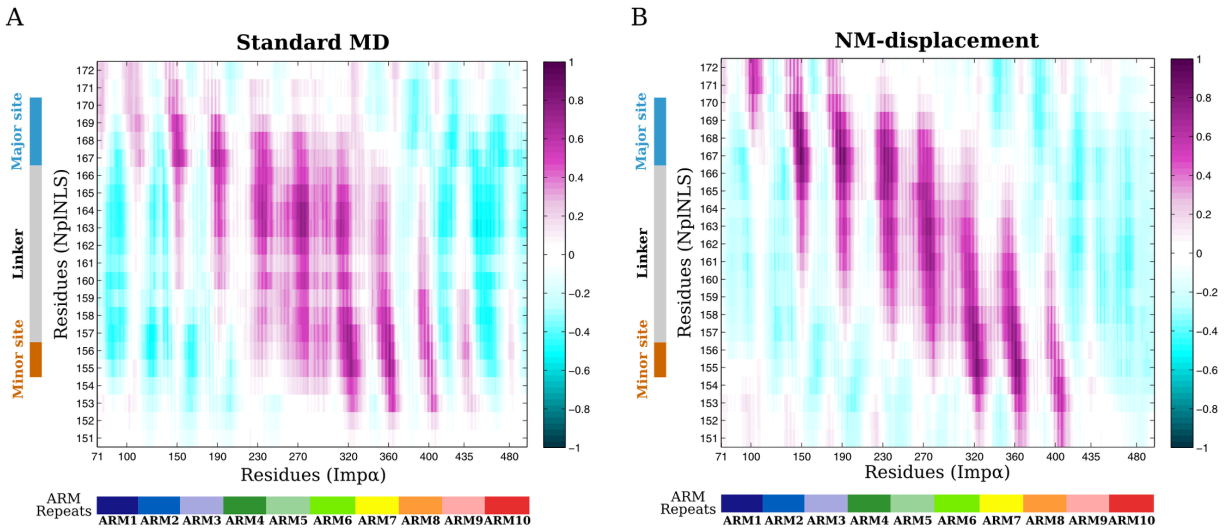


Fig 6 – Heatmap of cross-correlations between Impα and NplNLS. (A) Trajectories from standard MD simulations (300 ns ensemble) and (B) NM-displacement (ensemble from references structures 67,730 ps, 207,080 ps and 274,970 ps) were used for the calculation of correlations. A color bar indicates the degree of correlation from anti-correlated (negative values) to correlated (positive values) residues. The X and Y axes, respectively, show the position of each ARM repeat in relation to the Impα sequence and the NplNLS residues in contact with the protein binding sites.

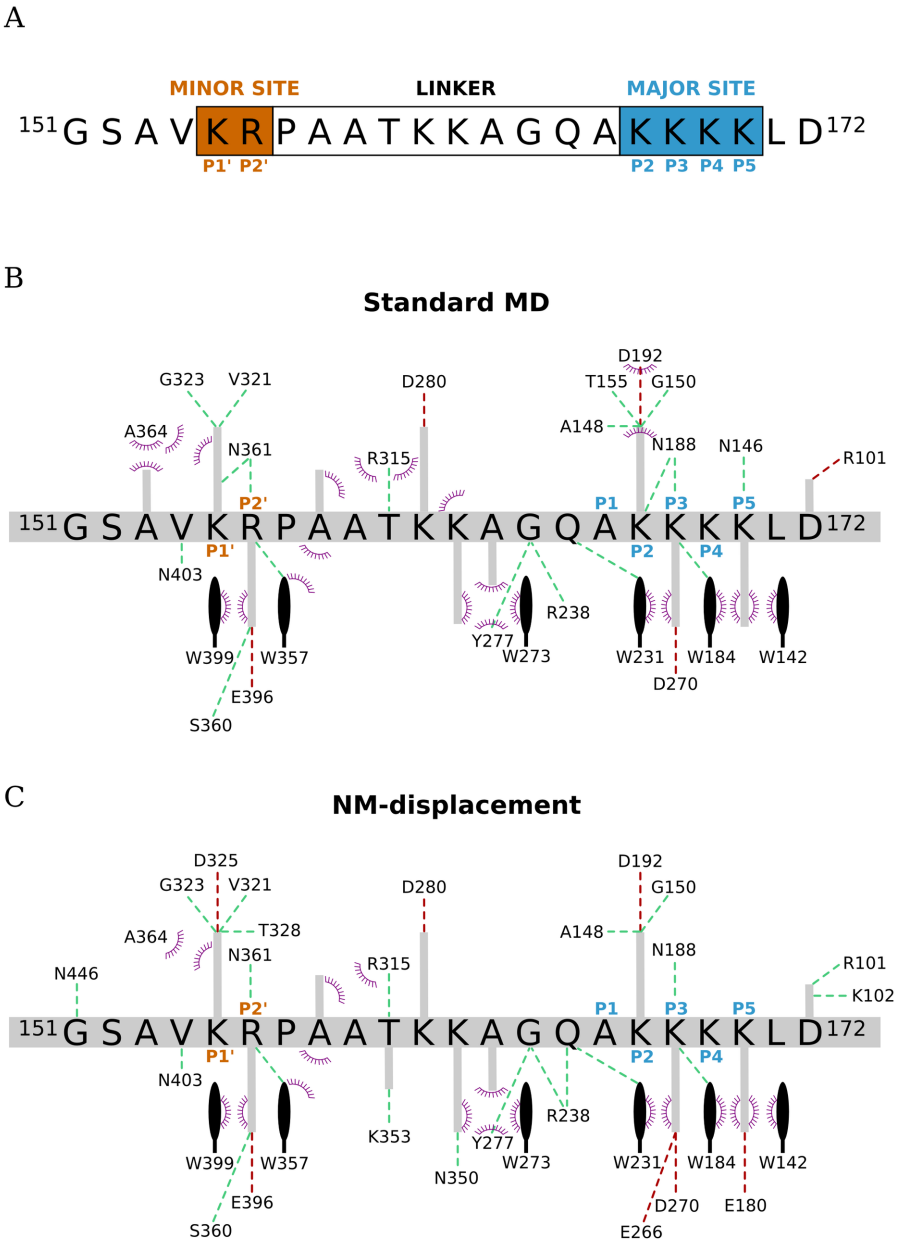
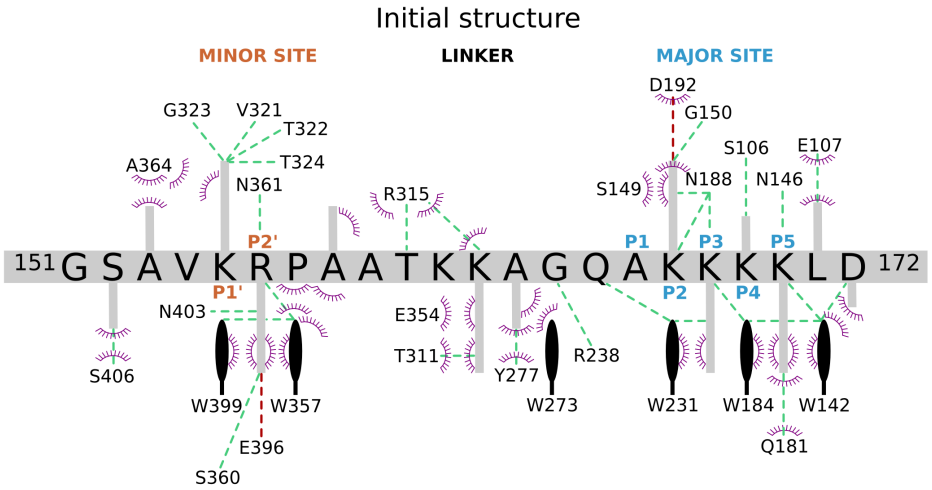


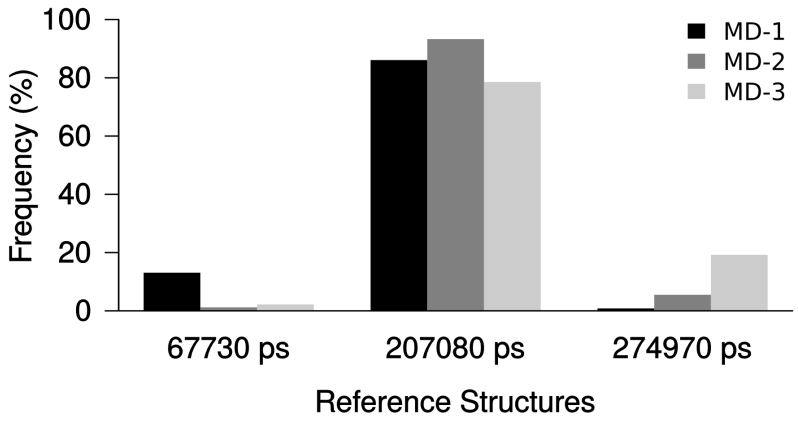
Fig 7 – Schematic representation of the interactions observed in the Impα-NplNLS interface. (A) The major (P2-P5) and minor (P1'-P2') sites and the linker region are indicated in the sequence of the NplNLS. (B) The standard MD (300 ns ensemble) and (C) NM-displacement (ensemble from reference structures 67,730 ps, 207,080 ps and 274,970 ps) interaction scheme is shown with salt bridges (red) and hydrogen bonds (green) as dashed-lines, and hydrophobic contacts are shown as arcs with radiating spokes. The important tryptophan residues that mediate the hydrophobic contacts are depicted in the scheme as black sticks. The main chain of the NplNLS is represented as a gray horizontal line with its respective amino acid sequence, together with side chains shown as perpendicular lines. Only interactions that had an occupancy rate $\geq$50% of the analyzed trajectories are indicated in the scheme.

3.9 Supporting Information

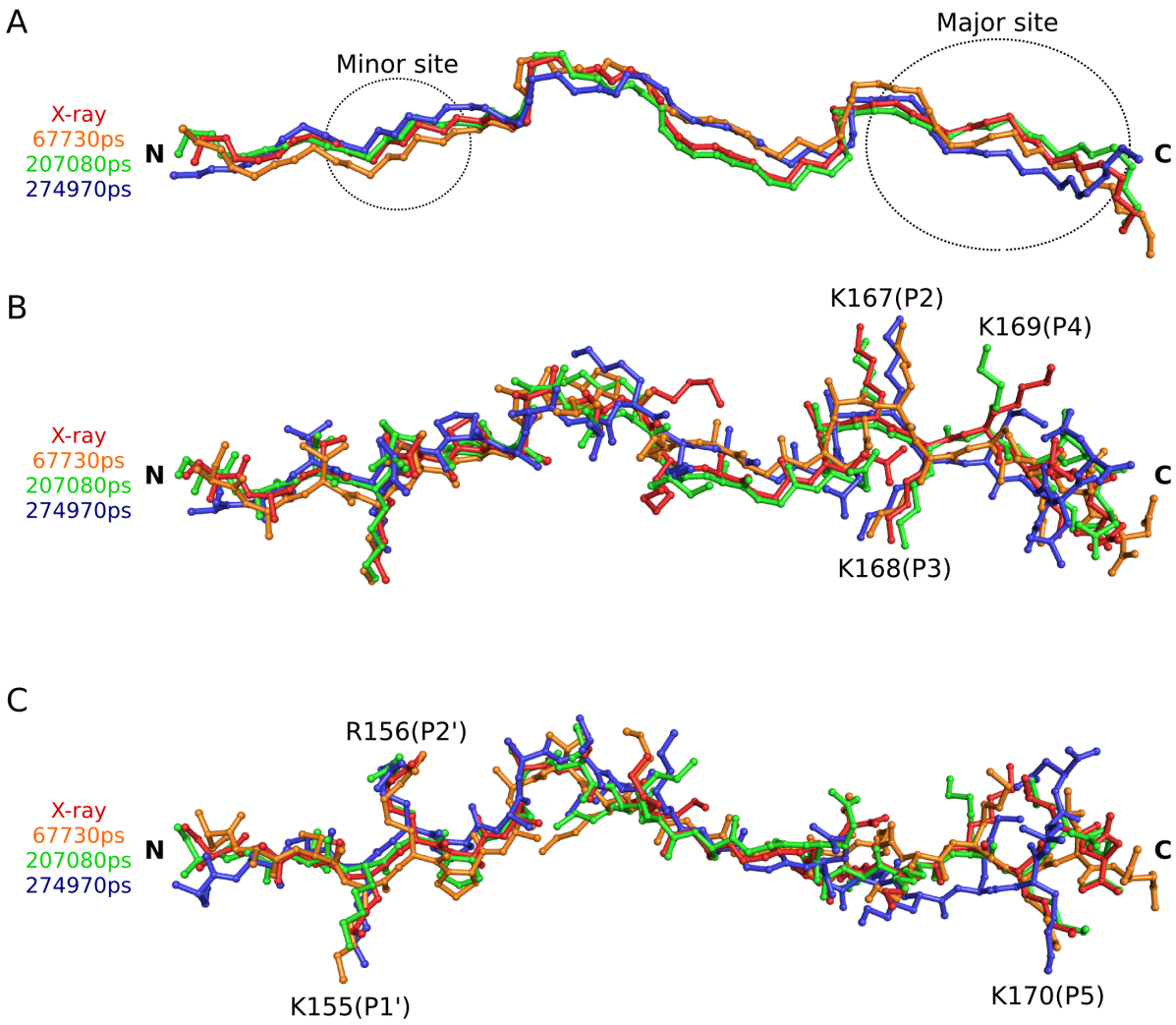
Figures



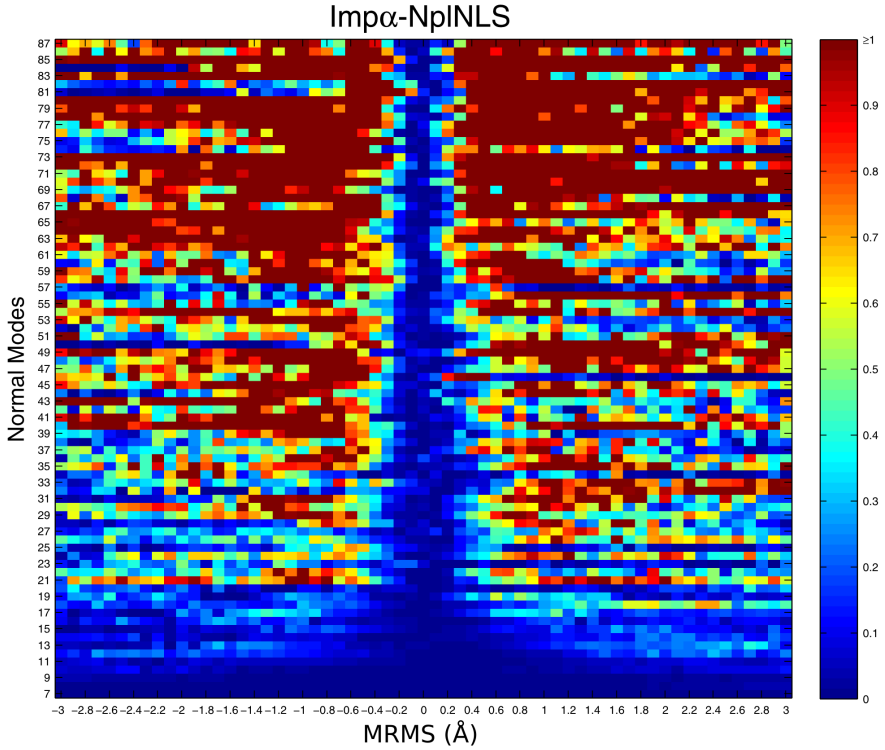
S1 Fig: Interactions from the starting structure. Scheme of interactions of the starting structure for MD simulations. The representation is similar to the description of Fig 7.



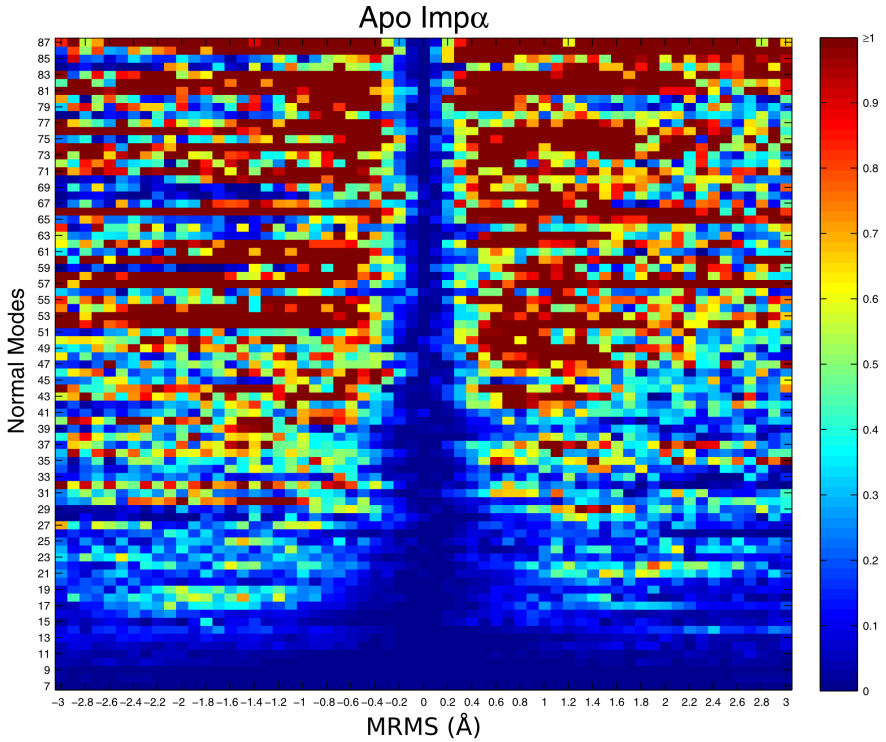
S2 Fig: Convergence analysis. Bar plot indicating the frequency of structures clustered within each reference (67,730 ps, 207,080 ps and 274,970 ps) from the MD simulations.



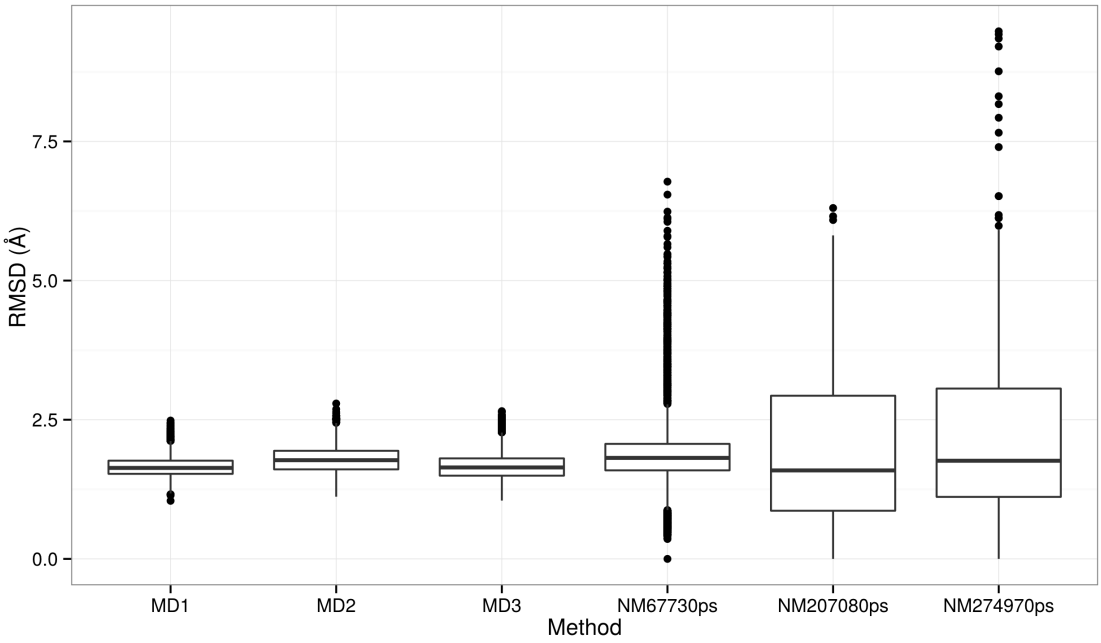
S3 Fig: Structural alignment of NplNLS. Structural representation of the NplNLSs from reference structures 67,730 ps (orange), 207,080 ps (green) and 274,970 ps (blue) aligned with the NplNLS from X-ray (red). (A) Stick diagram of the main chain, with major and minor sites indicated. (B-C) Stick diagram including side chains and the residues from each site.



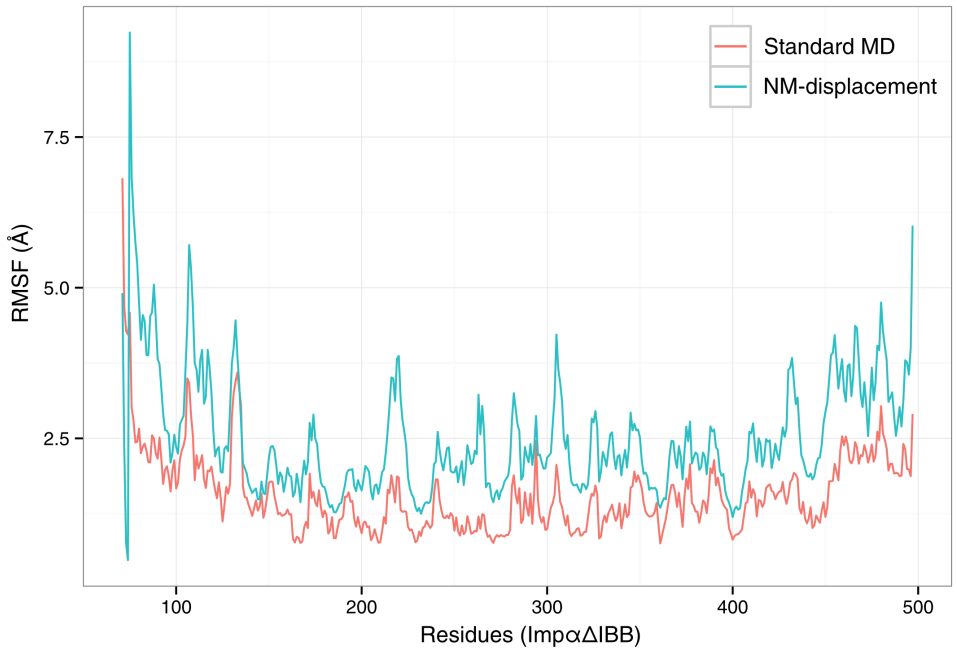
S4 Fig: Heatmap of the total restraint energy values from Impα-NplNLS. The values are from the structures generated from the NM-displacement of the most representative reference structure (207,080 ps) according to the convergence analysis. The X-axis is the displacement range, represented as values of mass-weighted root mean square (MRMS), and the Y-axis is the NM numbers. Lower values of energy (blue tons) indicate favorable conformations.



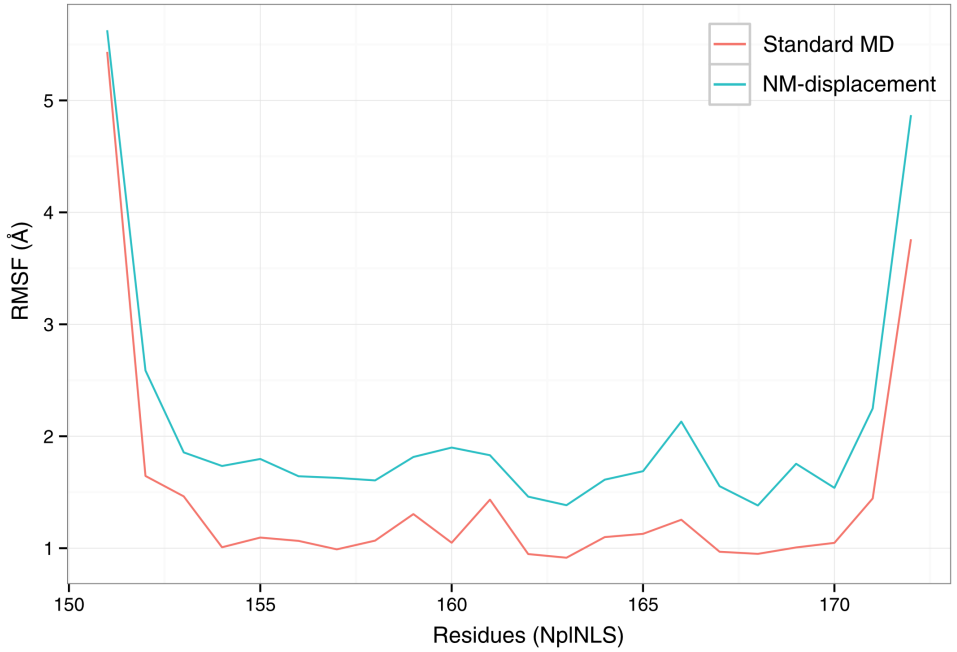
S5 Fig: Heatmap of the total restraint energy values from Apo Impα. The values are from the structures generated from the NM-displacement. The X-axis is the displacement range, represented as values of MRMS, and the Y-axis is the NM numbers. Lower values of energy (blue tons) indicate favorable conformations.



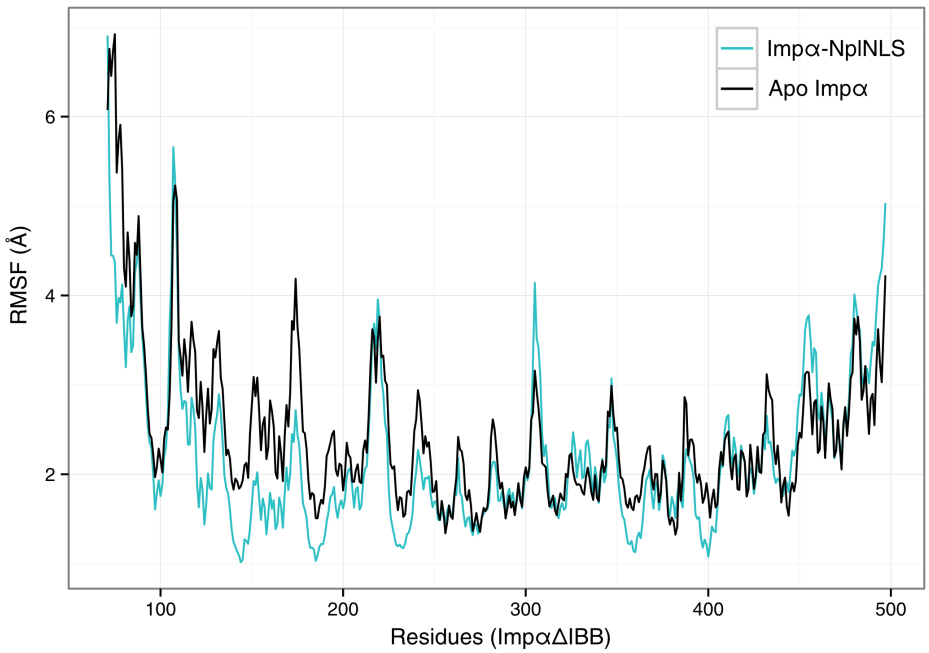
S6 Fig: Conformational exploration. Box-plot of the RMSD distribution from the trajectories of standard MD (MD-1, MD-2 and MD-3) and NM-displacement (67,730 ps, 207,080 ps and 274,970 ps).



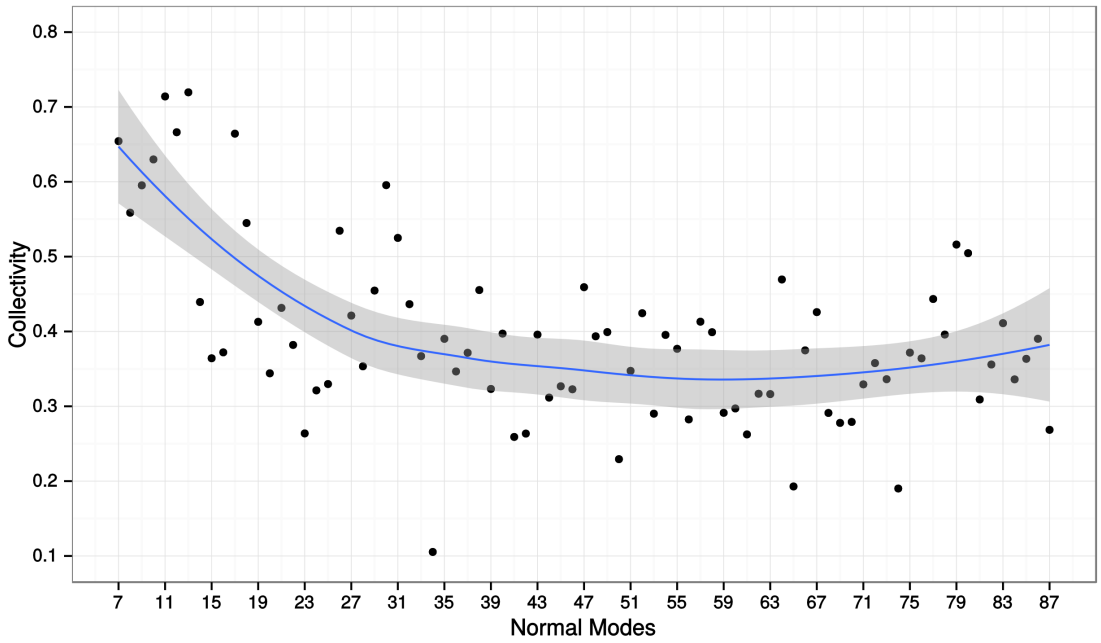
S7 Fig: Residue fluctuations of Impα–NplNLS. Residue fluctuations based on Cα RMSF of Impα from the ensemble trajectories of standard MD (red) and NM-displacement (cyan).



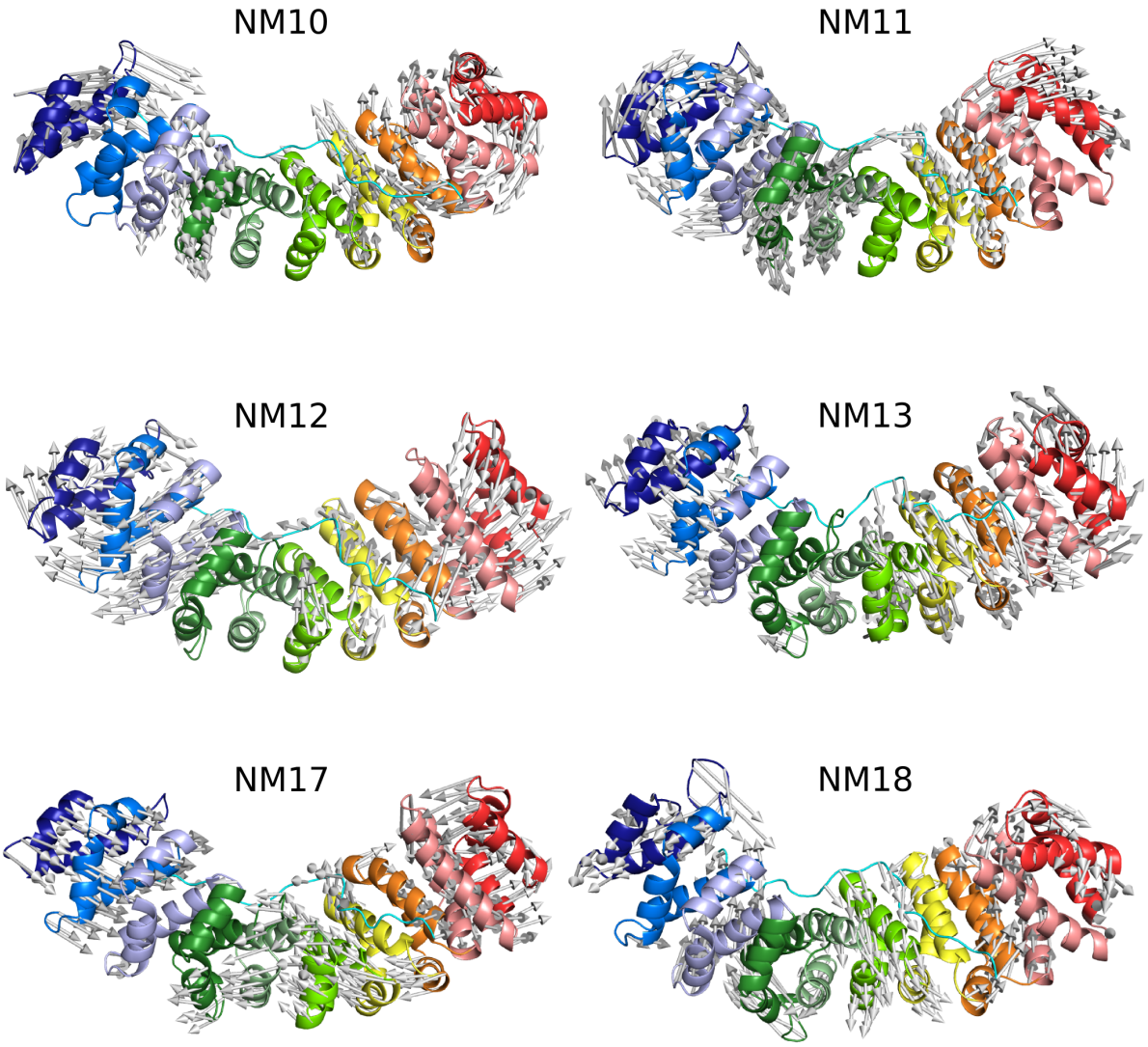
S8 Fig: Residue fluctuations of NplNLS. Residue fluctuations based on Cα RMSF of NplNLS from the ensemble trajectories of standard MD (red) and NM-displacement (cyan).



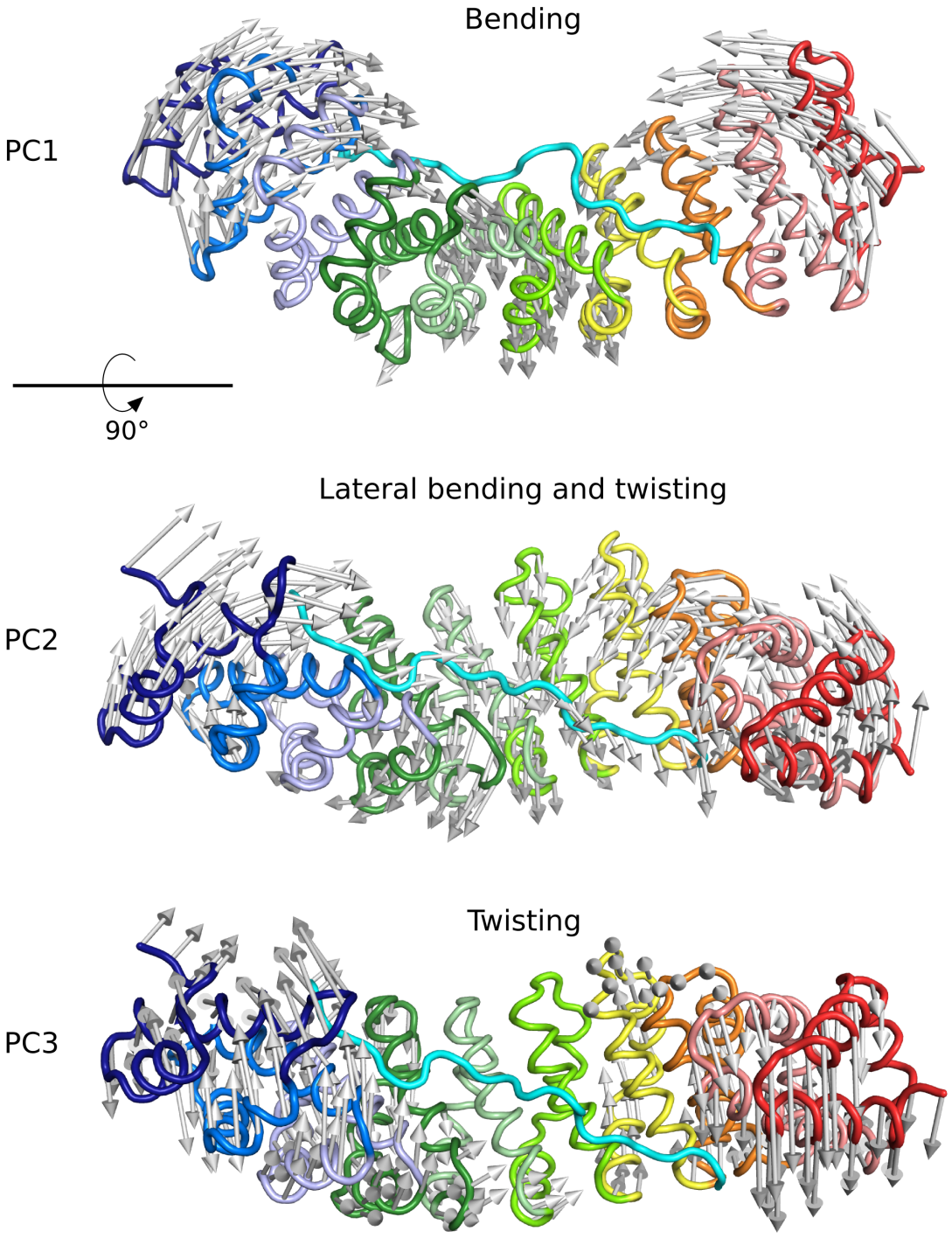
S9 Fig: Residue fluctuations of apo Imp$\alpha$$\Delta$IBB. Residue fluctuations based on Cα RMSF of Impα from the trajectories of the NM-displacement of Impα–NplNLS (from reference structure 207,080 ps; cyan) and Apo Imp$\alpha$$\Delta$IBB (black).



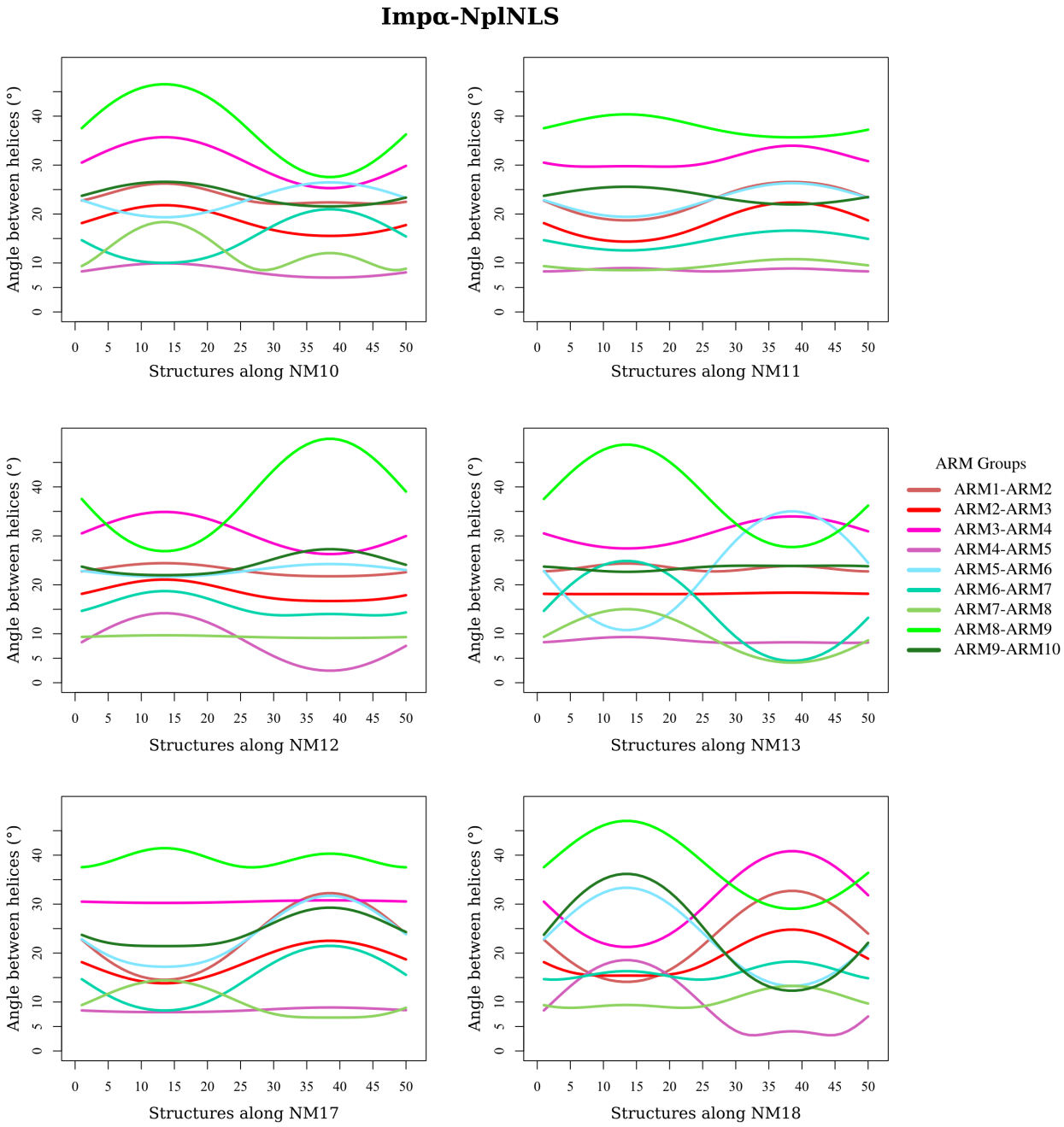
S10 Fig: Collectivity from Impα–NplNLS. The collectivity values are plotted for each NM, and a smoothed line is fitted (blue line), representing the data tendency. The shaded area is the confidence interval around the smoothed line calculated with the ggplot package (http://ggplot2.org/) in R.



S11 Fig: Main motions observed in NMs10-13, NM17 and NM18. The vector arrows indicating the motions are shown. The Impα is displayed in a cartoon diagram, with each ARM colored from blue to red, corresponding to N to C-terminals. The NplNLS (cyan) is in a cartoon representation and is positioned in an antiparallel configuration compared to Impα.



S12 Fig: Main motions observed in PCs1-3 from standard MD. The vector arrows indicating the motions are shown. The Impα is displayed in a Cα representation, with each ARM colored from blue to red, corresponding to N to C-terminals. The NplNLS (cyan) is in Cα representation and is positioned in an antiparallel configuration compared to Impα.



S13 Fig: Angles between helices of Impα–NplNLS for NMs10-13, NM17 and NM18. The ARM groups considered for each angle calculation are depicted with different color assignments, similar to Fig 5.

Movies

S1 Movie: Movie of NM7. Impα is shown in a cartoon diagram, and it is colored according to each ARM repeat, from the N-terminal (blue) to the C-terminal (red) ends of the protein. The NplNLS is represented in a cyan cartoon diagram positioned in an antiparallel configuration compared to Impα. To view this movie file, access the following link: <http://journals.plos.org/plosone/article/asset?unique&id=info: doi/10.1371/journal.pone.0157162.s001>

S2 Movie: Movie of NM8. Impα is shown in a cartoon diagram, and it is colored according to each ARM repeat, from the N-terminal (blue) to the C-terminal (red) ends of the protein. The NplNLS is represented in a cyan cartoon diagram positioned in an antiparallel configuration compared to Impα. To view this movie file, access the following link: <http://journals.plos.org/plosone/article/asset?unique&id=info: doi/10.1371/journal.pone.0157162.s002>

S3 Movie: Movie of NM9. Impα is shown in a cartoon diagram, and it is colored according to each ARM repeat, from the N-terminal (blue) to the C-terminal (red) ends of the protein. The NplNLS is represented in a cyan cartoon diagram positioned in an antiparallel configuration compared to Impα. To view this movie file, access the following link: <http://journals.plos.org/plosone/article/asset?unique&id=info: doi/10.1371/journal.pone.0157162.s003>

Tables

S1 Table: Backbone RMSD values from structural alignment. The reference structures are aligned with the Impα-NplNLS crystallographic structure (PDB ID 3UL1).

Backbone RMSD	67,730 ps	207,080 ps	274,970 ps
Impα	2.12439	1.15923	2.482
NplNLS	1.5432	0.759316	1.70461

S2 Table: Salt bridges occupancies. The occupancies of salt bridges between NplNLS and Impα in standard MD and NM-displacement. Interactions that were above 50% of occupancy are highlighted in gray.

Salt Bridges – Standard MD		
NLS	Impα	Occupancies (%)
R156	E396	99.97
D172	R101	64.98
D172	K102	44.85
E153	R117	0.47
K155	D325	2.16
K161	D280	51.68
K162	E354	7.24
K167	D192	96.82
K168	D270	4.00
K168	E266	1.50
K169	E107	4.75
K170	E180	20.04

Salt Bridges – NM displacement		
NLS	Impα	Occupancies (%)
R156	D433	0.06
R156	E396	99.94
D160	R117	0.38
D160	H203	0.06
D172	R101	33.29
D172	R106	0.04
D172	R117	0.02
D172	R227	0.84
D172	H177	0.02
D172	K102	26.99
D172	K108	0.36
E153	R117	1.63
E153	K108	0.06
K155	D325	93.03
K161	D280	99.77
K161	E354	0.02
K162	E354	43.61
K167	D192	98.81
K167	E107	0.04
K168	D270	74.51
K168	E266	100.00
K169	D113	0.08
K169	D192	0.04
K169	E107	33.08
K169	E266	0.88
K169	E306	0.04
K170	E180	80.12
K170	E266	0.33

S3 Table: Hydrogen bonds occupancies. The occupancies of hydrogen bonds between NplNLS (blue) and Impα (green) in standard MD and NM-displacement. Interactions that were above 50% of occupancy are highlighted in gray.

Hydrogen Bonds – Standard MD		
Donor	Acceptor	Occupancies (%)
R156-Side	E396-Side	99.94%
W357-Side	R156-Main	99.61%
G164-Main	Y277-Side	98.62%
R315-Side	T160-Main	98.45%
R156-Main	N361-Side	98.19%
K167-Side	D192-Side	96.58%
K168-Main	N188-Side	96.12%
K155-Side	V321-Main	93.05%
W184-Side	K168-Main	92.35%
W231-Side	Q165-Main	92.13%
N188-Side	K168-Main	85.79%
K167-Side	G150-Main	84.58%
N361-Side	R156-Main	84.33%
R238-Side	G164-Main	83.91%
K167-Side	A148-Main	77.50%
K155-Side	G323-Main	73.12%
R101-Side	D172-Side	65.59%
V154-Main	N403-Side	64.45%
K167-Main	N188-Side	64.25%
N146-Side	K170-Main	64.21%
K170-Main	N146-Side	62.96%
K155-Side	N361-Main	56.44%
K167-Side	T155-Side	52.82%
N403-Side	V154-Main	52.75%
K161-Side	D280-Side	51.89%
R156-Side	S360-Side	50.64%
W142-Side	K170-Main	47.02%
K170-Side	Q181-Side	46.69%
K102-Side	D172-Side	45.34%
K155-Main	N361-Side	44.13%
K167-Side	N188-Main	43.02%
A153-Main	N403-Side	38.06%
N235-Side	A166-Main	37.63%
W231-Side	A166-Main	32.86%
R315-Side	T160-Side	30.54%
K170-Side	W142-Side	29.99%
G191-Main	A166-Main	26.68%
S149-Side	K168-Main	25.98%
A163-Side	Y277-Side	25.12%
K155-Side	T322-Main	24.25%
K168-Side	N228-Side	23.82%
K155-Side	T328-Side	23.68%
W184-Side	K169-Main	22.01%
K167-Side	S149-Main	21.88%

Hydrogen Bonds – NM displacement		
Donor	Acceptor	Occupancies (%)
K168-Side	E266-Side	99.98%
R156-Side	E396-Side	99.92%
K161-Side	D280-Side	99.81%
K167-Side	D192-Side	98.81%
K167-Side	G150-Main	98.37%
R238-Side	Q165-Main	97.47%
K167-Side	A148-Main	96.53%
G164-Main	Y277-Side	96.46%
W357-Side	R156-Main	94.67%
R238-Side	G164-Main	93.39%
K155-Side	D325-Side	92.99%
R156-Side	S360-Side	92.18%
K155-Side	V321-Main	89.64%
R315-Side	T160-Main	88.91%
K155-Side	T328-Side	86.59%
W184-Side	K168-Main	85.40%
N361-Side	R156-Main	84.44%
W231-Side	Q165-Main	83.74%
K170-Side	E180-Side	80.15%
K162-Side	N350-Side	75.31%
K168-Side	D270-Side	74.27%
R156-Main	N361-Side	72.24%
K155-Side	G323-Main	72.07%
R101-Side	D172-Side	66.44%
V154-Main	N403-Side	64.96%
N188-Side	K168-Main	63.24%
K353-Side	T160-Side	58.60%
K102-Side	D172-Side	54.10%
G151-Main	N446-Side	52.38%
N403-Side	V154-Main	51.05%
W231-Side	A166-Main	46.61%
K162-Side	E354-Side	44.25%
N403-Side	R156-Side	42.22%
K170-Side	Q181-Side	42.20%
W142-Side	K170-Main	41.57%
A158-Main	W357-Side	41.53%
R156-Side	N361-Side	39.87%
A153-Main	N403-Side	39.35%
Q165-Side	D270-Side	37.95%
G151-Main	S406-Main	37.89%
K168-Main	N188-Side	37.41%
K167-Side	N188-Side	37.22%
G151-Main	D442-Side	36.80%
N146-Side	K170-Main	36.49%

S4 Table: Hydrophobic contacts occupancies. The occupancies of hydrophobic contacts between NplNLS and Impα in standard MD and NM-displacement. Interactions that were above 50% of occupancy are highlighted in gray.

Hydrophobic Contacts – Standard MD			Hydrophobic Contacts – NM displacement		
NLS	Impα	Occupancies (%)	NLS	Impα	Occupancies (%)
S152	S406	2.41	S152	S406	0.00
A153	A364	50.61	A153	A364	41.89
A153	N403	33.91	A153	N403	47.00
K155	T322	7.59	K155	T322	10.84
K155	A364	81.50	K155	A364	94.64
R156	W357	1.26	R156	W357	4.02
R156	W399	99.18	R156	W399	98.83
P157	W357	0.00	P157	W357	0.04
A158	R315	83.98	A158	R315	61.75
A158	W357	79.05	A158	W357	95.65
K162	W273	67.96	K162	W273	83.45
K162	T311	0.00	K162	T311	9.50
K162	R315	54.59	K162	R315	14.73
K162	E354	1.43	K162	E354	29.15
A163	W273	0.00	A163	W273	0.31
A163	Y277	70.85	A163	Y277	73.66
A166	N235	0.03	A166	N235	0.04
K167	S149	0.37	K167	S149	17.20
K167	D192	63.47	K167	D192	27.62
K167	W231	0.00	K167	W231	0.69
K168	W184	43.16	K168	W184	17.81
K168	W231	93.57	K168	W231	99.87
K169	N146	35.65	K169	N146	27.18
K170	W142	97.21	K170	W142	86.40
K170	Q181	42.04	K170	Q181	32.45
K170	W184	94.56	K170	W184	96.34
L171	S105	20.54	L171	S105	34.61
L171	R106	0.00	L171	R106	0.04
L171	E107	12.21	L171	E107	30.13
D172	W142	11.33	D172	W142	46.60

S5 Table: Radius of curvature and average angles for the crystallographic structures. PDB IDs: 1pjm, 1pjn, 3tpm, 3ukw, 3ukx, 3uky, 3ukz, 3ul1 and 3uvu.

X-ray PDB ID	Radius of curvature (Å)	Average angle (°)
1pjm	30.993009873	19.6785642989
1pjn	31.2163002693	19.6892134803
3tpm	31.2949958837	19.6517732094
3ukw	31.4596203567	19.2809651455
3ukx	31.1706523738	19.7180289396
3uky	31.2620465432	19.6859686016
3ukz	31.5329175935	19.5667871962
3ul1	30.959989878	19.8467057138
3uvu	31.33650329	19.573789442

4 Manuscrito II: Ku70

O capítulo trata do estudo estrutural envolvendo a proteína Impα complexada com a NLS da proteína Ku70. O objetivo deste estudo foi testar a hipótese da Ku70 apresentar uma NLS clássica do tipo bipartida. Para isso, geramos um modelo teórico do complexo NLS:Impα utilizando a conformação de uma NLS bipartida já conhecida como *template*. A fim de garantir uma busca conformacional rigorosa e viável para a nossa arquitetura computacional disponível, aplicamos a técnica de Dinâmica Molecular com Modos Normais Excitados (MDeNM). Apesar desta técnica possibilitar uma maior exploração conformacional, a amostragem gerada foi restrita à Impα e não permitiu uma avaliação adequada dos contatos na interface do complexo. Simulações adicionais são necessárias, no entanto, nossos dados indicam uma possível instabilidade do sistema, sugerindo que o peptídeo da Ku70 não se liga à Impα como uma NLS bipartida clássica.

Observação:

Conforme estabelecido pelo Conselho do Programa de Pós-Graduação em Ciências Biológicas (Genética), expresso na Instrução Normativa Nº01/2012, os resultados apresentados neste capítulo foram redigidos na forma de manuscrito (ainda em elaboração), que após sua conclusão, será submetido para publicação em periódico científico da área.

Conformational sampling of Ku70 modeled as a classical bipartite signal bound to Importin-α

Marcos Tadeu Geraldo[1,*], Agnes Alessandra Sekijima Takeda[1,2], David Perahia[3], Ney Lemke[1]

1 Laboratório de Bioinformática e Biofísica Computacional, Departamento de Física e Biofísica, Instituto de Biociências de Botucatu, UNESP – Universidade Estadual Paulista, Botucatu, SP, 18618-970, Brazil
2 Instituto de Biotecnologia (IBTEC), UNESP – Universidade Estadual Paulista, Botucatu, SP, 18607-440, Brazil
3 Laboratoire de Biologie et Pharmacologie Appliquée, Centre national de la recherche scientifique, École Normale Supérieure, Cachan, France

* mtgeraldo@gmail.com

4.1 Abstract

Ku70 protein is a key component participating of the DNA double-strand break repair by non-homologous end joining pathway. It forms a heterodimeric complex with Ku80 that binds to the DNA strands, allowing the access of the following proteins that directly mediates the repair. Ku70 is transported to the cell nucleus via the classical nuclear import pathway mediated by Importin-α (Impα). The first step of transport occurs by the direct binding of Impα to the classical nuclear localization sequence of Ku70 (Ku70NLS). The X-ray crystallographic structure shows that only the C-terminal residues of Ku70NLS interact with Impα, characterizing it as a monopartite NLS. However, mutagenesis data show that N-terminal residues are also essential for nuclear import, which indicates that the cNLS could be bipartite. Therefore, the aim of our work is to investigate the interactions of Impα-Ku70NLS complex combining standard molecular dynamics simulations (MD) and normal mode (NM) analysis. For this purpose, we used the recently developed methodology of MDeNM (Molecular Dynamics with excited Normal Modes) in which the directions of low-frequency NMs are privileged during a short MD. We used in this study the structure of Ku70NLS modeled as a bipartite cNLS and complexed with Impα. Our results indicate a possible instability of the complex, suggesting that Ku70NLS would not bind to Impα as a classical bipartite NLS.

4.2 Introduction

The DNA repair systems are critical for maintaining the integrity of the genetic material. Exchanges or error-base pairing or even breakage of the DNA strand can occur during the process in which DNA is subjected in the cell or during exposure to certain environmental conditions, including chemical agents. The mechanisms of DNA repair are: i) base excision repair (BER), ii) nucleotide excision repair (NER), iii) error-base repair (MMR) (KNUDSEN et al., 2009) and iv) repair by direct reversal of damage or double-strand break (DSB), separated into two major pathways, homologous recombination (HR) and non-homologous end joining (NHEJ) (CHRISTMANN et al., 2003).

The repair by non-homologous end joining (NHEJ) acts on double-stranded breaks of DNA. These double-stranded breaks can induce highly potential harmful effects on a larger scale, such as exchanges and chromosome breakage, and consequent cell death (PFEIFFER; GOEDECKE; OBE, 2000). From all the proteins involved in this repair system, the Ku70 (synonym: XRCC6) and Ku80 proteins (synonym: XRCC5) play a key role in DNA repair, due to their ability to compose a heterodimer and to bind DNA double-stranded break in order to facilitate its repair by non-homologous end joining by extremities adhesion (OCHI et al., 2010). The crystallographic structure of the complex was elucidated with Ku70-80 heterodimer isolated or associated with DNA (WALKER; CORPINA; GOLDBERG, 2001) forming an asymmetric ring that involves the double-stranded DNA and allows the association of other proteins involved in the repair process. Both proteins have nuclear localization sequences (NLSs) (KOIKE et al., 1999a; KOIKE et al., 1999b), thus allowing its translocation to the nucleus independently via the classical import pathway mediated by Importin-α (Impα) + Importin-β (Impβ) complex (KOIKE; SHIOMI; KOIKE, 2001).

The classical NLSs (cNLSs) contain one or two positively charged amino acid groups and are classified as monopartite cNLS - containing a single group of basic residues [(K/R)KX(R/K)]; and bipartite cNLS - containing two groups of basic residues separated by 10-12 residues variants [KRX$_{10\text{-}12}$KRXK] (which X represents any residue) (FONTES et al., 2003). In both cases, the cNLSs are recognized by Impα, the nuclear import receptor responsible for recognition of NLS to transport the proteins by the classical nuclear import pathway (STEWART, 2007). The recognition of Impα occurs by two binding sites, so-called major and minor site. These sites have important positions for NLS binding: the P2-P5 positions from major site and P1'-P2' from minor site. Classical monopartite NLSs bind preferentially into the major site (FONTES; TEH; KOBE, 2000), whereas classical bipartite NLSs bind into both sites (FONTES et al., 2003). These bipartite NLSs can also establish contacts with Impα by the 10-12 region, denominated as linker, that has been reported as important for the recognition of bipartite NLSs (MARFORI et al., 2012; BARROS et al., 2012; CHRISTIE et al., 2015).

Although many cNLSs have already been identified, the understanding of the their recognition remains limited (MARFORI et al., 2011). In the case of DNA repair proteins, Ku70NLS was initially considered as bipartite (KOIKE et al., 1999a). However, a recent work of Takeda et al. (2011) solved by X-ray crystallography the C-terminal structure of Ku70NLS peptide, suggesting that this peptide binds as a monopartite NLS because there was not electron density for its N-terminal region. One important aspect is the presence of positively charged residues in this region, indicating that these residues could bind to Impα. Moreover, a previous study of Koike et al. (1999b) demonstrated by fluorescence assays, using different constructs of Ku70NLS fusioned to an enhanced green fluorescent protein (EGFP), that the N-terminal region of the peptide is critical for the nuclear distribution, concluding that deletions at the N-terminal region impairs the NLS recognition by Impα.

In order to understand the interaction of Ku70NLS with Impα, our goal is to answer the following questions: (1) Can Ku70 interact with Impα as a bipartite model of NLS? (2) Are there non-classical interactions that promote the recognition of the complex?

To answer these questions we generated a bipartite model of Ku70NLS and then simulated this system using two approaches: standard molecular dynamics simulations (MD) and molecular dynamics with excited normal modes (MDeNM). In MDeNM, the vectors of previously selected normal modes are randomly combined and kinetically activated in order to extend the conformational space (COSTA et al., 2015). The advantages of this technique are: (1) different structural conformations with large sample are generated, thereby providing a support for the desired statistical analyses, and (2) calculation time is much shorter than conventional MD.

4.3 Methods

4.3.1 In silico modeling: Ku70NLS complexed with Impα

To model the Ku70 peptide (NLS: ^{537}EGKVTKRKHDNEGSGSKRPKVE^{558}) bound to the major and minor sites of Impα, we used as template the flap endonuclease 1 (FEN-1) NLS complexed with Impα (PDB ID: 3UVU). We used the Modeller v.9.10 program to perform the homology modeling, and based on a stereochemical analysis in the MolProbity program (CHEN et al., 2009) and the Ku70NLS occupation on the major and minor sites (Fig 1), we chose the best-model to be used as the input for all the molecular simulations.

4.3.2 Search of similar structures to Impα

We used Prody (BAKAN; MEIRELES; BAHAR, 2011) to perform a basic local alignment search over the GenBank database (www.ncbi.nlm.nih.gov/genbank/) to retrieve

sequences with high similarity to Impα that have their corresponding tertiary structure deposited on the PDB database (www.rcsb.org/). We gathered 36 X-ray structures with more than 95% sequence identity to Impα. Then, using CHARMM program v.36b1 (BROOKS et al., 2009), we performed the Principal Component Analysis (PCA) for the X-ray structures and calculated their vectoral overlap with the Impα normal modes (NMs). This analysis allowed the validation of the first NMs used in the MDeMD method.

4.3.3 MDeNM method procedure

The protonation state of the system's residues was verified in the PROPKA webserver (propka.org/) and no protonation change was required. Then, the topology and parameter files were generated with the CHARMM-GUI server (www.charmm-gui.org). The simulations were carried out in explicit solvent with periodic boundary conditions and the PME algorithm to treat electrostatic interactions. The simulation box was composed of 58,992 TIP3 water molecules (ensuring a minimum of 10 ångströms (Å) between the protein and the box's edges) and 59 Na^+ and 54 Cl^- counter ions to neutralize the system's charges, giving an ion concentration of 0.15 M. A minimization procedure was applied with harmonical force constraints for the backbone and side-chain atoms, using the Steepest Descent (SD) and Adopted Basis Newton-Raphson (ABNR) algorithms. Afterwards, the system was equilibrated, keeping the harmonical constraints, in a NVT ensemble for 100 ps with the Velocity Verlet integrator and the Nosé-Hoover method to keep the temperature constant to 300 kelvin (K). A 1 fs integration time step was used exclusively for the equilibration procedure to avoid possible instabilities during the simulation and consequent abnormal terminations. Finally, the harmonical constraints were removed and a 1 nanosecond (ns) MD was conducted with the leap-frog integrator with 2 fs integration time step. The pressure and temperature were kept constant (NPT ensemble) with the Nosé-Hoover thermostat algorithm. After this short MD, the atomic coordinates and velocities were stored to be subsequently used in the NM excitation procedure.

The final structure from the 1 ns MD was submitted to energy minimization using the conjugate-gradient (CG) method followed by the ABNR algorithm. Harmonic constraints were applied during the CG steps, being progressively decreased from 250 to 0 $kcal.mol^{-1}$Å^{-2}, then it was applied the ABNR algorithm with no constraints. NMs were computed with the VIBRAN module of CHARMM. The carbon-α (Cα) atomic fluctuations (based on root mean square fluctuations – RMSF) from each NM were calculated to select the highly flexible ones to be excited. An incremental temperature (T_i) was added to the current temperature for activating the vectors from the chosen NMs, according to Costa et al. (2015). Six different T_i (1, 3, 5, 7, 9 and 11K) were used and the results from a 300 replica were compared relative to the conformational exploration of the

NMs space. An appropriate T_i allows a large conformational search but keeps the system in a relaxed way. Finally, the best T_i to be used in the excitation procedure was chosen, and 2,000 replica were generated in this condition. Each replica is a distinct linear combination of the chosen NMs and three excitations of 2 ps were performed for each one.

All conformations generated during the excitation procedure were clustered, based on a 3 Å cut-off for backbone, using the *g_cluster* module of the GROMACS software to obtain representative backbone conformations. These selected conformations were submitted to 200 ps of standard MD simulations to remove any remaining excitation effects.

In addition, we performed a long conventional MD simulation (100 ns), applying the similar procedures of minimization and equilibration described above to set up the system, to compare with our results from the MDeNM method.

4.3.4 Data analysis

Projection onto the chosen NMs

The structures generated from the MDeNM procedure were projected onto the chosen NMs using the VMOD facility of the CHARMM program. This analysis is critical to ensure a well coverage of the conformational space.

RMSD and RMSF calculations

The global root-mean-square deviation (RMSD) for backbone atoms and the root-mean-square fluctuation (RMSF) for Cα atoms were calculated in the Wordom program (SEEBER et al., 2007) and the results were plotted using the R program (R Core Team, 2015).

Geometric analysis of Impα

The collective motions of Impα were described in terms of geometrical measurements, according to Geraldo et al. (2016). The radius of curvature from the plane formed by three vertices that corresponded to the plane of the observed bending motion, and the angle between α-helices 3 (H3) from neighboring ARM repeats were calculated to characterize the bending and twisting motions, respectively. In addition, we used the data from Geraldo et al. (2016) for Impα complexed with Nucleoplasmin-NLS (Impα-NplNLS) – a classical bipartite NLS – to compare with our geometrical data from the Impα-Ku70NLS model.

Maps of cross-correlations

The initial comparison of residue-residue contacts from the standard MD and MDeNM results was initially performed generating maps of cross-correlation to evaluate the associated movements between Ku70NLS and Impα. The cross-correlation calculations were performed in the Wordom software (SEEBER et al., 2007).

Evaluation of peptide-protein interactions

The trajectories from standard MD and short MDs after excitation were used to evaluate the contacts in the peptide-protein interface based on the occurrence of specific types of interactions. The presence of salt bridges and hydrogen bonds were determined in the VMD program. The criteria for considering these interactions was: donor-acceptor distance was $\leqslant$ 3.5 Å, and donor-hydrogen-acceptor angle deviation was $\leqslant$ 60 degrees for hydrogen bonds. In contrast, hydrophobic contacts were determined based on the LIGPLOT program, using the Ku70NLS-Impα model. Afterwards, based on this list of hydrophobic contacts, the distances of the closest carbon atom from the hydrophobic side-chains of each residue-pairs were determined from the MD trajectories to calculate the occupancies of each hydrophobic contact. The distance criteria adopted was $\leqslant$ 4 Å.

4.4 Results

4.4.1 Geometric analysis of Impα and selection of NMs for MDeNM

A quality vector analysis comparing the first three lowest-frequency modes revealed a bending and twisting motion for NM7 and NM8, respectively (Fig 2). The bending resembled a gateway motion-pattern along the concave surface of Impα, which is the region composed of conserved amino acids responsible for the NLS binding. The twisting from NM8 was closely related to changes in the torsion amplitude between adjacent ARM repeats along the entire protein. Finally, the motion from NM 9 could be simplified as a segregation of two main motions: a lateral bending pattern and a twisting (Fig 2).

The quantitative analysis determined the radius of curvature in opened and closed configurations of Impα and showed different amplitudes for the bending motion, in which Impα-Ku70NLS had movements with lower amplitudes compared to Impα-NplNLS (Fig 3A). Moreover, the average values for the angles between helices in the twisting motion were higher for Impα-NplNLS (Fig 3B). Finally, the specific analysis of the angles between helices showed a general reduction in the amplitude for most ARM pairs from the motions described as lateral-bending/twisting (NM9) and twisting (NM8) for Impα-Ku70NLS (Fig 4). The only exception were ARMs 6-7 that showed an increased amplitude in the twist.

These characterized motions could be also observed on PCA using the crystallographic structures of Impα bound to different types of NLSs. A total of 36 similar structures to Impα (over 95% of sequence identity) were retrieved based on the basic local alignment search over the GenBank database. Comparing NMs and PCs, we observed a large vector overlap between PC1-NM7, PC2-NM8 and PC3-NM9 (S1 Table). This result indicated the importance of selecting these NMs for the excitation procedure. The residue-fluctuation analysis over each NM (S1 Fig) indicated that most conformational changes occurred approximately until modes 12-13, therefore, these additional NMs were also selected to be kinetically excited.

4.4.2 NMs activation

A crucial step for the MDeNM method is the choice of an excitation temperature because this is directly related to the conformational exploration of the system. We first tested our system over different sets of temperature, always checking for the protein structural integrity. Using a 300 excitation replica (only a single degree of freedom is activated in each replica) for each temperature, we observed a significant increase in the range of conformational variation that was correlated to the increase in temperature (S2 Fig). Costa et al. (2015) observed that incremental temperatures above 50 K led to destabilization of the system. Therefore, we set 11 K as the top temperature and generated additional replica to reach different points in the NM space. After 2,000 replica we clearly observed a well-coverage in the bi-dimensional projection onto each pair of NMs, and to illustrate this result, the S3 Fig shows all the projections onto NM7.

Although the system reached different conformational states, most of them were close to the starting structure. Therefore, a selection for representative structures was performed for a subsequent run of short MD simulations. According to Costa et al. (2015), the MD run is important to avoid considering artifacts (e.g. some unstable/unfavorable conformation) that could be possibly generated during excitation. The clusterization procedure selected 343 representative conformations that showed an uniformed distribution in the bi-dimensional projection onto each pair of NMs (S4 Fig). These conformations were used for running 200 ps of MD simulations and all subsequent analyses were performed using the ensemble of all generated short-trajectories. The bi-dimensional projection onto each pair of NMs showed that the conformational space was well-explored (S5 Fig).

An initial comparison between these ensemble of trajectories and standard MD was performed using RMSD and RMSF calculations. The global RMSD values indicated that MDeNM had greater global conformational exploration compared to standard MD (S6 Fig). However, considering the Cα fluctuations most of the conformational changes observed in MDeNM occurred with Impα, whereas Ku70NLS was more stable and, in

contrast, the MD results showed significant fluctuations for most of the peptide, particularly around residue $_{\text{Ku70NLS}}S^{550}$ (S7 Fig).

4.4.3 Contacts in major and minor sites

A general view of possible contacts in Impα-Ku70NLS was determined by cross-correlations between amino acids. Standard MD showed well-bounded areas of positive correlations only in the major site (Fig 5). In contrast, MDeNM showed a large number of positive correlations along the entire interface of Impα-Ku70NLS.

Despite the great divergence of correlations, the interaction analysis revealed an overall similarity between the MD and MDeNM methods, including the minor (542-KRK-544) and major (553-KRPK-556) sites; furthermore, the number of contacts decreased in both methods compared to our initial structure (Fig 6). The conserved residues at P1' and P2' from the minor site were observed in MDeNM, such as $_{\text{Imp}\alpha}D^{325}$, $_{\text{Imp}\alpha}E^{396}$ and $_{\text{Imp}\alpha}W^{399}$. Standard MD also showed these contacts, except $_{\text{Imp}\alpha}D^{325}$ at P1'. The critical residue $_{\text{Imp}\alpha}D^{192}$ at P2 from the major site was determined in both methods, including important hydrophobic interactions such as $_{\text{Imp}\alpha}W^{231}$, $_{\text{Imp}\alpha}W^{184}$ and $_{\text{Imp}\alpha}W^{142}$. In contrast, sparse number of contacts were observed in the linker region (residues 545 to 552) that was restricted to residue $_{\text{Imp}\alpha}R^{315}$, which established a salt bridge with $_{\text{Ku70NLS}}D^{546}$.

4.5 Discussion

4.5.1 Ku70NLS may not bind as a classical bipartite NLS

The DNA repair machinery is highly important to maintain the integrity of the genetic material. Due to its active nuclear activities via participating in the repair of double-stranded breaks in DNA, the Ku70 began to be structurally studied in order to understand how its transport to the nucleus occurs. A mutagenesis study identified some of the key residues involved in the recognition by Impα (KOIKE et al., 1999b). From these data, the residues of the N-terminal region proved to be fundamental for nuclear localization. Later, structural data from X-ray crystallography partially solved the structure of Ku70NLS, suggesting a monopartite model of cNLS binding to Impα. In this scenario, our simulations showed residues in the N-terminal region making contacts in the minor site of Impα, especially by the MDeNM technique, represented by electrostatic interactions occurring at positions P1' and P2'. Moreover, the main contacts that have been reported in the literature for classical bipartite NLSs, such as FEN-1 (BARROS et al., 2012) and Nucleoplasmin (MARFORI et al., 2012; FONTES; TEH; KOBE, 2000) were also observed in the results from both methods. In particular, we observed the critical contact at P2 in the major site of Impα, establishing a strong electrostatic interaction involving $_{\text{Imp}\alpha}D^{192}$ and $_{\text{Ku70NLS}}K^{553}$. Contacts at positions P3–P5 were also

observed, highlighting the importance of the highly conserved hydrophobic contacts that were already observed in other cNLSs such as SV40 (FONTES; TEH; KOBE, 2000), Nucleoplasmin (MARFORI et al., 2012; FONTES; TEH; KOBE, 2000), retinoblastoma (FONTES et al., 2003) and FEN-1 (BARROS et al., 2012).

In contrast, the cNLS linker exhibited few contacts with Impα. Although the participation of the linker in the Impα recognition seems to vary between cNLSs, studies have shown the recurring participation of some residues such as $_{\text{Imp}\alpha}\text{R}^{315}$, $_{\text{Imp}\alpha}\text{R}^{238}$ and $_{\text{Imp}\alpha}\text{Y}^{277}$ in the recognition of Impα (FONTES; TEH; KOBE, 2000; FONTES et al., 2003; CONTI; KURIYAN, 2000; CHEN et al., 2005; CUTRESS et al., 2008; GIESECKE; STEWART, 2010; YANG et al., 2010; BARROS et al., 2012). Furthermore, the flanking residues from both sites seem to contribute to the cNLS:Impα interaction. These evidences indicate that our Ku70NLS model may not act as a classical bipartite NLS. In addition, most cNLSs span major and minor sites by 10-12 residues, however, in our model of Ku70NLS bound to Impα, the linker region would spans only by 9 residues. Studies show that longer linkers are tolerated (KOSUGI et al., 2009; LANGE et al., 2010), but Ku70 would be the first reported case of a smaller linker. Although we still observed interactions with high occupancies occurring in both sites of Impα, a smaller linker size probably did not stabilize the NLS into the Impα binding grooves, and as a consequence, we observed weak correlations between the residues on the interface of the complex from the MD simulation.

4.5.2 The conformational changes from MDeNM were limited to Impα

The first three NMs exhibited clear patterns of collective motion in accordance to the in silico data from another cNLS – Nucleoplasmin (GERALDO et al., 2016) – characterized by bending and twisting motions. The bending is defined as the opening and closing of Impα in the access region to the cNLS binding site, and the twisting is the torsional motion along the entire axis of Impα following its own curvature, and these patterns of motion may be directly related to the access of the NLS and its local adjustments into the binding grooves of Impα, respectively. The geometrical analysis showed lower amplitude values for the bending and twisting motions of Impα-Ku70NLS, indicating that Impα is trying to adapt its conformation and motions to the reduced size of the linker. In the specific analysis of angles between pairs of ARM repeats revealed an overall smaller amplitude for most pairs compared to the classical bipartite NLS from Nucleoplasmin. In fact, the patterns of amplitude were similar to the data from Geraldo et al. (2016) for the N-terminally truncated Impα in the absence of an NLS peptide (Apo Impα), including the higher oscillation between ARMs 6-7. Apo Impα has been revealed to be more flexible and unstable in relation to the bound Impα based on experimental and in silico studies (TAKEDA et al., 2013; FALCES et al., 2010; GERALDO et al., 2016); therefore, our data points out to a possible instability of the Impα-Ku70NLS complex.

The combined usage of NM and MD methods has been shown as a powerful way for improving the sampling of the conformational space without requiring a very high computational cost compared to traditional simulation techniques (COSTA et al., 2015; LAINE et al., 2011). The recent development of the MDeNM method matched these conditions in a comparison with a long standard MD, using hen egg lysozyme and HIV-1 protease as the studied examples (COSTA et al., 2015). In our results, the activated NMs in the MDeNM method were able to increase the conformational space compared to classical dynamics, based on the analysis of global backbone RMSD. Despite this general result favoring MDeNM, most Cα fluctuations were restricted to Impα, and almost no effects were observed in Ku70NLS. In fact, Ku70NLS fluctuated considerably in standard MD, reflecting a possible instability of the peptide in the modeled structural conditions used in our study. However, this result may also indicate that the simulation time used was not sufficient for the accommodation of Ku70NLS into the binding grooves of Impα. Nevertheless, we still would expect some fluctuations from the MDeNM method. Apparently, the vectors used during NM activation were not suitable to induce a separation of Ku70NLS and Impα for an appropriate evaluation of the complex stabilization upon binding. This may also reflect the results of cross-correlations observed, exhibiting very distinct patterns of contacts. A possible solution would be applying artificial NMs in opposite directions of the peptide-protein complex to induce its separation, thus promoting more conformational changes along the interface.

In conclusion, despite the simulation techniques applied in this study, the conformational sampling generated was restricted to Impα and did not allow the appropriate evaluation of the contacts in the Impα-Ku70NLS interface, especially for the MDeNM method. Additional simulations are required, however, our data indicates a possible instability of the system, suggesting that Ku70NLS could not bind to Impα as a classical bipartite NLS.

4.6 Acknowledgments

This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [grant numbers 2012/19447-2; 2014/21976-9] and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [grant number 142110/2012-4]. We are grateful to Prof Dr. Cesar Martins for providing computational access for running part of the simulations.

4.7 Figures

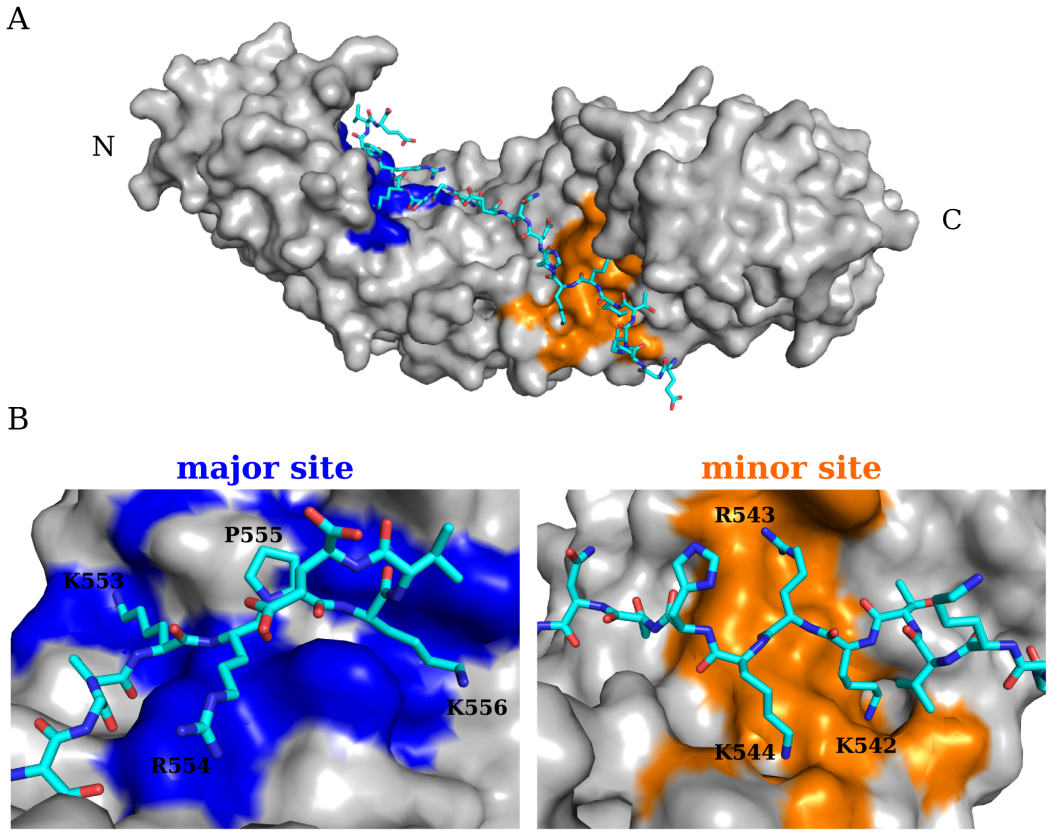


Fig 1 – The starting structure of Ku70NLS complexed with Impα for MD and MDeNM simulations. (A) The surface representation of Impα (gray) complexed with Ku70NLS (cyan stick representation). The Impα major (blue) and minor (orange) binding sites to cNLS are indicated. (B) The major and minor sites zoom indicating the residues with positively charged side chains positioned in the main pockets of the Impα binding core.

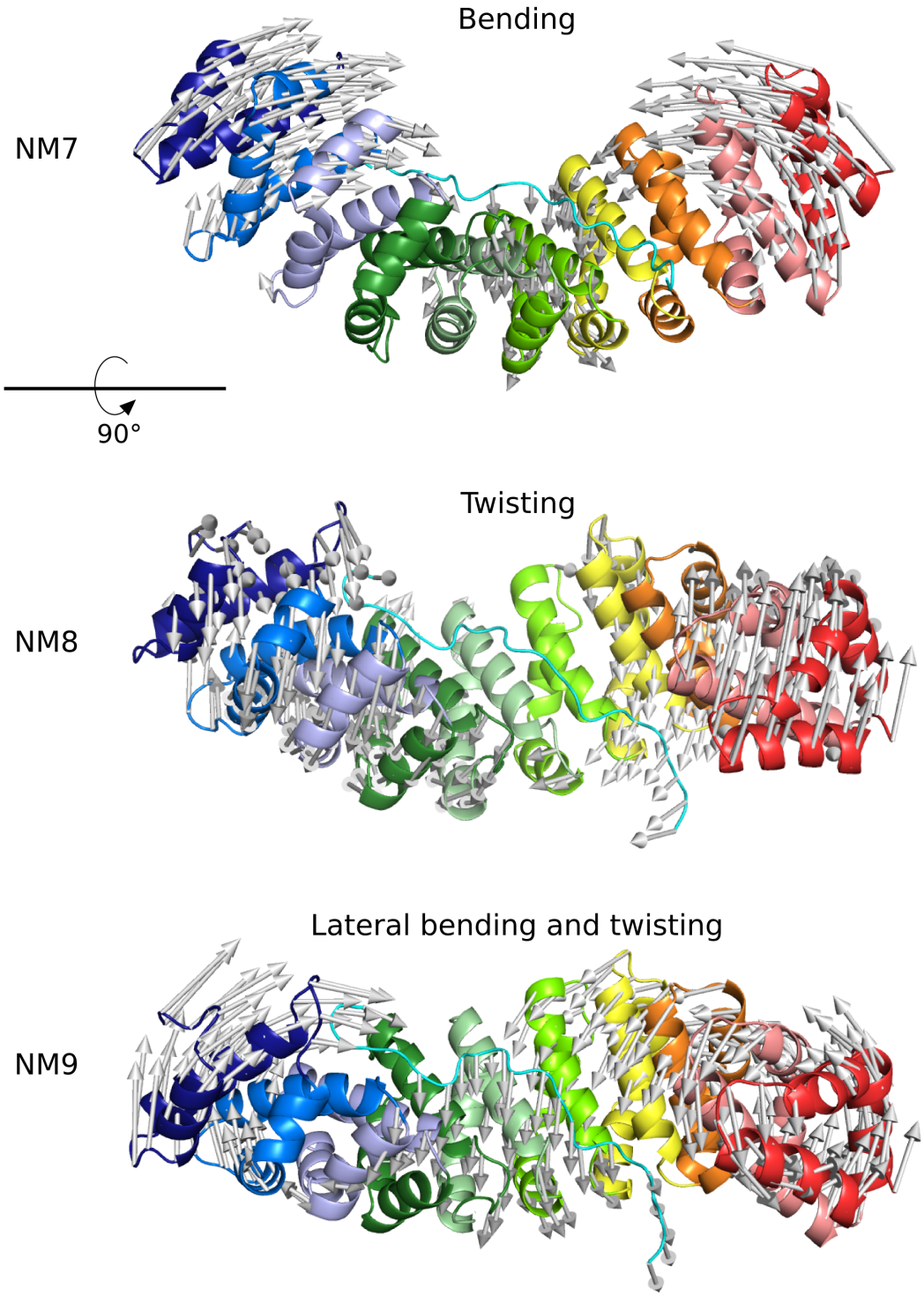


Fig 2 – Vectors from NMs 7-9. The cartoon representation of Impα colored based on each ARM repeat as a rainbow spectrum from N-terminus (blue) to C-terminus (red), and Ku70NLS (cyan cartoon model) positioned in an anti-parallel configuration compared to Impα. The vector arrows (gray) for NMs 7–9 are shown along with the motion described by these modes. NM7 is shown in a front view, whereas NM8 and NM9 are shown in an upper view (90° rotation in the X-axis).

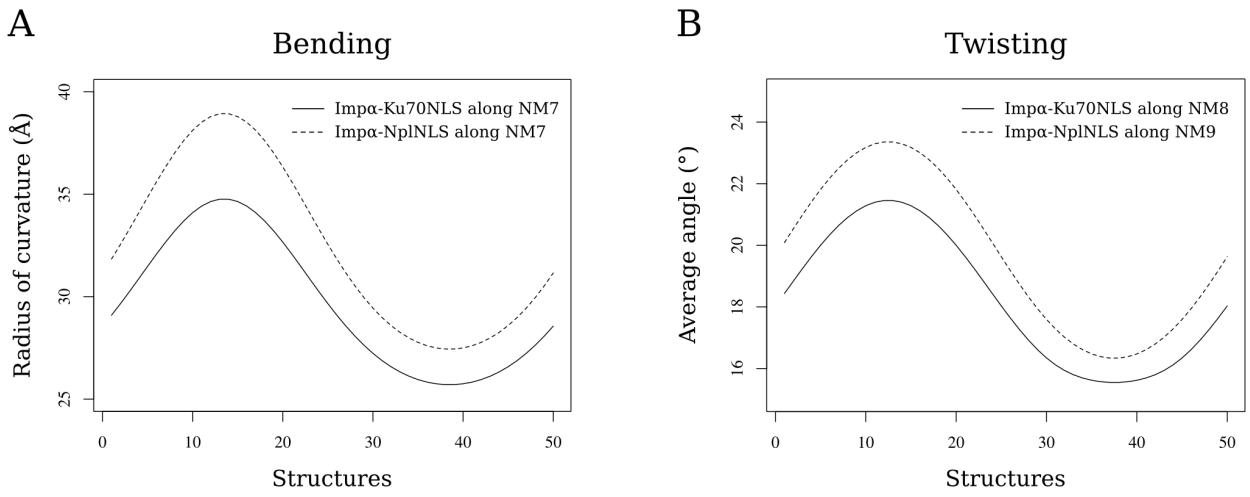


Fig 3 – Geometric analysis of the bending and twisting motions of Impα. (A) The bending motion was quantitatively characterized by the radius of curvature along NM7 for Impα-Ku70NLS (solid line) and Impα-NplNLS (dashed line), whereas (B) the twisting motion was quantitatively characterized by the average values for the angles between helices for Impα-Ku70NLS (solid line) and Impα-NplNLS (dashed line) along NM8 and NM9, respectively.

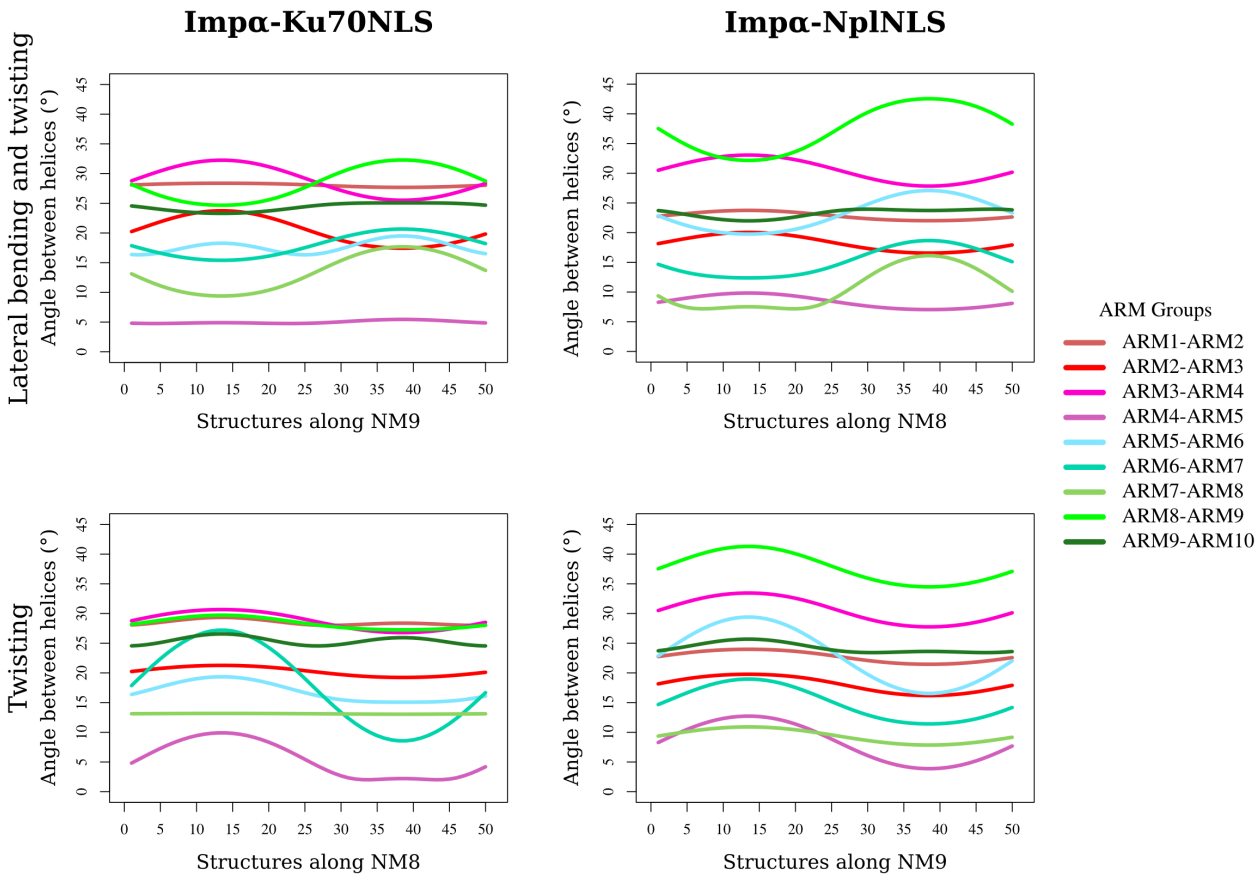


Fig 4 – Angle between helices of Impα. The angles between neighboring H3 pairs from the motions described as lateral-bending/twisting and twisting for Impα–Ku70NLS and Impα–NplNLS. The ARM groups considered for each angle calculation are depicted with different color assignments.

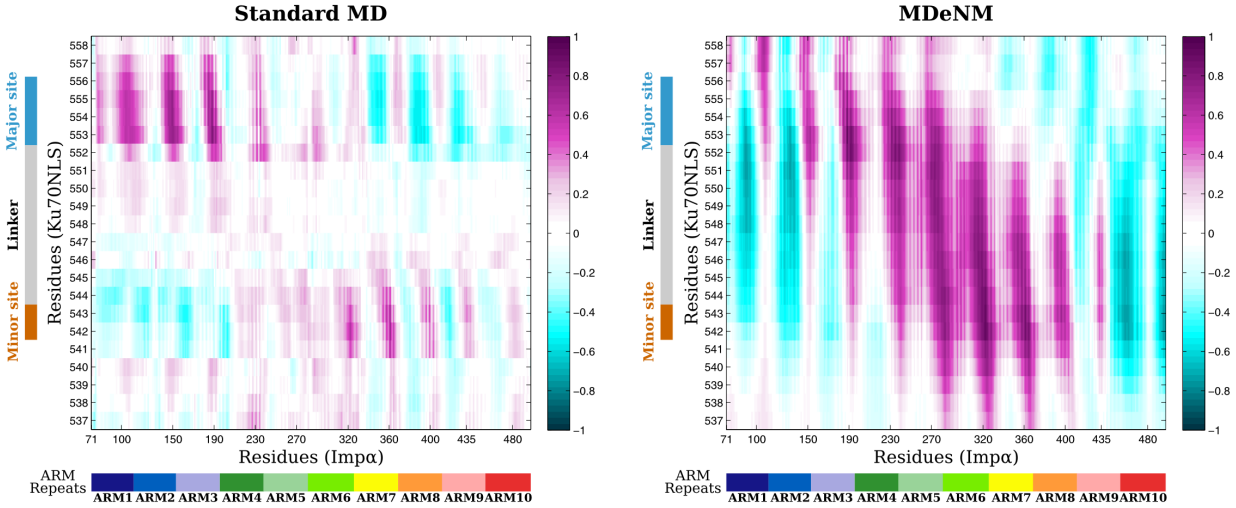


Fig 5 – Heatmap of cross-correlations between Impα and Ku70NLS. (A) Trajectories from standard MD (100 ns) and (B) short-MD simulations from MDeNM method were used for the calculation of correlations. A color bar indicates the degree of correlation from anti-correlated (negative values) to correlated (positive values) residues. The X and Y axes, respectively, show the position of each ARM repeat in relation to the Impα sequence and the Ku70NLS residues in contact with the protein binding sites.

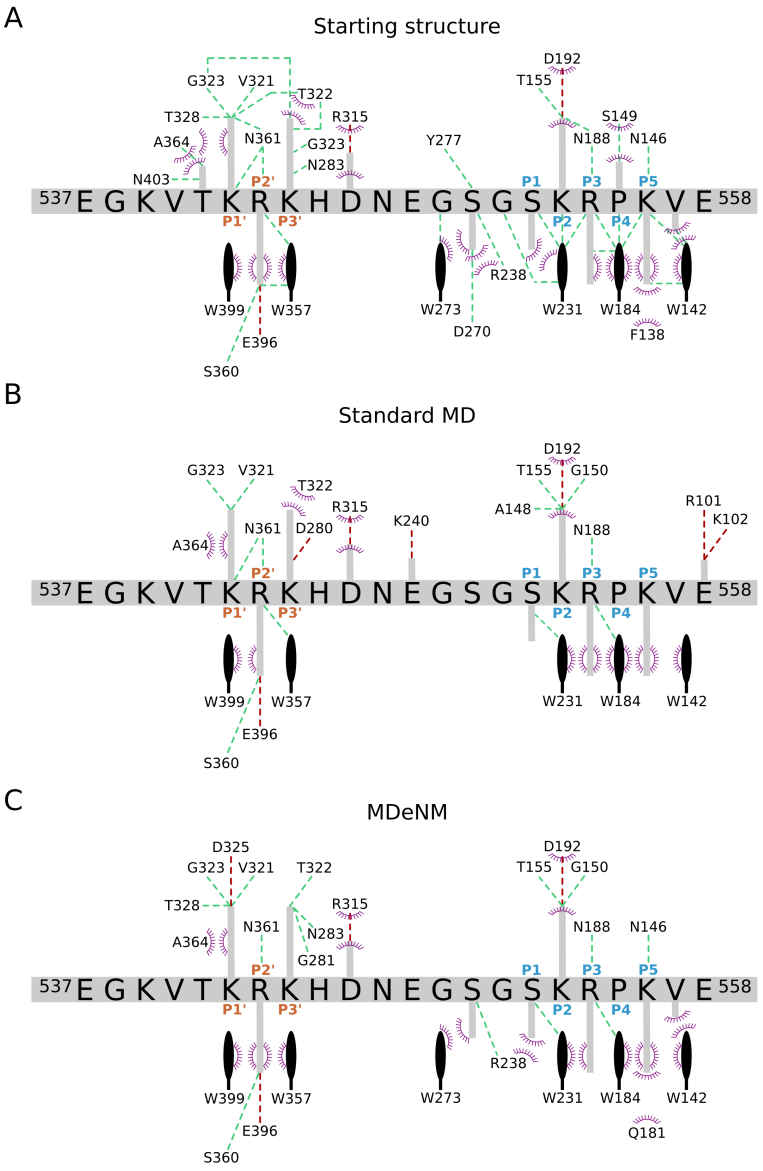
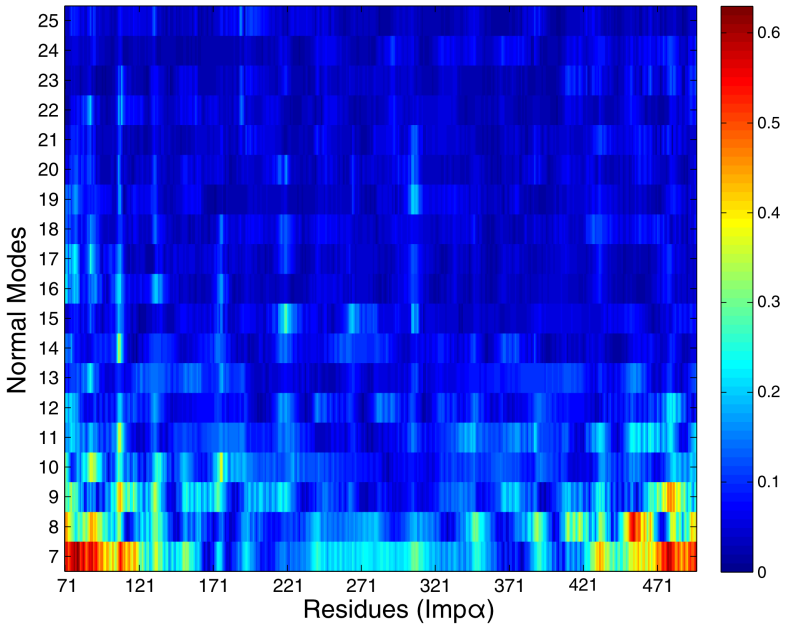


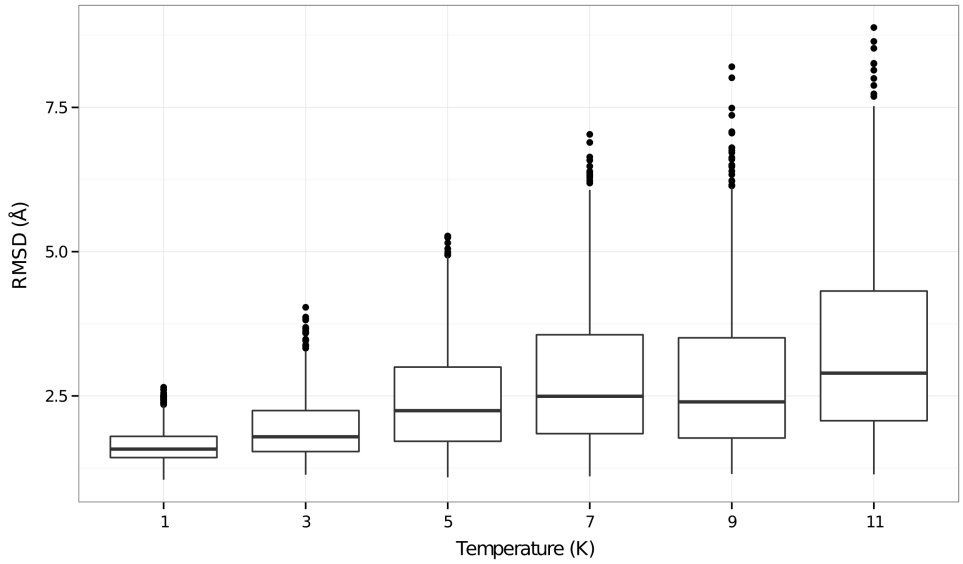
Fig 6 – Scheme of the interactions observed in the Ku70 and Impα interface. The main chain of Ku70NLS is represented as a gray horizontal line with its corresponding amino acid sequence along with side chains shown as perpendicular lines. The dashed lines stands for hydrogen bonds (green) and salt bridges (red). The hydrophobic contacts are indicated by arcs with radiating spokes (purple). The important tryptophan residues that mediate the hydrophobic contacts are depicted in the scheme as black sticks. The main positions from major (P2-P5) and minor (P1'-P2') sites are indicated in blue and orange, respectively. The interactions were analyzed in three different cases: (A) initial modeled conformation for Ku70 and Impα complex; (B) standard MD; (C) short MD simulations using the representative conformations after clustering the 2,000 replica of MDeNM excitation. Only interactions that had an occupancy rate $\geq$50% of the analyzed trajectories are indicated in the scheme.

4.8 Supporting Information

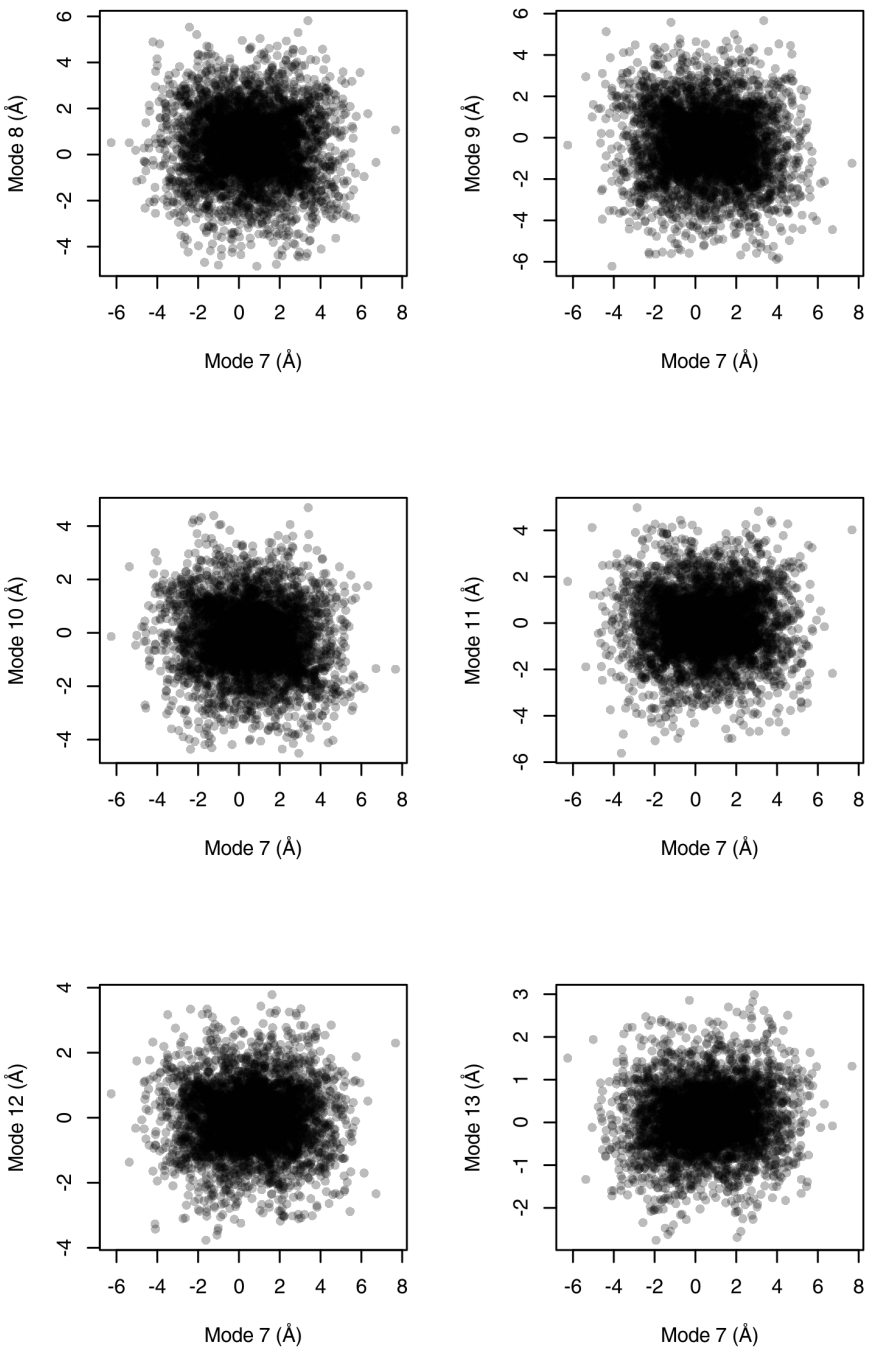
Figures



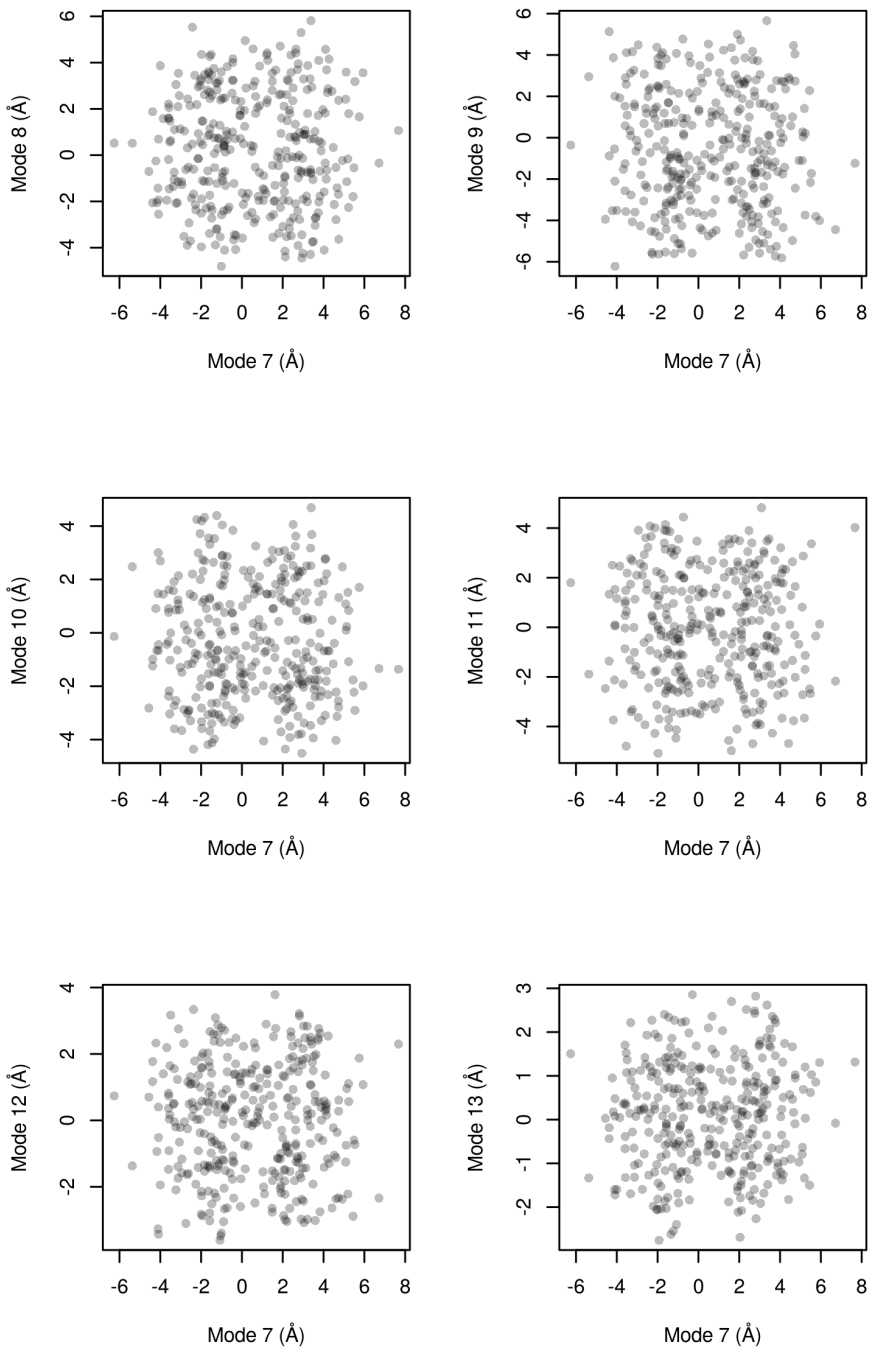
S1 Fig: Heatmap of Cα fluctuations for Impα. The Cα fluctuations are from NMs 7 to 25.



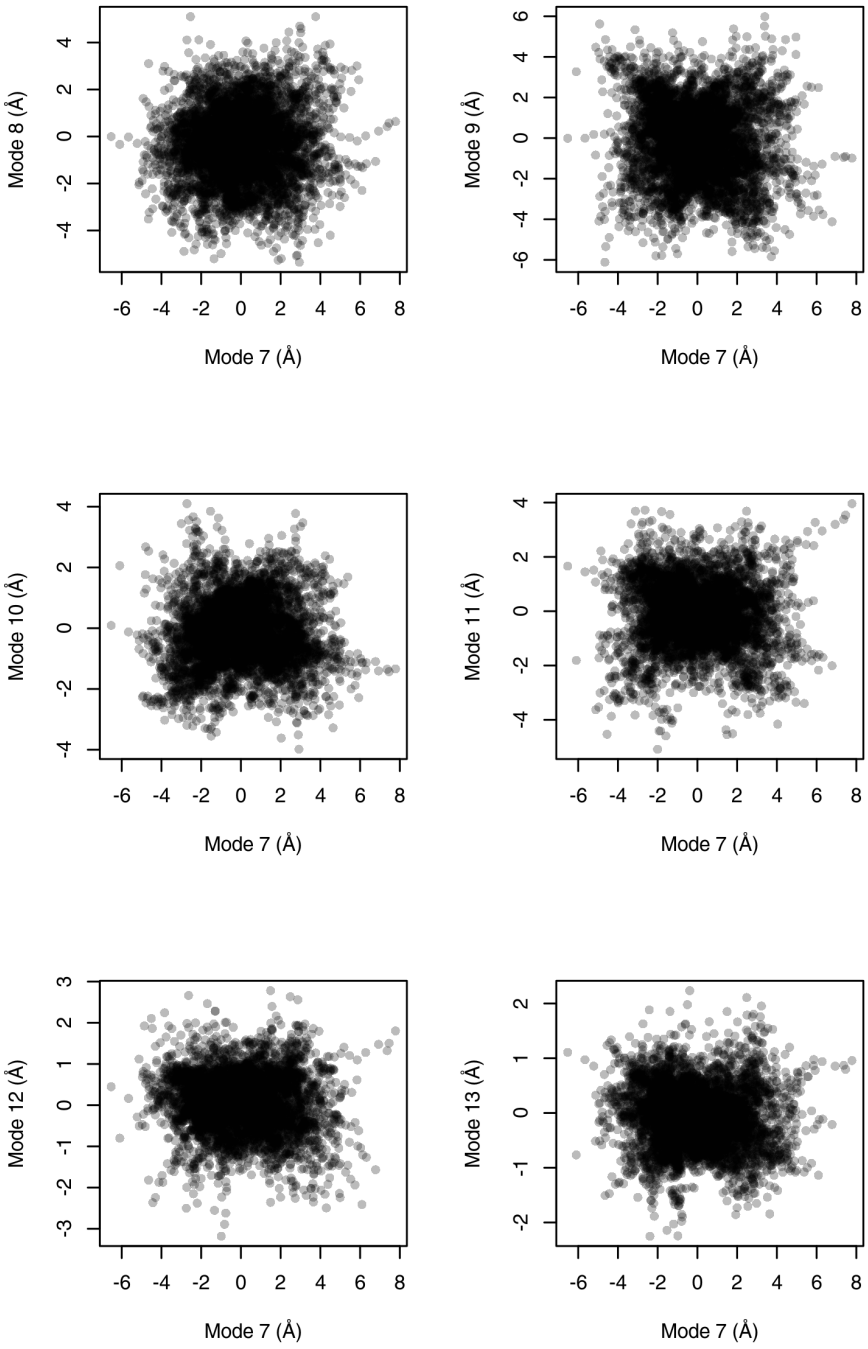
S2 Fig: RMSD distribution from different T_i temperatures. The box-plots are a representation of the structure distribution from the 300 replica of MDeNM excitation from 1 to 11 K.



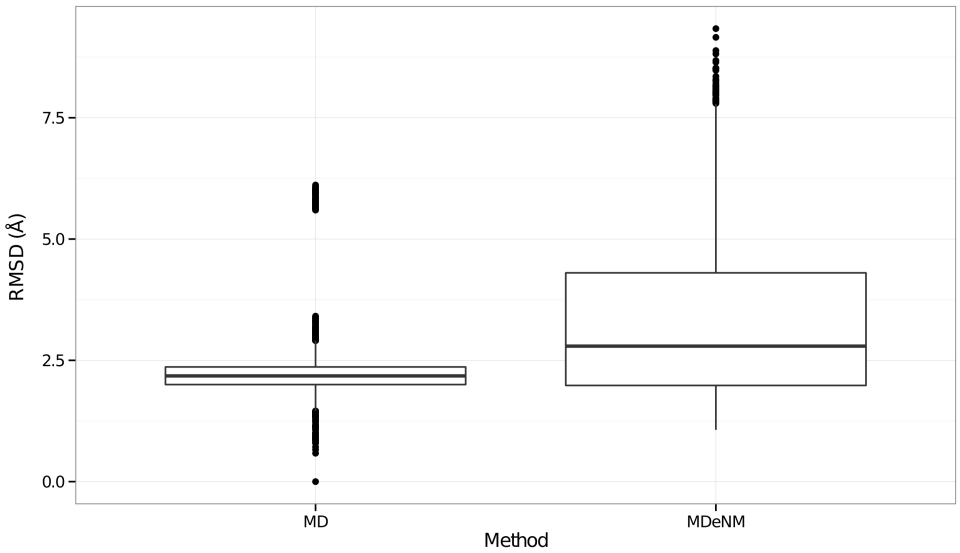
S3 Fig: Bi-dimensional scatter-plot of NM7 *versus* all activating modes. The projected coordinates were obtained from 2,000 replica of MDeNM excitation.



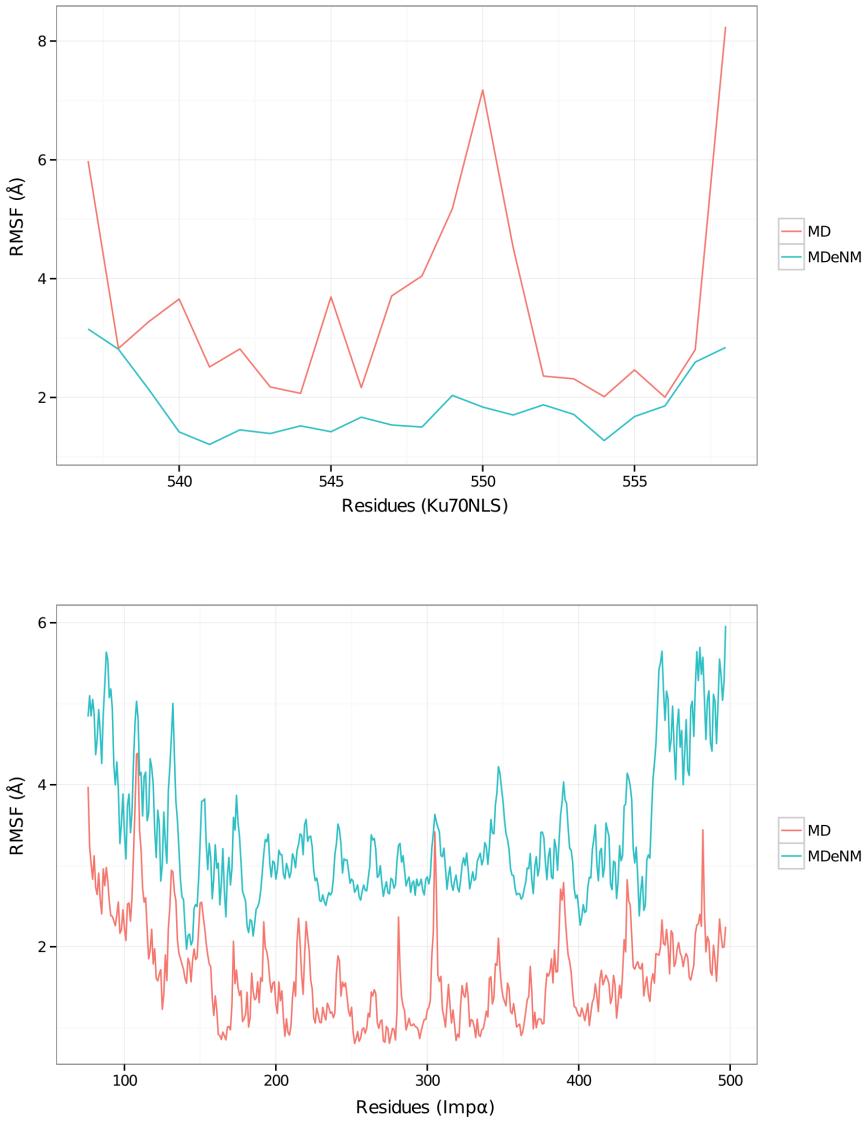
S4 Fig: Bi-dimensional scatter-plot of 343 representative conformations. The 343 representative conformations obtained from clustering the 2,000 replica of MDeNM excitation were projected onto NM7 *versus* all activating modes.



S5 Fig: Bi-dimensional scatter-plot from the short-MD simulations. The bi-dimensional scatter-plot of NM7 *versus* all activating modes.



S6 Fig: RMSD distribution from MD simulations. Box-plot from 100 ns of standard MD (*left*) and 2,000 replica of MDeNM excitation (*right*).



S7 Fig: RMSF plot from MD and MDeNM simulations. Cα fluctuations from the trajectories of standard MD (red line) and short-MD simulations (blue line). The fluctuations were analyzed separately for Ku70NLS (*up*) and Impα residues (*down*).

Tables

S1 Table: Percentage of overlap between vectors from PC and NM analysis. Higher values are highlighted in gray.

	Normal modes						
PC (X-rays)	NM7	NM8	NM9	NM10	NM11	NM12	NM13
PC1	70.304	0.1333	14.5188	0.0312	9.1886	0.0324	0.0011
PC2	5.0683	50.6864	6.5066	0.9943	9.8509	3.4208	3.6892
PC3	12.6236	16.5117	39.6636	0.5374	3.9670	0.6133	0.2657
PC4	0.9410	0.1060	0.0691	3.0426	10.6553	0.5339	0.2960
PC5	0.0005	2.0490	0.6768	2.5437	1.1145	0.0183	0.0114

5 Considerações finais

A abordagem utilizada nesta tese combinando Dinâmica Molecular e Análise de Modos Normais se mostrou uma boa estratégia para compreender tanto a dinâmica como as interações do nosso sistema estudado.

As análises do complexo Nucleoplasmina-Impα nos mostrou os possíveis movimentos apresentados durante o reconhecimento do peptídeo NLS. Apesar do nosso estudo enfatizar NLSs clássicas bipartidas, os resultados obtidos principalmente com os movimentos de alta amplitude da Impα podem também estar presentes e serem igualmente importantes para outras NLSs, incluindo NLSs clássicas monopartidas e NLSs não clássicas. Fica cada vez mais claro que as proteínas de formato solenóide normalmente apresentam alta flexibilidade, e essa característica parece ser funcionalmente importante por gerar plasticidade de ligação, possibilitando o reconhecimento de proteínas com tamanhos e formatos distintos.

Além disso, com os dados que geramos, novas análises podem ser realizadas, em especial, para entender melhor o papel da região *linker*. Uma abordagem interessante seria a realização de mutações ou deleções em sequências NLS e a posterior avaliação dos seus efeitos nas interações com a Impα. Uma análise semelhante que estamos realizando é com relação ao nosso estudo com a NLS da proteína Ku70. Apesar do nosso objetivo principal ser a caracterização da Ku70NLS como monopartida ou bipartida, estamos tentando avaliar um modelo de NLS com um possível *linker* de tamanho reduzido ao que a literatura afirma como tamanho mínimo. Portanto, esses resultados poderão nos fornecer dados importantes para a atualização da definição de NLS clássicas, e nos servir de base para outras NLSs ainda não estudadas.

Por fim, para complementar e compreender melhor os nossos dados de simulação para a Ku70, pretendemos realizar experimentos de Calorimetria de Titulação Isotérmica (ITC). A técnica de ITC é uma técnica biofísica que avalia interações bioquímicas entre moléculas em virtude de calor absorvido ou gerado por essa interação. Essa medida de calor permite uma determinação acurada de constantes de ligação (K_b), reações de estequiometria (n), entalpia (ΔH) e entropia (ΔS). Dessa maneira, é obtido um perfil termodinâmico da interação molecular em um único experimento. Por ser uma técnica que vai além das afinidades de ligação e possibilita a elucidação do mecanismo de interação molecular, como o número de sítios de ligação, a ITC é uma boa escolha para validação experimental das interações macromoleculares. Assim, poderemos complementar os nossos dados de simulação computacional referente à ligação entre Ku70NLS e Impα.

Referências

ADCOCK, S. A.; MCCAMMON, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. **Chemical reviews**, ACS Publications, v. 106, n. 5, p. 1589–1615, 2006. Citado na página 18.

ALEXANDROV, V. et al. Normal modes for predicting protein motions: a comprehensive database assessment and associated web tool. **Protein science**, Wiley Online Library, v. 14, n. 3, p. 633–643, 2005. Citado na página 36.

AMADEI, A.; LINSSEN, A.; BERENDSEN, H. Essential dynamics of proteins. **Proteins Structure Function and Bioinformatics**, n. 17, p. 412–25, 1994. Citado na página 19.

ANDRADE, M. A.; BORK, P. Heat repeats in the huntington's disease protein. **Nature genetics**, Nature Publishing Group, v. 11, n. 2, p. 115–116, 1995. Citado na página 11.

BAHAR, I. et al. Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. **Biochemistry**, ACS Publications, v. 36, n. 44, p. 13512–13523, 1997. Citado na página 19.

BAHAR, I.; RADER, A. Coarse-grained normal mode analysis in structural biology. **Current opinion in structural biology**, Elsevier, v. 15, n. 5, p. 586–592, 2005. Citado na página 19.

BAKAN, A.; MEIRELES, L. M.; BAHAR, I. Prody: protein dynamics inferred from theory and experiments. **Bioinformatics**, Oxford Univ Press, v. 27, n. 11, p. 1575–1577, 2011. Citado 2 vezes nas páginas 29 e 63.

BARDUCCI, A.; BUSSI, G.; PARRINELLO, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. **Physical review letters**, APS, v. 100, n. 2, p. 020603, 2008. Citado na página 19.

BARROS, A. C. de et al. Structural basis of nuclear import of flap endonuclease 1 (fen1). **Acta Crystallographica Section D: Biological Crystallography**, International Union of Crystallography, v. 68, n. 7, p. 743–750, 2012. Citado 6 vezes nas páginas 16, 25, 35, 62, 68 e 69.

BARROS, A. C. de 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**, Elsevier, 2016. Citado na página 34.

BERGONZO, C. et al. Multidimensional replica exchange molecular dynamics yields a converged ensemble of an rna tetranucleotide. **Journal of chemical theory and computation**, ACS Publications, v. 10, n. 1, p. 492–499, 2013. Citado na página 19.

BERNARDES, N. E. et al. Structure of importin-α from a filamentous fungus in complex with a classical nuclear localization signal. **PloS one**, Public Library of Science, v. 10, n. 6, p. e0128687, 2015. Citado na página 35.

BERNARDI, R. C.; PASCUTTI, P. G. Hybrid qm/mm molecular dynamics study of benzocaine in a membrane environment: how does a quantum mechanical treatment of both anesthetic and lipids affect their interaction. **Journal of Chemical Theory and Computation**, ACS Publications, v. 8, n. 7, p. 2197–2203, 2012. Citado na página 18.

BISCHOFF, F. R. et al. Rangap1 induces gtpase activity of nuclear ras-related ran. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 91, n. 7, p. 2587–2591, 1994. Citado na página 12.

BISCHOFF, F. R.; PONSTINGL, H. Catalysis of guanine nucleotide exchange on ran by the mitotic regulator rcc1. **Nature**, Nature Publishing Group, v. 354, n. 6348, p. 80–82, 1991. Citado na página 12.

BROOKS, B.; KARPLUS, M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 82, n. 15, p. 4995–4999, 1985. Citado 2 vezes nas páginas 33 e 36.

BROOKS, B. R. et al. Charmm: the biomolecular simulation program. **Journal of computational chemistry**, Wiley Online Library, v. 30, n. 10, p. 1545–1614, 2009. Citado 2 vezes nas páginas 27 e 64.

BRÜSCHWEILER, R. Collective protein dynamics and nuclear spin relaxation. **The Journal of chemical physics**, AIP Publishing, v. 102, n. 8, p. 3396–3403, 1995. Citado na página 28.

BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **The Journal of chemical physics**, AIP Publishing, v. 126, n. 1, p. 014101, 2007. Citado na página 26.

CHEN, M.-H. et al. Phospholipid scramblase 1 contains a nonclassical nuclear localization signal with unique binding site in importin α. **Journal of Biological Chemistry**, ASBMB, v. 280, n. 11, p. 10599–10606, 2005. Citado 4 vezes nas páginas 16, 25, 35 e 69.

CHEN, V. B. et al. Molprobity: all-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D: Biological Crystallography**, International Union of Crystallography, v. 66, n. 1, p. 12–21, 2009. Citado 2 vezes nas páginas 26 e 63.

CHOOK, Y.; BLOBEL, G. Karyopherins and nuclear import. **Current opinion in structural biology**, Elsevier, v. 11, n. 6, p. 703–715, 2001. Citado 2 vezes nas páginas 13 e 14.

CHRISTIE, M. et al. Structural biology and regulation of protein import into the nucleus. **Journal of molecular biology**, Elsevier, 2015. Citado 5 vezes nas páginas 13, 17, 24, 35 e 62.

CHRISTMANN, M. et al. Mechanisms of human dna repair: an update. **Toxicology**, Elsevier, v. 193, n. 1, p. 3–34, 2003. Citado na página 62.

CINGOLANI, G. et al. Structure of importin-β bound to the ibb domain of importin-α. **Nature**, Nature Publishing Group, v. 399, n. 6733, p. 221–229, 1999. Citado na página 24.

COLLEDGE, W. et al. Extensive mutagenesis of the nuclear location signal of simian virus 40 large-t antigen. **Molecular and cellular biology**, Am Soc Microbiol, v. 6, n. 11, p. 4136–4139, 1986. Citado na página 16.

CONTI, E.; KURIYAN, J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin α. **Structure**, Elsevier, v. 8, n. 3, p. 329–338, 2000. Citado 5 vezes nas páginas 12, 16, 25, 35 e 69.

CONTI, E. et al. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α. **Cell**, Elsevier, v. 94, n. 2, p. 193–204, 1998. Citado 3 vezes nas páginas 12, 15 e 24.

COOK, A. et al. Structural biology of nucleocytoplasmic transport. **Annu. Rev. Biochem.**, Annual Reviews, v. 76, p. 647–671, 2007. Citado na página 12.

COSTA, M. G. d. S. et al. Exploring free energy landscapes of large conformational changes: Molecular dynamics with excited normal modes. **Journal of Chemical Theory and Computation**, ACS Publications, 2015. Citado 5 vezes nas páginas 20, 63, 64, 67 e 70.

CRONSHAW, J. M. et al. Proteomic analysis of the mammalian nuclear pore complex. **The Journal of cell biology**, Rockefeller Univ Press, v. 158, n. 5, p. 915–927, 2002. Citado na página 11.

CUI, Q.; BAHAR, I. **Normal mode analysis: theory and applications to biological and chemical systems**. [S.l.]: CRC press, 2010. Citado 2 vezes nas páginas 19 e 20.

CUTRESS, M. L. et al. Structural basis for the nuclear import of the human androgen receptor. **Journal of cell science**, The Company of Biologists Ltd, v. 121, n. 7, p. 957–968, 2008. Citado 4 vezes nas páginas 16, 25, 35 e 69.

DELARUE, M.; SANEJOUAND, Y.-H. Simplified normal mode analysis of conformational transitions in dna-dependent polymerases: the elastic network model. **Journal of molecular biology**, Elsevier, v. 320, n. 5, p. 1011–1024, 2002. Citado na página 19.

DINGWALL, C.; LASKEY, R. Nucleoplasmin: the archetypal molecular chaperone. In: **Seminars in cell biology**. [S.l.: s.n.], 1990. v. 1, n. 1, p. 11–17. Citado na página 25.

DINGWALL, C. et al. The nucleoplasmin nuclear location sequence is larger and more complex than that of sv-40 large t antigen. **The Journal of cell biology**, Rockefeller Univ Press, v. 107, n. 3, p. 841–849, 1988. Citado 2 vezes nas páginas 15 e 24.

DOYLE, L. et al. Rational design of α-helical tandem repeat proteins with closed architectures. **Nature**, Nature Publishing Group, v. 528, n. 7583, p. 585–588, 2015. Citado na página 24.

DROR, R. O. et al. Biomolecular simulation: a computational microscope for molecular biology. **Annual review of biophysics**, Annual Reviews, v. 41, p. 429–452, 2012. Citado na página 18.

EMSLEY, P. et al. Features and development of coot. **Acta Crystallographica Section D: Biological Crystallography**, International Union of Crystallography, v. 66, n. 4, p. 486–501, 2010. Citado na página 29.

ESSMANN, U. et al. A smooth particle mesh ewald method. **The Journal of chemical physics**, AIP Publishing, v. 103, n. 19, p. 8577–8593, 1995. Citado na página 26.

ESWAR, N. et al. Comparative protein structure modeling using modeller. **Current protocols in bioinformatics**, Wiley Online Library, p. 5–6, 2006. Citado na página 26.

FAHRENKROG, B.; AEBI, U. The nuclear pore complex: nucleocytoplasmic transport and beyond. **Nature Reviews Molecular Cell Biology**, Nature Publishing Group, v. 4, n. 10, p. 757–766, 2003. Citado 3 vezes nas páginas 11, 13 e 14.

FALCES, J. et al. Recognition of nucleoplasmin by its nuclear transport receptor importin α/β: insights into a complete import complex. **Biochemistry**, ACS Publications, v. 49, n. 45, p. 9756–9769, 2010. Citado 2 vezes nas páginas 34 e 69.

FELDHERR, C. M.; AKIN, D.; COHEN, R. J. Regulation of functional nuclear pore size in fibroblasts. **Journal of cell science**, The Company of Biologists Ltd, v. 114, n. 24, p. 4621–4627, 2001. Citado na página 11.

FLOCH, A. G.; PALANCADE, B.; DOYE, V. Fifty years of nuclear pores and nucleocytoplasmic transport studies: multiple tools revealing complex rules. **Methods in cell biology**, v. 122, p. 1–40, 2013. Citado na página 11.

FLOQUET, N. et al. Conformational equilibrium of cdk/cyclin complexes by molecular dynamics with excited normal modes. **Biophysical journal**, Elsevier, v. 109, n. 6, p. 1179–1189, 2015. Citado na página 28.

FONTES, M. R. et al. Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-α. **Journal of Biological Chemistry**, ASBMB, v. 278, n. 30, p. 27981–27987, 2003. Citado 7 vezes nas páginas 16, 24, 25, 34, 35, 62 e 69.

FONTES, M. R.; TEH, T.; KOBE, B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-α. **Journal of molecular biology**, Elsevier, v. 297, n. 5, p. 1183–1194, 2000. Citado 9 vezes nas páginas 15, 16, 25, 30, 34, 35, 62, 68 e 69.

FORWOOD, J. K. et al. Quantitative structural analysis of importin-β flexibility: paradigm for solenoid protein structures. **Structure**, Elsevier, v. 18, n. 9, p. 1171–1183, 2010. Citado 2 vezes nas páginas 25 e 33.

GARCÍA, A. E.; HARMAN, J. G. Simulations of crp:(camp) 2 in noncrystalline environments show a subunit transition from the open to the closed conformation. **Protein science: a publication of the Protein Society**, Blackwell Publishing, v. 5, n. 1, p. 62, 1996. Citado na página 19.

GARCÍA, A. E.; SANBONMATSU, K. Y. Exploring the energy landscape of a β hairpin in explicit solvent. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 42, n. 3, p. 345–354, 2001. Citado na página 19.

GENHEDEN, S.; RYDE, U. Will molecular dynamics simulations of proteins ever reach equilibrium? **Physical Chemistry Chemical Physics**, Royal Society of Chemistry, v. 14, n. 24, p. 8662–8677, 2012. Citado na página 36.

GERALDO, M. T. et al. Bending-twisting motions and main interactions in nucleoplasmin nuclear import. **PLOS ONE**, Public Library of Science, v. 11, n. 6, p. e0157162, 2016. Citado 2 vezes nas páginas 65 e 69.

GIESECKE, A.; STEWART, M. Novel binding of the mitotic regulator tpx2 (target protein for xenopus kinesin-like protein 2) to importin-α. **Journal of Biological Chemistry**, ASBMB, v. 285, n. 23, p. 17628–17635, 2010. Citado 4 vezes nas páginas 16, 25, 35 e 69.

GIPSON, B. et al. Computational models of protein kinematics and dynamics: Beyond simulation. **Annual review of analytical chemistry**, Annual Reviews, v. 5, p. 273–291, 2012. Citado na página 18.

GÖRLICH, D. et al. A 41 amino acid motif in importin-alpha confers binding to importin-beta and hence transit into the nucleus. **The EMBO Journal**, Nature Publishing Group, v. 15, n. 8, p. 1810, 1996. Citado na página 24.

GÖRLICH, D. et al. Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. **Current Biology**, Elsevier, v. 5, n. 4, p. 383–392, 1995. Citado na página 24.

GÖRLICH, D.; KUTAY, U. Transport between the cell nucleus and the cytoplasm. **Annual review of cell and developmental biology**, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 15, n. 1, p. 607–660, 1999. Citado 3 vezes nas páginas 11, 13 e 14.

GÖRLICH, D. et al. Identification of different roles for rangdp and rangtp in nuclear protein import. **The EMBO journal**, Nature Publishing Group, v. 15, n. 20, p. 5584, 1996. Citado na página 12.

GÖRLICH, D. et al. Distinct functions for the two importin subunits in nuclear protein import. Nature Publishing Group, 1995. Citado na página 12.

GROSSFIELD, A.; ZUCKERMAN, D. M. Quantifying uncertainty and sampling quality in biomolecular simulations. **Annual reports in computational chemistry**, Elsevier, v. 5, p. 23–48, 2009. Citado na página 36.

HALPERIN, I. et al. Principles of docking: An overview of search algorithms and a guide to scoring functions. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 47, n. 4, p. 409–443, 2002. Citado na página 18.

HAYWARD, S.; KITAO, A.; BERENDSEN, H. J. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 27, n. 3, p. 425–437, 1997. Citado 2 vezes nas páginas 34 e 36.

HENZLER-WILDMAN, K.; KERN, D. Dynamic personalities of proteins. **Nature**, Nature Publishing Group, v. 450, n. 7172, p. 964–972, 2007. Citado na página 18.

HESS, B. P-lincs: A parallel linear constraint solver for molecular simulation. **Journal of Chemical Theory and Computation**, ACS Publications, v. 4, n. 1, p. 116–122, 2008. Citado na página 26.

HESS, B. et al. Lincs: a linear constraint solver for molecular simulations. **Journal of computational chemistry**, Citeseer, v. 18, n. 12, p. 1463–1472, 1997. Citado na página 26.

HINSEN, K. Analysis of domain motions by approximate normal mode calculations. **Proteins Structure Function and Genetics**, v. 33, n. 3, p. 417–429, 1998. Citado na página 20.

HINSEN, K.; THOMAS, A.; FIELD, M. J. Analysis of domain motions in large proteins. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 34, n. 3, p. 369–382, 1999. Citado na página 20.

HODEL, M. R.; CORBETT, A. H.; HODEL, A. E. Dissection of a nuclear localization signal. **Journal of Biological Chemistry**, ASBMB, v. 276, n. 2, p. 1317–1325, 2001. Citado 2 vezes nas páginas 16 e 25.

HUANG, J.; MACKERELL, A. D. Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. **Journal of computational chemistry**, Wiley Online Library, v. 34, n. 25, p. 2135–2145, 2013. Citado na página 26.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. Vmd: visual molecular dynamics. **Journal of molecular graphics**, Elsevier, v. 14, n. 1, p. 33–38, 1996. Citado na página 29.

ICHIYE, T.; KARPLUS, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 11, n. 3, p. 205–217, 1991. Citado 2 vezes nas páginas 33 e 36.

IHAKA, R.; GENTLEMAN, R. R: a language for data analysis and graphics. **Journal of computational and graphical statistics**, Taylor & Francis, v. 5, n. 3, p. 299–314, 1996. Citado na página 29.

JOLLIFFE, I. T. Principal component analysis, ser. **Springer Ser. Statist., 2nd ed. New York: Springer**, 2002. Citado na página 19.

JORGENSEN, W. L. et al. Comparison of simple potential functions for simulating liquid water. **The Journal of chemical physics**, AIP Publishing, v. 79, n. 2, p. 926–935, 1983. Citado na página 26.

KAJAVA, A. V. Review: proteins with repeated sequence—structural prediction and modeling. **Journal of structural biology**, Elsevier, v. 134, n. 2, p. 132–144, 2001. Citado na página 33.

KALDERON, D. et al. Sequence requirements for nuclear location of simian virus 40 large-t antigen. Nature Publishing Group, 1984. Citado na página 15.

KALDERON, D. et al. A short amino acid sequence able to specify nuclear location. **Cell**, Elsevier, v. 39, n. 3, p. 499–509, 1984. Citado na página 24.

KAPPEL, C. et al. An unusual hydrophobic core confers extreme flexibility to heat repeat proteins. **Biophysical journal**, Elsevier, v. 99, n. 5, p. 1596–1603, 2010. Citado 2 vezes nas páginas 24 e 25.

KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. **Nature Structural & Molecular Biology**, Nature Publishing Group, v. 9, n. 9, p. 646–652, 2002. Citado na página 18.

KESKIN, O. et al. Relating molecular flexibility to function: a case study of tubulin. **Biophysical journal**, Elsevier, v. 83, n. 2, p. 663–680, 2002. Citado na página 33.

KIM, M. K.; JERNIGAN, R. L.; CHIRIKJIAN, G. S. An elastic network model of hk97 capsid maturation. **Journal of Structural Biology**, Elsevier, v. 143, n. 2, p. 107–117, 2003. Citado na página 20.

KIRBY, T. W. et al. Nuclear localization of the dna repair scaffold xrcc1: Uncovering the functional role of a bipartite nls. **Scientific reports**, Nature Publishing Group, v. 5, 2015. Citado na página 34.

KITAO, A.; GO, N. Investigating protein dynamics in collective coordinate space. **Current opinion in structural biology**, Elsevier, v. 9, n. 2, p. 164–169, 1999. Citado na página 19.

KNUDSEN, N. Ø. et al. Nuclear translocation contributes to regulation of dna excision repair activities. **DNA repair**, Elsevier, v. 8, n. 6, p. 682–689, 2009. Citado na página 62.

KOBE, B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin α. **Nature Structural & Molecular Biology**, Nature Publishing Group, v. 6, n. 4, p. 388–397, 1999. Citado 3 vezes nas páginas 12, 24 e 34.

KOBE, B.; KAJAVA, A. V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. **Trends in biochemical sciences**, Elsevier, v. 25, n. 10, p. 509–515, 2000. Citado 3 vezes nas páginas 11, 24 e 33.

KOIKE, M. et al. Ku80 can translocate to the nucleus independent of the translocation of ku70 using its own nuclear localization signal. **Oncogene**, v. 18, n. 52, p. 7495–7505, 1999. Citado 2 vezes nas páginas 62 e 63.

KOIKE, M. et al. The nuclear localization signal of the human ku70 is a variant bipartite type recognized by the two components of nuclear pore-targeting complex. **Experimental cell research**, Elsevier, v. 250, n. 2, p. 401–413, 1999. Citado 3 vezes nas páginas 62, 63 e 68.

KOIKE, M.; SHIOMI, T.; KOIKE, A. Dimerization and nuclear localization of ku proteins. **Journal of Biological Chemistry**, ASBMB, v. 276, n. 14, p. 11167–11173, 2001. Citado na página 62.

KOSUGI, S. et al. Six classes of nuclear localization signals specific to different binding grooves of importin α. **Journal of Biological Chemistry**, ASBMB, v. 284, n. 1, p. 478–485, 2009. Citado 4 vezes nas páginas 13, 24, 35 e 69.

KRISSINEL, E.; HENRICK, K. Inference of macromolecular assemblies from crystalline state. **Journal of molecular biology**, Elsevier, v. 372, n. 3, p. 774–797, 2007. Citado na página 26.

KUTAY, U.; GÜTTINGER, S. Leucine-rich nuclear-export signals: born to be weak. **Trends in cell biology**, Elsevier, v. 15, n. 3, p. 121–124, 2005. Citado na página 12.

LAI, M.-C. et al. A human importin-β family protein, transportin-sr2, interacts with the phosphorylated rs domain of sr proteins. **Journal of Biological Chemistry**, ASBMB, v. 275, n. 11, p. 7950–7957, 2000. Citado na página 16.

LAINE, E. et al. Mutation d816v alters the internal structure and dynamics of c-kit receptor cytoplasmic region: implications for dimerization and activation mechanisms. **PLoS Comput Biol**, v. 7, n. 6, p. e1002068–e1002068, 2011. Citado na página 70.

LAIO, A.; PARRINELLO, M. Escaping free-energy minima. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 99, n. 20, p. 12562–12566, 2002. Citado na página 19.

LANGE, A. et al. Expanding the definition of the classical bipartite nuclear localization signal. **Traffic**, Wiley Online Library, v. 11, n. 3, p. 311–323, 2010. Citado na página 69.

LEE, S. J. et al. The adoption of a twisted structure of importin-β is essential for the protein-protein interaction required for nuclear transport. **Journal of molecular biology**, Elsevier, v. 302, n. 1, p. 251–264, 2000. Citado na página 33.

LEE, S. J. et al. Structural basis for nuclear import complex dissociation by rangtp. **Nature**, Nature Publishing Group, v. 435, n. 7042, p. 693–696, 2005. Citado na página 24.

LEUNG, S. W. et al. Dissection of the karyopherin α nuclear localization signal (nls)-binding groove: Functional requirements for nls binding. **Journal of Biological Chemistry**, ASBMB, v. 278, n. 43, p. 41947–41953, 2003. Citado na página 15.

LYMAN, E.; ZUCKERMAN, D. M. Ensemble-based convergence analysis of biomolecular trajectories. **Biophysical journal**, Elsevier, v. 91, n. 1, p. 164–172, 2006. Citado na página 27.

MA, J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. **Structure**, Elsevier, v. 13, n. 3, p. 373–380, 2005. Citado na página 20.

MACKERELL, A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. **The journal of physical chemistry B**, ACS Publications, v. 102, n. 18, p. 3586–3616, 1998. Citado na página 26.

MAKKERH, J. P.; DINGWALL, C.; LASKEY, R. A. Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. **Current Biology**, Elsevier, v. 6, n. 8, p. 1025–1027, 1996. Citado 2 vezes nas páginas 16 e 25.

MARFORI, M. et al. Structural basis of high-affinity nuclear localization signal interactions with importin-α. **Traffic**, Wiley Online Library, v. 13, n. 4, p. 532–548, 2012. Citado 10 vezes nas páginas 12, 16, 24, 25, 30, 34, 35, 62, 68 e 69.

MARFORI, M. et al. Molecular basis for specificity of nuclear import and prediction of nuclear localization. **Biochimica et Biophysica Acta (BBA)-Molecular Cell Research**, Elsevier, v. 1813, n. 9, p. 1562–1577, 2011. Citado 2 vezes nas páginas 16 e 63.

MARSILI, S. et al. Orac: A molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. **Journal of computational chemistry**, Wiley Online Library, v. 31, n. 5, p. 1106–1116, 2010. Citado na página 19.

MATSUURA, Y.; STEWART, M. Structural basis for the assembly of a nuclear export complex. **Nature**, Nature Publishing Group, v. 432, n. 7019, p. 872–877, 2004. Citado na página 24.

MATSUURA, Y.; STEWART, M. Nup50/npap60 function in nuclear protein import complex disassembly and importin recycling. **The EMBO journal**, EMBO Press, v. 24, n. 21, p. 3681–3689, 2005. Citado na página 24.

MOROIANU, J.; BLOBEL, G.; RADU, A. Nuclear protein import: Ran-gtp dissociates the karyopherin alphabeta heterodimer by displacing alpha from an overlapping binding site on beta. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 93, n. 14, p. 7059–7062, 1996. Citado 2 vezes nas páginas 12 e 24.

OCHI, T. et al. Structural biology of dna repair: spatial organisation of the multicomponent complexes of nonhomologous end joining. **Journal of nucleic acids**, Hindawi Publishing Corporation, v. 2010, 2010. Citado na página 62.

ONUCHIC, J. N.; LUTHEY-SCHULTEN, Z.; WOLYNES, P. G. Theory of protein folding: the energy landscape perspective. **Annual review of physical chemistry**, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 48, n. 1, p. 545–600, 1997. Citado na página 18.

OROZCO, M. A theoretical view of protein dynamics. **Chemical Society Reviews**, Royal Society of Chemistry, v. 43, n. 14, p. 5051–5066, 2014. Citado na página 18.

PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied physics**, AIP Publishing, v. 52, n. 12, p. 7182–7190, 1981. Citado na página 27.

PEMBERTON, L. F.; PASCHAL, B. M. Mechanisms of receptor-mediated nuclear import and nuclear export. **Traffic**, Wiley Online Library, v. 6, n. 3, p. 187–198, 2005. Citado 3 vezes nas páginas 11, 13 e 14.

PFEIFFER, P.; GOEDECKE, W.; OBE, G. Mechanisms of dna double-strand break repair and their potential to induce chromosomal aberrations. **Mutagenesis**, Oxford Univ Press, v. 15, n. 4, p. 289–302, 2000. Citado na página 62.

PUMROY, R. A. et al. Molecular determinants for nuclear import of influenza a pb2 by importin α isoforms 3 and 7. **Structure**, Elsevier, v. 23, n. 2, p. 374–384, 2015. Citado 2 vezes nas páginas 33 e 34.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2015. Disponível em: <http://www.R-project.org>. Citado na página 65.

RADU, A.; BLOBEL, G.; MOORE, M. S. Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 92, n. 5, p. 1769–1773, 1995. Citado na página 24.

REXACH, M.; BLOBEL, G. Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. **Cell**, Elsevier, v. 83, n. 5, p. 683–692, 1995. Citado na página 12.

RIGGLEMAN, B.; WIESCHAUS, E.; SCHEDL, P. Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a drosophila segment polarity gene. **Genes & development**, Cold Spring Harbor Lab, v. 3, n. 1, p. 96–113, 1989. Citado na página 24.

ROBBINS, J. et al. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. **Cell**, Elsevier, v. 64, n. 3, p. 615–623, 1991. Citado 2 vezes nas páginas 15 e 24.

ROMO, T. et al. Singular value decomposition analysis of time-averaged crystallographic refinement. **Proteins**, v. 22, p. 311–321, 1995. Citado na página 19.

ROUT, M. P. et al. The yeast nuclear pore complex composition, architecture, and transport mechanism. **The Journal of cell biology**, Rockefeller Univ Press, v. 148, n. 4, p. 635–652, 2000. Citado na página 11.

ROUT, M. P.; WENTE, S. R. Pores for thought: nuclear pore complex proteins. **Trends in cell biology**, Elsevier, v. 4, n. 10, p. 357–365, 1994. Citado na página 11.

SCHLICK, T. et al. Algorithmic challenges in computational molecular biophysics. **Journal of Computational Physics**, Elsevier, v. 151, n. 1, p. 9–48, 1999. Citado na página 18.

SCHRÖDINGER, L. The pymol molecular graphics system, version 1.3 r1; schrödinger, llc. **New York**, 2010. Citado na página 30.

SEEBER, M. et al. Wordom: a program for efficient analysis of molecular dynamics simulations. **Bioinformatics**, Oxford Univ Press, v. 23, n. 19, p. 2625–2627, 2007. Citado 3 vezes nas páginas 29, 65 e 66.

SENN, H. M.; THIEL, W. Qm/mm methods for biological systems. In: **Atomistic approaches in modern biology**. [S.l.]: Springer, 2007. p. 173–290. Citado na página 18.

SETTANNI, G. et al. Effects of ligand binding on the mechanical properties of ankyrin repeat protein gankyrin. **PLoS Comput Biol**, Public Library of Science, v. 9, n. 1, p. e1002864, 2013. Citado na página 24.

SKJAERVEN, L.; MARTINEZ, A.; REUTER, N. Principal component and normal mode analysis of proteins; a quantitative comparison using the groel subunit. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 79, n. 1, p. 232–243, 2011. Citado na página 36.

SOTHISELVAM, S. et al. Macrolide antibiotics allosterically predispose the ribosome for translation arrest. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 111, n. 27, p. 9804–9809, 2014. Citado na página 18.

STEWART, M. Molecular mechanism of the nuclear protein import cycle. **Nature Reviews Molecular Cell Biology**, Nature Publishing Group, v. 8, n. 3, p. 195–208, 2007. Citado 5 vezes nas páginas 11, 12, 13, 14 e 62.

SUGITA, Y.; OKAMOTO, Y. Replica-exchange molecular dynamics method for protein folding. **Chemical physics letters**, Elsevier, v. 314, n. 1, p. 141–151, 1999. Citado na página 19.

SUHRE, K.; SANEJOUAND, Y.-H. Elnemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. **Nucleic acids research**, Oxford Univ Press, v. 32, n. suppl 2, p. W610–W614, 2004. Citado na página 20.

SUN, F. et al. Atomic model of rabbit hemorrhagic disease virus by cryo-electron microscopy and crystallography. **Biophysical Journal**, Elsevier, v. 104, n. 2, p. 414a, 2013. Citado na página 18.

TAKEDA, A. A. et al. Structural basis of importin-α-mediated nuclear transport for ku70 and ku80. **Journal of molecular biology**, Elsevier, v. 412, n. 2, p. 226–234, 2011. Citado na página 63.

TAKEDA, A. A. et al. Biophysical characterization of the recombinant importin-α from neurospora crassa. **Protein and peptide letters**, Bentham Science Publishers, v. 20, n. 1, p. 8–16, 2013. Citado 3 vezes nas páginas 34, 36 e 69.

TAMA, F. et al. Building-block approach for determining low-frequency normal modes of macromolecules. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 41, n. 1, p. 1–7, 2000. Citado na página 19.

TAMA, F.; SANEJOUAND, Y.-H. Conformational change of proteins arising from normal mode calculations. **Protein engineering**, Oxford Univ Press, v. 14, n. 1, p. 1–6, 2001. Citado 2 vezes nas páginas 19 e 28.

TAMA, F. et al. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 100, n. 16, p. 9319–9323, 2003. Citado na página 20.

TSALLIS, C.; STARIOLO, D. A. Generalized simulated annealing. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 233, n. 1, p. 395–406, 1996. Citado na página 19.

VALADIE, H. et al. Dynamical properties of the mscl of escherichia coli: a normal mode analysis. **Journal of molecular biology**, Elsevier, v. 332, n. 3, p. 657–674, 2003. Citado na página 20.

WALKER, J. R.; CORPINA, R. A.; GOLDBERG, J. Structure of the ku heterodimer bound to dna and its implications for double-strand break repair. **Nature**, Nature Publishing Group, v. 412, n. 6847, p. 607–614, 2001. Citado na página 62.

WEIS, K.; RYDER, U.; LAMOND, A. I. The conserved amino-terminal domain of hsrp1 alpha is essential for nuclear protein import. **The EMBO journal**, European Molecular Biology Organization, v. 15, n. 8, p. 1818, 1996. Citado na página 12.

YANG, J.; WANG, Y.; CHEN, Y. Gpu accelerated molecular dynamics simulation of thermal conductivities. **Journal of Computational Physics**, Elsevier, v. 221, n. 2, p. 799–804, 2007. Citado na página 18.

YANG, S. N. et al. Probing the specificity of binding to the major nuclear localization sequence-binding site of importin-α using oriented peptide library screening. **Journal of Biological Chemistry**, ASBMB, v. 285, n. 26, p. 19935–19946, 2010. Citado 5 vezes nas páginas 15, 16, 25, 35 e 69.

ZHAO, G. et al. Mature hiv-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. **Nature**, Nature Publishing Group, v. 497, n. 7451, p. 643–646, 2013. Citado na página 18.