Importin α2 associates with chromatin via a novel DNA binding domain

Kazuya Jibiki[1], Takashi Kodama[2, 5], Atsushi Suenaga[1,3], Yota Kawase[1], Noriko Shibazaki[3], Shin Nomoto[1], Seiya Nagasawa[3], Misaki Nagashima[3], Shieri Shimodan[3], Renan Kikuchi[3], Noriko Saitoh[4], Yoshihiro Yoneda[5], Ken-ich Akagi[5, 6], Noriko Yasuhara[1, 3*]

1 Graduate School of Integrated Basic Sciences, Nihon University, Setagaya-ku, Tokyo, Japan

2 Laboratory of Molecular Biophysics, Institute for Protein Research, Osaka University, Osaka, Japan

3 Department of Biosciences, College of Humanities and Sciences, Nihon University, Setagaya-ku, Tokyo, Japan

4 Division of Cancer Biology, The Cancer Institute of JFCR, Tokyo, Japan

5 National Institute of Biomedical Innovation, Health and Nutrition (NIBIOHN), Osaka, Japan

6 present address: Environmental Metabolic Analysis Research Team, RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa, Japan

* Corresponding author. E-mail: yasuhara@chs.nihon-u.ac.jp

Abstract

The nuclear transport of functional proteins is important for facilitating appropriate gene expression. The proteins of the importin α family of nuclear transport receptors operate via several pathways to perform their nuclear protein import function. Additionally, these proteins are also reported to possess other functions, including chromatin association and gene regulation. However, these non-transport functions are not yet fully understood. Here, we report novel molecular characteristics of importin α involving the binding to multiple regions of chromatin. We identified the importin α DNA binding domain-the Nucleic Acid Associating Trolley pole domain (NAAT domain) as helical structures within the N terminal IBB Domain. We propose a "stroll and locate" model to explain how importin α associates with double-strand DNA. This is the first study to show that importin α interacts with chromatin via novel DNA binding domain.

Introduction

The importin α family is a class of nuclear transport receptors that mediates protein translocation into the eukaryotic cell nucleus through the nuclear pore (Goldfarbet et al, 2004). Proteins are generally synthesised in the cytoplasm, so nuclear proteins, such as transcription factors, have to be transported into the nucleus via transport receptors such as importins. Importin α proteins recognise their transport cargo proteins by the protein's nuclear localisation signal (NLS), which is mainly composed of basic amino acids (Kalderon et al,1984; Lange et al , 2007), and it imports them by forming a trimeric complex with importin β1 and the cargo protein (Görlich & Mattaj ,1996; Oka & Yoneda 2018; Imamoto et al ,1995). The protein is then released from the importins by the binding of Ran-GTP to importin β1, which facilitates the dissociation of importin β1 from the importin α-cargo protein complex (Görlich & Mattaj ,1996), and by the binding of Nup50 or CAS to importin α, which facilitates the dissociation of importin α from its cargo (Matsuura& Stewart , 2005; Kutay et al 1997; Lindsay et al, 2002).

Architecturally, importin α family proteins consist of three domains; 1) N-terminal importin β binding (IBB) domain which interacts with importin β1 or otherwise binds in an autoregulatory fashion to the NLS binding sites of importin α2 itself (a.a 1-69); 2) a stable helix repeat domain called armadillo (ARM) repeats including two NLS binding sites (a.a 69-392); and 3) the C terminal region, including the Nup50 and CAS binding domain (a.a 392-529). (Kobe,1999; Kaylen&Gino ,2010; Miyamoto et al, 2016)

The importin α family proteins are expressed from several family genes in mammalian cells. Their expression varies widely depending on the cell types with cargo specificity, thereby regulating the protein activity in the nucleus through selective nuclear protein transport (Hu et al, 2010; Mihalas et al, 2015; Young et al, 2011). In this study, we designate the importin α family

proteins as importin α1 (KPNA1, NPI1, importin α5 in humans), importin α2 (KPNA2, Rch1, importin α1 in humans), importin α3 (KPNA3, Qip2, importin α4 in humans), importin α4 (KPNA4, Qip1, importin α3 in humans), importin α6 (KPNA6, NPI2, importin α7 in humans) and importin α8 (KPNA7).

Importin α proteins have been shown to act as multi-functional proteins in cellular activities in addition to their NLS transport receptor function for selective nuclear transport. Their additional functions traverse spindle assembly, lamin polymerisation, nuclear envelope formation, protein degradation, cytoplasmic retention, gene expression, cell surface function and mRNA-related functions (Miyamoto et al ,2016). Additionally, importin α family members are also known to accumulate in the nucleus under certain stress condition, such as heat shock and oxidative stress wherein they bind to a DNase-sensitive nuclear component (Kodiha et al,2004 ; Furuta et al,2004; Miyamoto et al, 2004) and regulate transcription of specific genes, such as STK35 (Yasuda et al, 2012).

Importin α2 is known to play roles in maintenance of undifferentiated state of mouse ES cells. The mechanism by which importin α2 influences ES cells fate is not yet fully understood. For example, Oct3/4 is an autoregulatory gene (Niwa et al, 2000; Niwa ,2007), so one possible model involves the upregulation of protein expression of Oct3/4 by its own enhanced nuclear import. This could explain why importin α2 expression is necessary for ES cell maintenance, as it is the main nuclear transporter of Oct3/4 in the undifferentiated state. However, Oct3/4 is known to induce differentiation when expressed in excess (Niwa et al, 2000; Niwa ,2007), and a nuclear export accelerated mutant of Oct3/4 still maintained the undifferentiated state of ES cells (Oka et al,2013), suggesting that the accumulation of Oct3/4 molecules within the nucleus may not affect its expression level by autoregulation. One activity of importin α2 that could possibly influence gene expression in ES cells is its interaction with chromatin. We hypothesised that importin α2 may also interact with chromatin of undifferentiated ES cells to influence gene expression levels. Undifferentiated mouse ES cells are particularly appropriate to study importin α functions, as a single family member, importin α2, is predominantly expressed over other family proteins.

In the present study, we tried to investigate the molecular mechanism of the importin α chromatin association using mouse ES cells and revealed that importin α2 bound to multiple regions in the undifferentiated mouse ES cell genome through direct DNA binding. Furthermore, importin α2 directly bound the upstream region of Oct3/4 gene through a novel chromatin associating domain in the IBB Domain. We also found that the association of importin α2 with chromatin was multi-mode and that the protein was able to stroll around the DNA.

This is the first study to reveal importin α as a direct DNA binding protein with a novel DNA binding domain.

Results

1. Importin α2 interacts with the genomic region of ES cells

Expression of importin α2 is highly and predominantly maintained in mouse undifferentiated ES cells (Yasuhara et al, 2007). We first checked the nuclear distribution of importin α2 in undifferentiated mouse ES cells. Endogenous importin α2 was localised both in the cytoplasm and the nucleus (Fig. 1A), as determined by immunostaining assays. The use of GFP fused importin α2 confirmed this localisation in the cytoplasm and the nucleus of undifferentiated ES cells. The nuclear localisation was more apparent than endogenous distribution when the protein was strongly expressed, while the distribution of control GFP was dispersed (Fig. 1B, C). As importin α2 play essential roles in the maintenance of undifferentiated state of ES cells (Yasuhara et al, 2007; Li et al, 2008; Young et al, 2011;Yasuhara et al, 2013), we focussed to study the mechanism and the function of nuclear localisation of importin α2.

We tested whether importin α2 binds to chromatin in ES cells. Previous studies indicated an effect of importin α2 on the expression of Oct3/4 (Yasuhara et al, 2007; Li et al, 2008; Young et al, 2011;Yasuhara et al, 2013), so we selected the Oct3/4 gene POU5F1 upstream sequence as a candidate for model fragment DNA for identifying the molecular action of importin α on chromatin. Two different 600 bp DNA sequences of mouse Oct3/4 gene were chosen for *in vitro* binding assays (we call these "upstream-1" and "upstream-2", where upstream-1 includes the conserved distal enhancer CR4 domain and upstream-2 includes the proximal enhancer domain). The binding potential of importin α2 to the two POU5F1 upstream regions was first confirmed by ChIP-quantitative PCR (qPCR) with importin α2 antibody in the undifferentiated ES cells. Primer sets to amplify the first 200bp of each upstream region were used in importin α2 ChIP-qPCR (Fig. 2A). As a result, upstream-1 and upstream-2 stably detected positive PCR amplification from importin α2 ChIP samples in independent assays (Fig.2B, EV1-3). These results suggested that multiple POU5F1 genomic region potentially interacts with importin α2. Therefore, we further adopted these regions to identify the binding domain of importin α2 to DNA.

2. importin α2 directly binds the DNA via the IBB domain

First, we checked whether the importin α2 protein directly binds to DNA by performing gel shift assays using the recombinant importin α2 proteins and double strand DNA obtained from the mouse ES cell genome (fig. 3, fig. EV4A). The two upstream 600 bp DNA sequences of the Oct3/4 gene (Fig. 2A) were used in this assay. The wild type importin α2 interacted with both of the upstream fragments of Oct3/4 genomic DNA, confirming that the DNA fragment from this region effectively bound importin α2 *in vitro* (Fig. 3B, C).

Next, we determined the DNA binding region of importin α2 by conducting further gel shift assays using the recombinant importin α2 mutant proteins lacking functional domains. We focused on the three major functional domains of importin α2, 1) the IBB domain, (a.a 1-69); 2) the main body (a.a 70-392); and 3) the C terminal region (a.a 393-529) (Fig. 3A, EV4A). The recombinant mutant proteins including the IBB domain showed affinity for both the upstream-1 and upstream-2 600 bp DNA sequences of the Oct3/4 gene, whereas the ΔIBB domain mutant (70-529) caused hardly any shift compared to the native DNA band (Fig. 3B, C). Here, the IBB domain that has been characterized as a coordinator of nuclear transport and autoinhibition, was also necessary for DNA binding of importin α2 *in vitro*.

3. Basic amino acid clusters within IBB domain contribute to the binding of importin α2 to DNA

To elucidate the portion of IBB domain responsible for DNA binding, we also examined whether the sequence homology to other nucleic acid binding proteins was present in the IBB domain sequence of importin α. A protein BLAST search performed in the protein and nucleic acid structure database (PDB) revealed that only the importin IBB domain families showed significant homology to the whole region of the IBB domain of importin α2 (data not shown). However, a BLAST search under search parameters adjusted for short input sequences showed local short motifs that were frequently matched against the other sequences in PDB. The DNA-binding and RNA-binding proteins were then extracted from the hit sequences, and revealed 24,830 of 156,365 Protein Data bank (PDB) entries that showed homology to KPNA2_22-51 (KDSTEMRRRRIEVNVELRKAKKDEQMLKRR). Among these were 675 entries of complexes containing both DNA and protein. In addition, 799 entries were complexes containing both RNA and protein. Surprisingly, these accounted for about 13% of all the DNA-protein complexes (5071 entries) in the PDB and about 38% of all RNA-protein complexes (2104 entries), respectively. These hits included 27 entries containing the tetra-R motif found in IBB domain in complexes with DNA (Fig. EV5) and 36 entries in complexes with RNA (Fig. EV6). Therefore, the basic sequence propensity of the importin α IBB domain was a common character in nucleic acid-binding proteins. Furthermore, a clear diversity was observed in the interaction pattern of the tetra-R motif with the nucleic acid, in the secondary structure of the region containing the motif in the complex and in the target nucleic acid sequence (Fig. EV5, 6). The manner in which the Tetra R motif directly contacts DNA was roughly divided into three types of interactions. The first was as a part of the α-helix that binds deeply to the main grooves of DNA, the second was as a part of the extended strand that binds to the minor grooves and the third was as a part of α-helix riding on nucleic acid phosphate backbones. Taking these results together, the tetra-R motif appeared to be a utility player in the interaction with nucleic acids. According to these results and

also based on the known affinity of basic amino acids to DNA (Zhang et al,2019), importin α2 may interact with DNA via the basic amino acid rich portion within the IBB domain.

Then we aligned the importin α family IBB domain to check for conservation of the basic amino acid in the IBB domain (Fig. 3D, E). While some variations in amino acid sequence were seen, three clusters of basic amino acids including the tetra-R motif were essentially conserved among the importin α family proteins. The differences were as follows. The first cluster 28RRRR in importin α2 was common in importin α1 and importin α6, while other importin α proteins showed exchanges in one or two amino acid with R at both ends. The second cluster (39RKAKK in importin α2) also differed, as importin α1 and importin α6 shared a common RKQKR sequence and importin α3 and importin α4 shared RKNKR, whereas importin α8 had RKTRK. The third cluster (49KRR in importin α2) was mostly conserved except importin α4 had KKR. This indicates that importin α family proteins commonly share the composition of the basic clusters within the IBB domain while possessing small differences.

Thus, we focused on three basic amino acid clusters, 28RRRR, 39RKAKK and 49RRR within the S24-R51 of IBB domain and designed amino acid substitution mutants to clarify the importin α2 DNA binding domain. All the R and K amino acids within each cluster were exchanged to A, as shown in Fig 4A, to create the 28A4, 39A5, 49A3 mutants, respectively. The three IBB domain mutant importin α2 proteins retained their nuclear transport activity for either of importin α/β dependent cargo proteins, SV40-NLS or Oct 3/4 in the *in vitro* transport assay with importin β1 (Fig. EV7). The 28A4 mutant showed weakened accumulation of transport cargo for SV40-NLS but enhanced accumulation for Oct 3/4, and 39A5 mutation showed a weakened transport efficiency for either of the tested cargos. Note that the similar mutants were tested reportedly in yeast importin α1 to measure their importin β binding efficiency (Harreman et al,2003), where either the mutants corresponding to 28A4 or 39A5 reduced the binding to importin β and 49A3 contrarily enhanced the binding. Although this could also hold true for mouse importin α2, our transport assay shows that the mutants were still functional to import cargo proteins together with importin β with differential affinity depending on the cargo types.

The R and K to A substitution mutant recombinant proteins were also used in gel shift analysis to test the binding affinity against the same Oct3/4 gene DNA sequences tested for the wild type importin α2. The 28A4 mutant had almost undetectable for binding to the DNA fragment, whereas the 39A5, 49A3 mutants showed reduced binding for the DNA fragments and retained a weak affinity for both the upstream-1 and upstream-2 600 bp Oct3/4 gene sequences (Fig. 5B, C). As expected, the basic amino acids within importin α2 IBB domain functioned in DNA binding and the 28RRRR cluster was essential.

The conserved basic amino acid clusters within the IBB domain were slightly different between each importin α families. This indicates the possibility that importin α proteins

commonly possess DNA-associating activity similar to importin α2, with some specificities in binding modes due to the small differences within the associating motif. In the homology search for IBB domain, RNA-protein complexes were presented as homology proteins against IBB domain. Hence, there is possibility that IBB domain binds not only to DNA but to RNA as well.

4. importin α DNA associating model

We next attempted to determine how importin α2 binds to DNA. The amino acids 13L–16F and 23S–48K of the IBB domain are located in the helical state in the crystal structure of the complex with importin β (PDB ID: 1QGK) (Fig. 4A, D). By contrast, the IBB domain of importin α has been reported as a missing part in many crystal structures (for example, PDB ID: 5TBK; 5HUW; 5HUY; 5V5P; 5V5O; 5W4E), except in the complex with importin β. This suggests that the IBB domain adapts multi-conformational states or disordered conformation when it is not bound to importin β. Therefore, we predicted the IBB domain intrinsic three-dimensional structure by analysing the amino acid sequence through the web server SCRATCH, which is constructed based on machine learning methods (Cheng et al, 2005). This analysis revealed that the central part (24S–51R) of the IBB domain tends to easily form an α-helix structure, at least after being induced by other molecules (Fig. 4E). Accordingly, the results of Circular Dichroism (CD) measurements of the KPNA2_1-69 peptide showed that the KPNA2_1-69 peptide has an intrinsic helix-rich structure (the estimated helix content was approximately 33.7%) and addition of and a double-stranded DNA (upstream-1 600bp) slightly increased the helix content to 40.5% (Fig. 4F, G and EV8-10). These values corresponded to those of secondary structure prediction, suggesting the possibilities that the basic amino acid clusters within IBB domain interacts with DNA in α-helix conformation as is in the crystal structure with importin β (Kobe,1999).

We next investigated the binding pattern of IBB domain to DNA by docking and MD simulations using the α-helical structure of IBB in the complex with importin β (PDB ID: 1QGK). To predict the binding feature of importin α2 and DNA, we explored the IBB domain-DNA complex and α-helix-DNA interaction at atomic level by constructing a model structure of the IBB domain α-helix-DNA complex using AutoDock vina (Trott & Olsen, 2010) and performing a 30 ns MD simulation of the docked model structure to optimise the structures (see supporting information for the method for the computational experiment). The energetic and conformational analyses of the MD trajectory revealed two α-helix/DNA binding modes (mode A and B) (Fig. 5A-D and Fig. EV11-15). The IBB domain N-terminal α-helix was fitted into a major groove of canonical B-form DNA strands for the mode A (Fig. 5A, B), while it was placed on DNA in parallel for the mode B (Fig. 5C, D). These binding modes correspond to the two binding manners in α-helical states out of three tetra-R motif binding types mentioned in the BLAST search results in section 3. The electrostatic interactions between negatively charged phosphate groups in DNA

and positively charged patches (28RRRR, 39RKAKK and 49KRR) on the surface of the α-helix were important in both modes, and the α-helix did not directly read specific bases in the DNA. Interestingly, the α-helix moved on the DNA from state A to B during our short time-scale simulation.

We assessed the validity of our model structures (mode A and B) by introducing mutations in the positive electrostatic patches on the α-helix surface for both mode A and B structures and repeating the experiment (28A4, 39A5 and 49A3). We calculated the binding free energies between DNA and these variants and compared the calculated binding affinities with our experimental results (see supporting information for details of the methodology). The calculated binding free energies indicated a weakening of the binding affinities for these variants as evident by the disappearance of a part of the positive electrostatic patch on the α-helix surface (Table 1).

Furthermore, to elucidate the binding strength and stoichiometry of importin α to DNAs, we performed one-dimensional NMR titration experiments of chemically synthesized 15bp double-stranded DNA (GCA GAT GCA TAA CCG) which is the core sequence of SOX-POU binding region in upstream-1 region including POU5F1 distal enhancer with chemically synthesized peptide (KPNA2_21-50, the core part of the arginine rich region in mouse importin α2) (Fig. 5E, F and EV16, 17). The chemical shift changes in imino proton region with addition of peptide is shown in Fig 5E. We also examined the interaction between KPNA2_21-50 and a randomized sequence DNA duplex (GCG GAC CAC TAG ACG) which has the same length and base composition as SOX-POU core sequence (Fig. 5F). As the results of the NMR titration experiments, several peaks in the DNA imino proton region of 1D spectra was apparently perturbed by the addition of peptide. Since these spectral changes seemed to show slightly different properties up to about 100 μM and over 150 μM, therefore, we assumed two-step changes and analyzed the spectra of the entire imino proton region by non-linear least square fitting to obtain the dissociation constant and stoichiometry for the binding.

The KPNA2_21-50 peptide showed two-step (mode) binding with these DNAs (Fig. EV18, 19). The estimated binding strengths were $K_{d1} = 1.6 \times 10^{-7}$ and $K_{d2} = 1.2 \times 10^{-5}$ with respect to the SOX-POU sequence, and the stoichiometry was about 1: 2 (DNA: peptide) for both binding steps. For the randomized sequence DNA duplex, the estimated binding strengths were $K_{d1} = 2.1\text{x}10^{-6}$ and $K_{d2} = 1.1\text{x}10^{-5}$, and the stoichiometry (DNA: peptide) was about 1: 1.6 for both binding steps. In the NMR titration experiments, several peaks in the DNA imino proton region of 1D spectra were apparently perturbed by the addition of peptide. The changes in the chemical shifts without remarkable resonance line widths broadening indicate that binding interactions occur in the fast exchange regime on the NMR chemical shift timescale, which in turn suggests that the binding dissociation constant, K_d, would have to fall in the 10^{-6} M range or more (Williamson,2013). This held true for both the weak bindings to randomised sequence, suggesting the binding was almost

diffusion limited, whereas the estimated dissociation constant for SOX2/POU was out of this range in the case of relatively strong binding mode, because the estimated dissociation constant for SOX2-POU DNA by fitting of the NMR titration data was slightly higher than that for the random sequence DNA. This implies that the on-rate for binding to SOX2-POU was significantly facilitated and that despite the fact that the apparent overall binding seemed to be relatively strong, each microscopic steps of the binding event had a short half-life.

The finding of multiple binding modes with similar binding energies in computational docking model analyses has been interpreted as experimental confirmation that these multi-conformations correspond to actual specific or not-specific binding modes (Kozakov et al,2014). This is consistent with the framework of a dimension-reduced facilitated diffusion model for the search for promoters of DNA-binding proteins on chromatin (Hannon et al,1986). This physicochemical model indicates that the DNA binding proteins randomly bind to multiple sites on the DNA and diffuse by strolling or hopping along the DNA chain until they reach their specific functional sites. This reduction in the dimensionality (3D to 1D) of the search for the target enhances the search efficiency and, in fact, DNA-binding proteins are known to reach target sites at a rate that is two orders of magnitude larger than the value estimated from three-dimensional diffusion (Riggs et al,1970). This mechanism necessarily requires both non-selective weak and specific binding properties for the DNA-binding protein, and the energy variation among multi-conformational modes of the protein–DNA binding must be within the range of $\sigma < 2kBT$ to achieve reasonable diffusion and association rates (Slutsky& Mirny,2004). The sliding rate depends on the degree of positive charge clustering of the specific binding region of the nucleic acid binding protein (Vuzman&Levy Y,2012). Therefore, the basic sequence of the IBB domain is expected to provide a favourable physicochemical propensity for strolling.

Importin α IBB domain contains conserved basic amino acid site at N-terminal short helix (13-22 of in importin α2) besides 28RRRR, 39RKAKK and 49RRR (Fig. 5D). Although those basic amino acid site were not applied to mutation analysis in this study, basic amino acid site at N-terminal short helix touched DNA when IBB domain binds to DNA either at mode A and mode B or when it changes the state from mode A to mode B in MD simulation. Here, we designated the DNA binding domain of importin α, the four conserved basic sites within IBB domain, as the "Nucleic Acid Associating Trolley pole" (NAAT) domain. Note that the NAAT domain do not include or locate closely to the NLS binding sites, importin α is capable of binding DNA and cargo protein simultaneously.

5. Relationship between the DNA binding activity and the nuclear transport activity of importin α

Since importin α plays a major role in nuclear transport of cargo proteins, elucidation of the

relationship between the transport activity and the DNA binding activity is important in considering its physiological significance in cells.

To clarify this issue, we assessed intracellular localization of GFP-fused mutants of importin α2 in ES cells (Fig. 6). We focused on the localization of the mutant 28A4, because it is the one that clearly lost DNA binding ability in the gel shift assay, (Fig. 4B, C and Fig. 5F, G), while it retaining its transport activity (Fig. EV7). We also used ED-28A4 mutant that lacked both the DNA binding and NLS binding activities.

In our experimental condition using ES cells, the GFP fused importin α2 28A4 showed cytoplasmic localization with weak signals in the nucleus (Figure 6A~E), in contrast to the wild type that was localized mainly in the nucleus and partially in the cytoplasm (Fig. 4A). Since the mutation in 28RRRR retained transport activity as shown by our transport assays (Fig. EV67), importin α2 28A4 is indicated to shuttle through the nuclear pore without retaining in the nucleus. Moreover, it has been reported that importin α can migrate into the nucleus without interaction with importin β in yeast and HeLa cells (Miyamoto et al,2002). Thus, even in the condition of reduced affinity to importin β due to this mutation (Harreman et al,2003), the free importin α2 28A4 is capable of entering into the nucleus. Taken together, our results indicate that nuclear retention via NAAT domain has essential role for the nuclear specific distribution.

The GFP-fusion of cargo transport deficient ED mutant (Gruss et al,2001) showed nuclear localization (Fig. 6B), whereas fusion proteins of ED-28A4 mutant that has neither the DNA binding and NLS binding activities showed a similar distribution in the cytoplasm to that shown by 28A4 mutant, demonstrating that the nuclear localization of the ED mutant was abrogated by the additional 28A4 mutation (Fig. 6D). Combined with these results, it can be concluded that the nuclear distribution of the ED mutant depended predominantly on the DNA binding activity of importin α2 via the basic amino acid cluster on the NAAT domain and that the chromatin binding are essentially independent of cargo transport activity, and vice versa.

Intriguingly, the nuclear localization was more intense for ED mutant than that of wild type in the ES cells (Fig. 6B), as described in a previous study using HeLa cells (Yasuda et al, 2012). This can be interpreted in terms of antagonistic action of IBB autoinhibitory mechanism on the chromatin binding. The importin α IBB domain, which includes the NAAT domain, masks its own major and minor NLS binding sites when it is free of cargo and importin β binding (PDB-1IAL, (Kobe,1999), PDB-4B8J, (Chang et al,2012), and Fig EV18). Considering the three dimensional structure, this autoinhibitory binding requires NLS-like basic amino acids in the IBB Domain, including the 49KRR (corresponding to mouse importin α2 structure PDB: 1IAL) in the NAAT domain. The ED mutant was designed so as to impair its NLS binding ability by double inversion of charged residues at their NLS binding sites, D192K/E396R in mouse importin α2 (Yasuda et al, 2012). From the structural view of the autoinhibitory state of IBB (PDB-1IAL,

Kobe,1999, PDB-4B8J, Chang et al,2012. and Fig EV 18). Considering the three dimensional structure, this autoinhibitory binding requires NLS-like basic amino acids in the IBB Domain, including the 49KRR (corresponding to mouse importin α2 structure PDB: 1IAL) in the NAAT domain. The ED mutant was designed so as to impair its NLS binding ability by double inversion of charged residues at their NLS binding sites, D192K/E396R in mouse importin α2 (Yasuda et al, 2012). From the structural view of the autoinhibitory state of IBB (PDB-1IAL, Kobe,1999, PDB-4B8J, Chang et al,2012 and Fig. EV18), these mutations should also destabilise the IBB binding to NLS binding sites in the autoinhibitory state because of the electrostatic repulsion between the cluster of positively charged amino acids of the NAAT domain and the arginine and lysine introduced at the NLS binding sites in ED mutant. Consequently, in the ED mutant, the IBB/NAAT domain may easily stick out of the main body of importin α detaching from the NLS binding pocket and interact with chromatin DNA. In this context, it is natural that the nuclear localisation of ED mutant was significantly enhanced compared to that of wild type by the enhanced interaction with chromatin DNA through the NAAT domain.

6. Importin α2 interacts with genomic DNA with weak sequence specificity

From the results of Blast search and MD simulation, it was assumed that NAAT domain does not seem to possess distinct sequence specificities, but rather has the property of interacting with various regions on genomic DNA. To confirm this hypothesis, we prepared mixtures of genomic DNA fragments that has a size of approximately 600bp or 900bp by digestion of chromatin from undifferentiated ES cells (Fig. EV19) and assayed whether importin α2 binds to these DNA mixtures by gel shift assays (Fig. 7A, B). In the electrophoretic mobility gel shift assays, the whole fraction of both 600bp and 900bp fragments were completely shifted by addition of wild type importin α2, indicating that importin α2 has affinity for almost all sequences of genomic DNA. The ED mutant also bound both size of DNA fraction, while the DNA binding deficient mutants showed only a weak interaction. Taken together, we concluded that importin α2 has potential to bind broad region of genomic DNA via the NAAT domain (Fig. 8).

Discussion

Importin α is known to associate DNase sensitive contents in the nucleus and the nuclear accumulation is accelerated under stress conditions such as heat shock or oxidative stress (Furuta et al, 2004; Miyamoto et al, 2004). Although the trigger controlling this process under heat shock conditions is not yet clear, it is proposed that a mutual control mechanism exists in which the normal protein transport pathway is suppressed for the urgent transport pathways that facilitate the specific import of heat shock proteins (Miyamoto et al, 2004; Furuta et al, 2004; Yasuda et al, 2012), such as Hsp70 via Hikeshi (Kose et al, 2012). Quite recently, the mechanism of importin

α nuclear retention under heat shock condition was partially revealed as a temperature-dependent modification of importin α protein (Ogawa& Imamoto , 2018). It was suggested that dysfunction of importin α as a transport receptor was the dominant determinant of the NLS transport suppression (Ogawa& Imamoto , 2018). Taken together with our observation that the reduction of NLS binding ability couples the enhancement of nuclear localisation, the shift of nuclear transport pathway could be brought about by increase of nuclear retention of importin α via NAAT domain that is consequence of the reduction of NLS binding activity by the heat-induced modification of importin α.

It should also be noted that the relationship between importin α chromatin association and transcriptional regulation is an important aspect to further understand the role of importin α-chromatin interaction. It is known that importin α carries various cargos (Pumroy&Cingolani,2015) including Oct3/4 (Yasuhara et al, 2007; Li et al, 2008; Young et al, 2011;Yasuhara et al, 2013) that binds to the cis element SOX2-POU. In the present study, it was demonstrated that importin α actually bound to the upstream sequence of Oct3/4 gene by the results of the ChIP-qPCR assays, and it was also shown by the NMR titration experiment that importin α bound to SOX2-POU element with enhanced binding specificity compared to random sequence DNA. These results indicated that importin α not only carried transcriptional regulators, but also interacted directly with the target DNA sequences of the cargo proteins. Interestingly, it has been reported that nuclear retention of importin α coordinated HeLa cell fate through changes in gene expression (Yasuda et al, 2012). It might be plausible that the binding of importin α to chromatin may affect the structural changes and/or phase separation of chromatin through cargo proteins such as chromatin modification factors.

As all members of importin α family seem to have NAAT domain (Fig. 3D) and have slightly different amino acid sequences to each other, variation of the physiological effects of importin α chromatin association via NAAT domain is also of considerable interest. Expression patterns of importin α family members have marked differences in different cell types, which points to the roles in cell fate determination. For example, regulation of development and alterations in expression profile of different importin α subtypes during differentiation in ES cells, and during spermatogenesis (reviewed in Miyamoto et al,2012, Loveland et al,2015; Yasuhara&Yoneda,2017), etc. have been studied so far. Furthermore, different importin αs have different cargo specificities. Therefore, it is natural to expect that the interwoven variations in these factors could enable rigorous regulations.

Conversely, it is fully conceivable that the interaction between importin α and chromatin is modified by the presence of cargo proteins such as transcriptional regulatory factors and that the modifications may be responsible for the regulation of cellular physiological functions. If the chromatin association is influenced by its cargos, the composite effects of the variation of

importin α types and of the cargos may be related to general transcription regulations and a variety of physiological functions. These problems will have to be reckon with in detail in future.

So far, there are reports on many intracellular functions of importin α by virtue of its binding to various proteins. For example, Nup50 dissociates NLS from importin α (Matsuura&Stewart,2005), RAN (Lee et al,2005) and RBBP4 dissociate importin β from α (Tsujii et al,2015), and CAS interacts with importin α to export it (Herold et al,1998;Tsujii et al,2015). Therefore, to understand the physiological role of the association of importin α with chromatin via DNA binding ability which was unveiled in this study, the balance among the above functions will play an important role in regulatory mechanisms of physiological condition that determine the cell fate.

As revealed in the present study, chromatin binding of importin α showed two distinct properties: first, it can interact with the vast majority of the genomic sequences, and second, it has moderate sequence selectivity probably to stay at specific sites (Fig. 2 and 5E, F).

Concerning the first point, the characteristics of the binding ability of NAAT domain could be an advantage for efficient delivery of itself or cargos to various sites of chromatin depending on the situation. Some of DNA binding proteins have been known to be able to locate their target sites amid myriad off-target sequences within millions to billions base pairs at remarkably rapid rates(Leven& Levy, 2019), and various theoretical and experimental studies have been conducted in order to characterize the mechanism of the rapid target site search(Berg et al,1981; Halford&Marko,2004; Sokolov et al,2005; Hu et al,2006; Wunderlich&Mirny,2008; Halford,2009;Kolomeisky,2011;Bauer&Metzler,2012;Brackley et al,2013; Normanno et al,2015 Shvets et al,2018).

As a mechanism of such a rapid target search, the facilitated diffusion (Berg et al,1981) is one of the predominant models. Although the concept of facilitated diffusion still has been the subject of controversies Halford,2009; Normanno,2015), the model has an increasing body of supporting evidence. (Normanno,2015)

Along with the facilitated diffusion model, the target site location on long DNAs by DNA-binding proteins is thought to consist of an initial collision with a random non-specific site and, probably, consequent 1D-3D combined search which is combination of three types of movements so-called "sliding", "hopping", and "jumping", and the balance of these behaviours seems to determine the efficiency of the target search (Normanno,2015). In the process, it is considered to be preferable that a protein (or a protein complex) has at least two DNA-binding surfaces or more to perform an intersegmental transfer from a DNA site to a relatively distant anther site. (Halford,2009) This property was exactly shown in the exchange of binding modes demonstrated by MD simulation for NAAT domain binding to DNA. In addition, theoretical consideration suggests that electrostatic interaction may play an important role in this facilitated diffusion

(Halford,2009;Leven& Levy, 2019) and for the facilitated diffusion nearby off-target DNA sequences, a large positively charged patch on a protein surface is essential.(Leven& Levy, 2019). Having four positively charged patches covering over a half of the entire surface, the property of the NAAT domain exactly satisfies these requirements necessary for the facilitated diffusion. Analysis of NMR titration also suggested that the process consists of relatively accelerated association and rapid dissociation as is described in section 4.

Combined with the fact that NAAT domain interacted with a wide range of sequences shown in the present study (Fig. 5E, F and 7), the association of importin α with chromatin is suggested to obey the principle of facilitated diffusion.

Regarding the second point which is concerning the binding specificity, it is also pointed out that there is a trade-off between the search step efficiency and the recognition step that usually exhibits a relatively strong and specific binding (Leven& Levy, 2019). The results of the ChIP-qPCR assay and NMR titration experiments (Fig. 2 and 5E, F) apparently showed the sequence dependent preference for the binding. Analysis of NMR titration also suggested that the process consists of relatively accelerated association and rapid dissociation. In addition, being transferring cargos including many DNA binding proteins, the sequence specificity of importin α binding to DNA could be augmented by the cargos. All these findings indicates that the binding property of NAAT to DNA is considered eligible for both searching and recognising the target on the huge size of sequence space.

Summing up, we propose that "stroll and locate" motion of importin α associating with chromatin via NAAT domain (Fig. 8) would be related to a wide variety of cellular physiological processes.

Materials and methods

Cell culture

Mouse ES cell lines were cultured as follows. Bruce 4 cell line were maintained in DMEM supplemented with 15% FCS and ESGRO (Millipore) on 0.1% gelatin coated dish at 10% CO_2.

Immunostaining

For immunostaining, cells were seeded on 0.1% gelatin coated cover slip, and then fixed in 3.7% formalin (Nacalai Tesque Inc.) in PBS. Cells were permeabilized using 0.5% Triton X100 in PBS (Nacalai Tesque Inc.), and blocked in 3% Skim milk in PBS (Nacalai Tesque Inc.). The 1st antibody for KPNA2 (importin α2) (rat polyclonal, MBL) and the second antibody Alexa488 conjugated anti-rat IgG (Thermo Fisher Scientific) were suspended in Can get signal (Toyobo), in concentrations according to the manufacturer's instructions. DNA were stained with DAPI

(Nacalai Tesque Inc.) or TO-PRO3-stain (Thermo Fisher Scientific), and images were obtained by conforcal microscopy A1 (Nikon) or LSM510 (ZEISS).

Plasmid construction and transfection

The full length and mutant pEGFP constructs of importin αs were cloned as previously described (Yasuda et al., EMBO 2012). pEGFP- importin α2 28A4, 39A5, 49A and importin α2 ED28A4, 39A5, 49A were constructed using KOD-plus-Mutagenesis Kit (Toyobo, Tokyo, Japan). Primers used were as follows.

28A4 Fwd-GCTGCTATAGAAGTTAATGTGGAACTCAGGAAA,

28A4 Rev-AGCAGCCATTTCTGTGCTGTCCTTCCC,

39A5 Fwd-AGCGGCAGCGAGTTCCACATTAAC,

39A5 Rev-GCTGCCGATGAGCAGATGCTG,

49A4 Fwd-AACGTCAGCTCCTTTCCTGATGAT,

49A4 Rev-AGCAGCAGCCAGCATCTGCTCATCTT,

pEGFP vectors were transfected into ES cells using lipofectamine 3000 (Thermo Fisher Scientific) following the manufacturer's protocol, and pictures were obtained using confocal microscope (ZEISS International). Line plots were performed by ImageJ (W. Rasband; http://rsb.info.nih.gov/ij/).

GST- importin α2 28A4,ED28A4 were generated by inserting the BamHI-EcoRI PCR fragments into the BamHI–EcoRI sites of the Escherichia coli N-terminal GSTfused protein expression vector pGEX-6P-2 (GE Healthcare). Recombinant proteins were obtained as described in the supplemental information (2.1.).

The upstream fragments of mouse POU5F1 were obtained as follows. The target fragments were obtained from Bruce4 cell genome DNA with targeting primers,

EcoRV upstream-1 Fwd-TACGATATCCACATCTGTTTCAAGCTAGTTCTA

EcoRV upstream-1-1 Rev- TACGATATCTGAATCTTCCGTTTCCTCC

EcoRV upstream-2 Fwd-TACGATATCGAGAATTATCAGGAGTTCAAGG

EcoRV upstream-2-1 Rev-TACGATATCACTTCCTGCTCCCCA

and the fragments were inserted into pBlueScript vector, followed by PCR using primer sets, upstream-1 Fwd / upstream-1-2 Rev, upstream-2 Fwd / upstream-2-2 Rev (sequences are described below to amplify for gel shift assays).

Chromatin immunoprecipitation

Chromatin immunoprecipitation assays were performed as previously described (Semba et al., 2017) with some modifications. Culture of 1.0×10^7 ES cells were crosslinked in 0.5% formaldehyde containing media for 5 minutes followed by termination of reaction with 0.125M

Glycin. Cell pellets were then suspended in RIPA buffer (50 mM Tris-HCl pH8.0, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 0.5% Sodium Deoxychorate and protease inhibitor cocktail) and were sonicated for 5 seconds 20 times at 40% maximum amplitude in level 2 with SONIFIER 250 (Branson) for ChIP-qPCR. The cell lysates were then diluted to 100ng/µl DNA concentration after centrifugation at 15000g, 10 minutes and were applied to IP using specific antibodies. The antibodies were anti-KPNA2 (importin α2) (rat monoclonal, originally produced as previously described) or normal rat IgG (Jackson) as control. The cushion antibody to bind rat IgG to Dynabeads M-280 anti-Rabbit IgG (Thermo Fisher Scientific) was anti-rat rabbit IgG (Jackson). The IP solutions were incubated at 4°C overnight with rotation, and the beads were washed with ChIP buffer (10 mM Tris-HCl pH8.0, 200 mM KCl, 1 mM CaCl2, 0.5% NP-40 and protease inhibitor cocktail), wash buffer (10 mM Tris-HCl pH8.0, 500 mM KCl, 1 mM CaCl2, 0.5% NP-40 and protease inhibitor cocktail) and TE buffer. Reverse crosslinking was achieved by mixing the beads with ChIP Elution buffer (50 mM Tris-HCl pH8.0, 10 mM EDTA, 1% SDS), followed by incubating with 2% proteinase K (Nacalai Tesque Inc.) at 50°C, 1hr. Obtained DNA was purified using Wizard SV Gel and PCR Celan up system (Promega).

Quantitative PCR

Quantitative PCR assays after ChIP were performed using THUNDERBIRD SYBR qPCR Mix (TOYOBO), and Thermal Cycler Dice Real Time System II (TAKARA BIO) following to the manufactures protocols with primers below.

upstream-1 Fwd-CACATCTGTTTCAAGCTAGTTCTAAGAA

upstream-1-2 Rev-CAACCTTGTCTTATGGATTGTTCTCTT

upstream-2 Fwd-ATGAAGACTACCATCAAGAGACACC

upstream-2-2 Rev-TTGTCTGTCTGCTCCTACACCAT

Genomic DNA sharing

Bruce 4 cells at 60-70% confluent on gelatin coated 10 cm culture dish were lysed in ChIP Elution Buffer (50mM Tris-HCl pH8.0, 10mM EDTA, 1% SDS). Cell lysates were shared for 5 seconds 6 times at output level 4 with Handy Sonic (TOMY) followed by 7cycles of 30 seconds on and 60 seconds off at high level with Bioruptor II (BM Equipment). Then cell lysates were incubated more than 3 hours with 250mM NaCl and 0.25mg/ml Proteinase K (Nacalai Tesque Inc.). Obtained DNA was purified using FastGene Gel/PCR Extraction Kit (Nippon Genetics Co.,Ltd.) and applied to electrophoresis in 2% agarose gel as shown in figure EV6. The 600bp and 900bp DNA fragment fractions were cut from the agarose gel (cutting images shown in figure EV6) and were purified using FastGene Gel/PCR Extraction Kit (Nippon Genetics Co.,Ltd.).

Gel shift assay

The binding of protein and DNA fragments was detected by gel shift assay as follows. The DNA fragments were biotinylated using Biotin 3' End DNA Labeling Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The biotinylated DNA fragments were applied to gel shift assay using LightShift® Chemiluminescent EMSA Kit (Thermo Fisher Scientific) and Chemiluminescent Nucleic Acid Detection Module (Thermo Fisher Scientific) according to the manufacturer's instructions. 17.24 pmol or 34.48 pmol recombinant proteins (see also Fig. EV4) and 1.57fmol DNA fragments were mixed to allow binding. Protein/DNA solution was applied to electrophoresis at 200V constant pressure current in 4% polyacrylamide gel. The DNA fragments were then transferred to the nylon membrane and was combined with the Streptavidin-Horseradish Peroxidase Conjugate. The Streptavidin-Horseradish Peroxidase Conjugate was detected using Pxi4 (Syngene) or G:BOX mini (Syngene).

Homology modeling of N-terminal domain of importin α2 proteins

Homology modeling of importin α2 IBB domain was performed using SWISS-MODEL with default parameters. Amino acid sequences of wild type and mutated IBB domain of mouse importin α2 were applied as target sequence, and PDB ID: 1QGK was selected as a template.

CD spectroscopy.

Far-ultraviolet CD spectra were collected on a Jasco J-820 spectropolarimeter at room temperature with 50 mM importin α in 20mM phosphate buffer at pH 7.0. using a 0.05-cm path-length cuvette. Sixty-four scans were averaged and the blank spectrum was subtracted from the sample spectra to calculate ellipticity.

NMR measurement

NMR spectra were acquired on a Bruker Avance II 800 MHz spectrometer equipped with a cryogenically cooled proton optimized triple resonance NMR 'inverse' probe (TCI) (Bruker Biospin, Germany). All spectra were acquired at 25ºC (298 K) with 256 scans using the p3919gp pulse program. Acquisition and spectrum processing were performed using Topspin3.2TM software. Chemically synthesized peptide and DNAs were dissolved in 50mM Tris-Cl buffer solution containing 10% D2O, pH 7.9 and used for the measurements in 4mmϕ Shigemi tubes. The amino acid sequence of the chemically synthesized peptide was KDSTEMRRRRIEVNVELRKAKKDEQMLKRR (Kpna2_21-50), and the sequences of 15 bp of DNAs were GCA GAT GCA TAA CCG (OCT-SOX2) and its randomized control sequence GCG GAC CAC TAG ACG.

Acknowledgments

This work was supported by the JSPS KAKENHI to Noriko Yasuhara (18H04870, 15K07069, 25116008) and Noriko Saitoh (18H05531, 18K19310) and by a Nihon University Multidisciplinary Research Grant for 2018.

We thank Drs. Naoki Horikoshi, Saki Hirata, Yuichiro Semba, S.J. Nogami, Hiroyuki Taguchi, Rashid Mehmood, Hitoshi Kurumizaka, Yasuyuki Ohkawa and Hiroshi Kimura for kind suggestions and discussions about this work.

Author contributions

Experiments were conducted as follows. Cell imaging: KJ, NS, NS, NY. ChIP-qPCR: SS, RK, NY. Mutant construction and analysis: KJ, SN, YK, SS, NS, NY. Homology modeling and Bioinformatics analysis: TK, KJ, NY. MD analysis: AS, MN. Circular Dichroism measurement and analysis:TK. NMR measurement and analysis: TK, KA. Conceptual input and supervision: KJ, TK, AS, NS, YY, NY. Project design and writing: KJ, TK, AS, NY.

Reference

Bauer, M. and R. Metzler (2012). Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys J* 102(10): 2321-2330.

Berg, O. G., R. B. Winter and P. H. von Hippel (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20(24): 6929-6948.

Brackley, C. A., M. E. Cates and D. Marenduzzo (2013). Intracellular facilitated diffusion: searchers, crowders, and blockers. *Phys Rev Lett* 111(10): 108101.

Chang, C. W., R. L. Couñago, S. J. Williams, M. Bodén and B. Kobe (2012). Crystal structure of rice importin-α and structural basis of its interaction with plant-specific nuclear localization signals. *Plant Cell* 24(12): 5074-5088.

Cheng, J., A. Z. Randall, M. J. Sweredoski and P. Baldi (2005). SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res* 33(Web Server issue): W72-76.

Furuta, M., S. Kose, M. Koike, T. Shimi, Y. Hiraoka, Y. Yoneda, T. Haraguchi and N. Imamoto (2004). Heat-shock induced nuclear retention and recycling inhibition of importin alpha. *Genes Cells* 9(5): 429-441.

Goldfarb, D. S., A. H. Corbett, D. A. Mason, M. T. Harreman and S. A. Adam (2004). Importin alpha: a multipurpose nuclear-transport receptor. *Trends Cell Biol* 14(9): 505-514.

Gruss, O. J., R. E. Carazo-Salas, C. A. Schatz, G. Guarguaglini, J. Kast, M. Wilm, N. Le Bot, I. Vernos, E. Karsenti and I. W. Mattaj (2001). Ran induces spindle assembly by reversing the inhibitory effect of importin alpha on TPX2 activity. *Cell* 104(1): 83-93.

Görlich, D. and I. W. Mattaj (1996). Nucleocytoplasmic transport. *Science* 271(5255): 1513-1518.

Halford, S. E. (2009). An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem Soc Trans* 37(Pt 2): 343-348.

Halford, S. E. and J. F. Marko (2004). How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* 32(10): 3040-3052.

Hannon, R., E. G. Richards and H. J. Gould (1986). Facilitated diffusion of a DNA binding protein on chromatin. *EMBO J* 5(12): 3313-3319.

Harreman, M. T., M. R. Hodel, P. Fanara, A. E. Hodel and A. H. Corbett (2003). The auto-inhibitory function of importin alpha is essential in vivo. *J Biol Chem* 278(8): 5854-5863.

Herold, A., R. Truant, H. Wiegand and B. R. Cullen (1998). Determination of the functional domain organization of the importin alpha nuclear import factor. *J Cell Biol* 143(2): 309-318.

Hu, J., F. Wang, Y. Yuan, X. Zhu, Y. Wang, Y. Zhang, Z. Kou, S. Wang and S. Gao (2010). Novel importin-alpha family member Kpna7 is required for normal fertility and fecundity in the mouse. *J Biol Chem* 285(43): 33113-33122.

Hu, T., A. Y. Grosberg and B. I. Shklovskii (2006). How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys J* 90(8): 2731-2744.

Imamoto, N., T. Shimamoto, T. Takao, T. Tachibana, S. Kose, M. Matsubae, T. Sekimoto, Y. Shimonishi and Y. Yoneda (1995). In vivo evidence for involvement of a 58 kDa component of nuclear pore-targeting complex in nuclear protein import. *EMBO J* 14(15): 3617-3626.

Kalderon, D., W. D. Richardson, A. F. Markham and A. E. Smith (1984). Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* 311(5981): 33-38.

Kobe, B. (1999). Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. *Nat Struct Biol* 6(4): 388-397.

Kodiha, M., A. Chu, N. Matusiewicz and U. Stochaj (2004). Multiple mechanisms promote the inhibition of classical nuclear import upon exposure to severe oxidative stress. *Cell Death Differ* 11(8): 862-874.

Kolomeisky, A. B. (2011). Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys Chem Chem Phys* 13(6): 2088-2095.

Kose, S., M. Furuta and N. Imamoto (2012). Hikeshi, a nuclear import carrier for Hsp70s, protects cells from heat shock-induced nuclear damage. *Cell* 149(3): 578-589.

Kozakov, D., K. Li, D. R. Hall, D. Beglov, J. Zheng, P. Vakili, O. Schueler-Furman, I. C. h. Paschalidis, G. M. Clore and S. Vajda (2014). Encounter complexes and dimensionality reduction in protein-protein association. *Elife* 3: e01370.

Kutay, U., F. R. Bischoff, S. Kostka, R. Kraft and D. Görlich (1997). Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor. *Cell* 90(6): 1061-1071.

Lange, A., R. E. Mills, C. J. Lange, M. Stewart, S. E. Devine and A. H. Corbett (2007). Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J Biol Chem* 282(8): 5101-5105.

Lee, S. J., Y. Matsuura, S. M. Liu and M. Stewart (2005). Structural basis for nuclear import complex dissociation by RanGTP. *Nature* 435(7042): 693-696.

Leven, I. and Y. Levy (2019). Quantifying the two-state facilitated diffusion model of protein-DNA interactions. *Nucleic Acids Res* 47(11): 5530-5538.

Li, X., L. Sun and Y. Jin (2008). Identification of karyopherin-alpha 2 as an Oct4 associated protein. *J Genet Genomics* 35(12): 723-728.

Lindsay, M. E., K. Plafker, A. E. Smith, B. E. Clurman and I. G. Macara (2002). Npap60/Nup50 is a tri-stable switch that stimulates importin-alpha:beta-mediated nuclear protein import. *Cell* 110(3): 349-360.

Lott, K. and G. Cingolani (2011). The importin β binding domain as a master regulator of nucleocytoplasmic transport. *Biochim Biophys Acta* 1813(9): 1578-1592.

Loveland, K. L., A. T. Major, R. Butler, J. C. Young, D. A. Jans and Y. Miyamoto (2015). Putting things in place for fertilization: discovering roles for importin proteins in cell fate and spermatogenesis. *Asian J Androl* 17(4): 537-544.

Matsuura, Y. and M. Stewart (2005). Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling. *EMBO J* 24(21): 3681-3689.

Mihalas, B. P., P. S. Western, K. L. Loveland, E. A. McLaughlin and J. E. Holt (2015). Changing expression and subcellular distribution of karyopherins during murine oogenesis. *Reproduction* 150(6): 485-496.

Miyamoto, Y., P. R. Boag, G. R. Hime and K. L. Loveland (2012). Regulated nucleocytoplasmic transport during gametogenesis. *Biochim Biophys Acta* 1819(6): 616-630.

Miyamoto, Y., M. Hieda, M. T. Harreman, M. Fukumoto, T. Saiwaki, A. E. Hodel, A. H. Corbett and Y. Yoneda (2002). Importin alpha can migrate into the nucleus in an importin beta- and Ran-independent manner. *EMBO J* 21(21): 5833-5842.

Miyamoto, Y., K. L. Loveland and Y. Yoneda (2012). Nuclear importin α and its physiological importance. *Commun Integr Biol* 5(2): 220-222.

Miyamoto, Y., T. Saiwaki, J. Yamashita, Y. Yasuda, I. Kotera, S. Shibata, M. Shigeta, Y. Hiraoka, T. Haraguchi and Y. Yoneda (2004). Cellular stresses induce the nuclear accumulation of importin alpha and cause a conventional nuclear import block. *J Cell Biol* 165(5): 617-623.

Miyamoto, Y., K. Yamada and Y. Yoneda (2016). Importin α: a key molecule in nuclear transport and non-transport functions. *J Biochem* 160(2): 69-75.

Niwa, H. (2007). How is pluripotency determined and maintained? *Development* 134(4): 635-646.

Niwa, H., J. Miyazaki and A. G. Smith (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 24(4): 372-376.

Normanno, D., L. Boudarène, C. Dugast-Darzacq, J. Chen, C. Richter, F. Proux, O. Bénichou, R. Voituriez, X. Darzacq and M. Dahan (2015). Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat Commun* 6: 7357.

Ogawa, Y. and N. Imamoto (2018). Nuclear transport adapts to varying heat stress in a multistep mechanism. *J Cell Biol* 217(7): 2341-2352.

Oka, M., T. Moriyama, M. Asally, K. Kawakami and Y. Yoneda (2013). Differential role for transcription factor Oct4 nucleocytoplasmic dynamics in somatic cell reprogramming and self-renewal of embryonic stem cells. *J Biol Chem* 288(21): 15085-15097.

Oka, M. and Y. Yoneda (2018). Importin α: functions as a nuclear transport factor and beyond. *Proc Jpn Acad Ser B Phys Biol Sci* 94(7): 259-274.

Pumroy, R. A. and G. Cingolani (2015). Diversification of importin-α isoforms in cellular trafficking and disease states. *Biochem J* 466(1): 13-28.

Riggs, A. D., S. Bourgeois and M. Cohn (1970). The lac repressor-operator interaction. 3. Kinetic studies. *J Mol Biol* 53(3): 401-417.

Semba, Y., A. Harada, K. Maehara, S. Oki, C. Meno, J. Ueda, K. Yamagata, A. Suzuki, M. Onimaru, J. Nogami, S. Okada, K. Akashi and Y. Ohkawa (2017). Chd2 regulates chromatin for proper gene expression toward differentiation in mouse embryonic stem cells. *Nucleic Acids Res* 45(15): 8758-8772.

Shvets, A. A., M. P. Kochugaeva and A. B. Kolomeisky (2018). Mechanisms of Protein Search for Targets on DNA: Theoretical Insights. *Molecules* 23(9).

Slutsky, M. and L. A. Mirny (2004). Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys J* 87(6): 4021-4035.

Sokolov, I. M., R. Metzler, K. Pant and M. C. Williams (2005). Target search of N sliding proteins on a DNA. *Biophys J* 89(2): 895-902.

Trott, O. and A. J. Olson (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2): 455-461.

Vuzman, D. and Y. Levy (2012). Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol Biosyst* 8(1): 47-57.

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* 73: 1-16.

Wunderlich, Z. and L. A. Mirny (2008). Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Res* 36(11): 3570-3578.

Yasuda, Y., Y. Miyamoto, T. Yamashiro, M. Asally, A. Masui, C. Wong, K. L. Loveland and Y. Yoneda (2012). Nuclear retention of importin α coordinates cell fate through changes in gene expression. *EMBO J* 31(1): 83-94.

Yasuhara, N., N. Shibazaki, S. Tanaka, M. Nagai, Y. Kamikawa, S. Oe, M. Asally, Y. Kamachi, H. Kondoh and Y. Yoneda (2007). Triggering neural differentiation of ES cells by subtype switching of importin-alpha. *Nat Cell Biol* 9(1): 72-79.

Yasuhara, N., R. Yamagishi, Y. Arai, R. Mehmood, C. Kimoto, T. Fujita, K. Touma, A. Kaneko, Y. Kamikawa, T. Moriyama, T. Yanagida, H. Kaneko and Y. Yoneda (2013). Importin alpha subtypes determine differential transcription factor localization in embryonic stem cells maintenance. *Dev Cell* 26(2): 123-135.

Yasuhara, N. and Y. Yoneda (2017). Importins in the maintenance and lineage commitment of ES cells. *Neurochem Int* 105: 32-41.

Young, J. C., A. T. Major, Y. Miyamoto, K. L. Loveland and D. A. Jans (2011). Distinct effects of importin α2 and α4 on Oct3/4 localization and expression in mouse embryonic stem cells. *FASEB J* 25(11): 3958-3965.

Zhang, J., Z. Ma and L. Kurgan (2019). Comprehensive review and empirical analysis of hallmarks of DNA-, RNA- and protein-binding residues in protein chains. *Brief Bioinform* 20(4): 1250-1268.

Figure legends

Figure 1.

(A) Endogenous importin α2 localisation was shown by immunofluorescence in undifferentiated ES cells. green / importin α2, blue / DNA.

(B, C) GFP and GFP-fused importin α2 were expressed in undifferentiated ES cells. blue / DNA.

Figure 2

Importin α2 binds the upstream region of the POU5F1 gene in undifferentiated mouse ES cells. ChIP-qPCR analysis with a specific antibody for importin α2 were performed to confirm the binding of importin α2 to the POU5F1 gene.

(A) The primer sets used in the experiments are shown.

(B) Results of ChIP-qPCR for primers POF5F1 upstream-1 and upstream -2. The relative CT value against a ChIP control sample prepared using normal IgG are listed as mean values of relative ratio for B with error bars in three independent experiments.

Figure 3

Importin α2 bound DNA *in vitro*.

(A) Full length and deleted recombinant importin α2 used in gel shift (*in vitro* binding) assays.

(B-C) The genomic DNA sequence from the importin α2 bound region in Oct3/4, determined by the ChIP-qPCR analysis in Fig. 2, were selected and applied to these *in vitro* binding assays. DNA sequences from the Oct3/4 upstream region (upstream-1 and -2) were located as described in Fig.2 A. Naked DNAs of upstream-1 (B) and upstream-2 (C) were used (also see Fig. S2).

(D) Conservation of basic amino acid sites of the importin α IBB domains. Amino acid sequences of mouse importin α (KPNA) families were aligned using ClustalW (DDBJ). The basic amino acid clusters without gaps are presented in red boxes and a cluster with gaps is presented in a brown box.

(E) Sequence logo of IBB domains created with WebLogo.

Figure 4

(A) The basic amino acid clusters in the IBB domain of importin α2 were mutated as indicated. The helical portion explained in (D) are indicated in box no.1 and 2.

(B, C) Recombinant wild type and mutant importin α proteins were applied to *in vitro* binding assays using the naked DNAs for upstream-1 (B) and upstream-2(C) regions of the Oct 3/4 gene. See also Fig. EV4 and EV5.

(D) The location of the basic clusters indicated on the helix structures of IBB in the complex with importin β (PDB ID: 1QGK). The box no.1 and 2 corresponds to those presented in (A).

(E) The amino acid sequence was analysed to predict the structural features through the web server SCRATCH.

(F) CD spectroscopy to predict helix content in IBB (mouse importin α2 KPNA 1-69 peptide) in the absence and presence of dsDNA (upstream-1 600bp). a) 6.6μM of peptide only, b) 6.6μ of

DNA, c) mixture of the peptide and DNA, d) difference spectrum obtained from c by subtracting b.

(G) The predicted secondary structure contents for the peptide in the absence and presence of the DNA. The upper panel shows the structures in the absence of the DNA and the lower shows the structures in the presence of the DNA. The unit of vertical axis is ellipticity described in mdeg.

Figure 5

(A-D) Two model structures of the importin α2 IBB domain/DNA complex were obtained from molecular docking and MD simulation: mode A (A and C) and mode B (B and D). DNAs are shown in the surface model and coloured according to atoms (carbon: grey, oxygen: red, nitrogen: blue, phosphorus: yellow) (C-F). The α-helices of importin α are shown in a ribbon model (A and B) or a surface model (C and D) and are coloured green (A-D). Three positively charged patches in the a-helix, 28RRRR, 39RKAKK and 49KRR, are shown in sticks coloured in magenta, cyan and orange, respectively. N-terminal basic residues that formed short helixes (13R, 16R, 18K, 20K and 22K) are also shown as sticks and coloured in yellow.

(E, F) ^{1}H-1D NMR titration experiments. The spectral changes in the imino proton region of the DNA induced by addition of indicated concentrations of peptide, Kpna2_21-50, the core part of the arginine rich region in mouse importin α2, are shown. (E) The spectral change for 50μM of the 15bp double-stranded DNA of SOX-POU binding region in upstream of POU5F1. (F) The change for 50μM of a 15bp double-stranded randomized sequence DNA. The peptide concentrations are indicated in each figure.

Figure 6

(A-E) The DNA binding-deficient mutant of importin α2 abrogated the nuclear localisation of the protein. GFP-fused importin α2 was expressed in undifferentiated ES cells. GFP-wild type importin α2 (A), importin α2 ED mutant (B), importin α2 28S4 mutant (C) and importin α2 ED-28A4 mutant (D) are shown with control GFP (E). Blue colour staining shows DNA. The distributions of GFP-fused proteins of two cells for each sample were measured by line plots.

Figure 7

(A, B) Importin α2 multiply bound genomic DNA of ES cells. Digested genomic DNA were purified from undifferentiated mouse ES cells and were applied to gel electrophoresis. Fractions either around 600bp or 900bp were purified. Gel cutting image is shown by white boxes (F). Recombinant wild type and mutant importin α proteins were applied to *in vitro* binding assays using the purified 600bp DNAs (G) and 900bp DNAs (H).

Figure 8

The importin α DNA associating model.

Importin α2 can possibly exist in an autoregulatory form (upper-left) or in a cargo binding form (upper-right) in the nucleus. In these forms, the Nucleic Acid Associating Trolley pole domain (NAAT domain) is in a free mode. When importin α2 approaches chromatin (with or without cargo or interacting proteins), the NAAT domain forms helix structures and associates with DNA predominantly with a tetra-R motif. Importin α2 associates with DNA in at least two modes and shifts between the modes by strolling on the DNA or hopping to neighbouring DNA. The schematic chart shows the NAAT domain and R-quartet motif of importin α2.

Tables

Table 1. Calculated binding free energies (ΔG_{bind}) between IBB domain and DNA in the MD simulations

ΔG_{bind} (kJ/mol)	Mode A[a]	Mode B[b]
Wild type	-526.05 ± 36.28	-602.96 ± 27.91
28A4	-296.06 ± 40.04	-431.71 ± 20.42
39A5	-342.54 ± 10.21	-426.85 ± 15.31
49A3	-432.00 ± 21.46	-508.82 ± 19.56

[a] Values for wild type was averaged over a 7-ns period ranging from 11 to 18 ns, those for variants were averaged over last 2 ns (8 ~ 10 ns).

[b] Values for wild type was averaged over a 5-ns period ranging from 25 to 30 ns, those for variants were averaged over last 2 ns (8 ~ 10 ns).

Figure 1

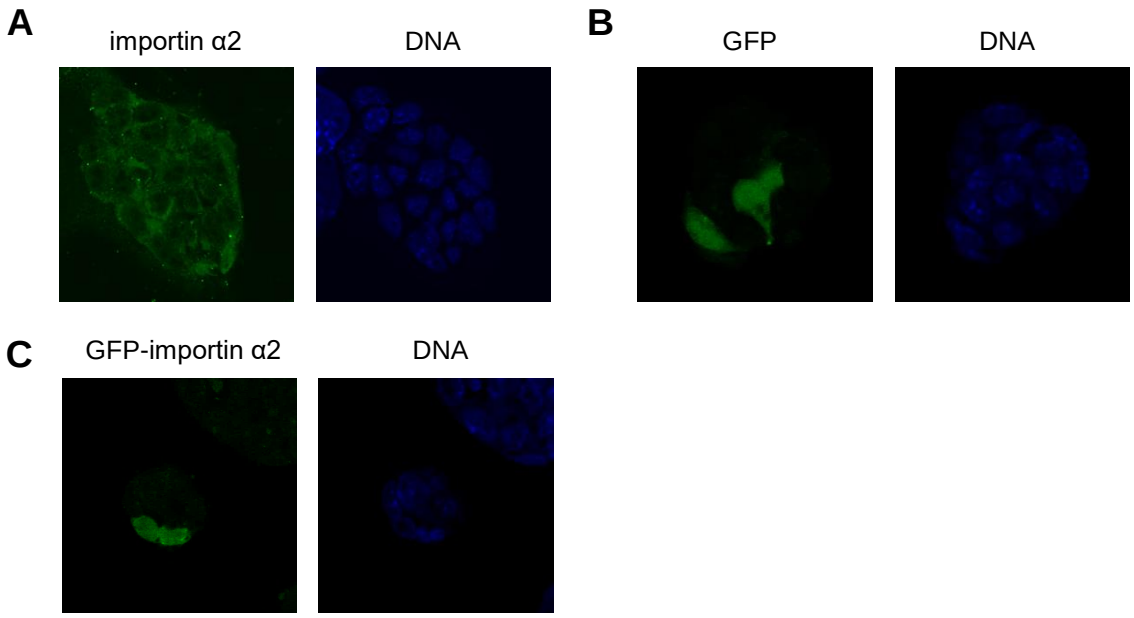


Figure 2

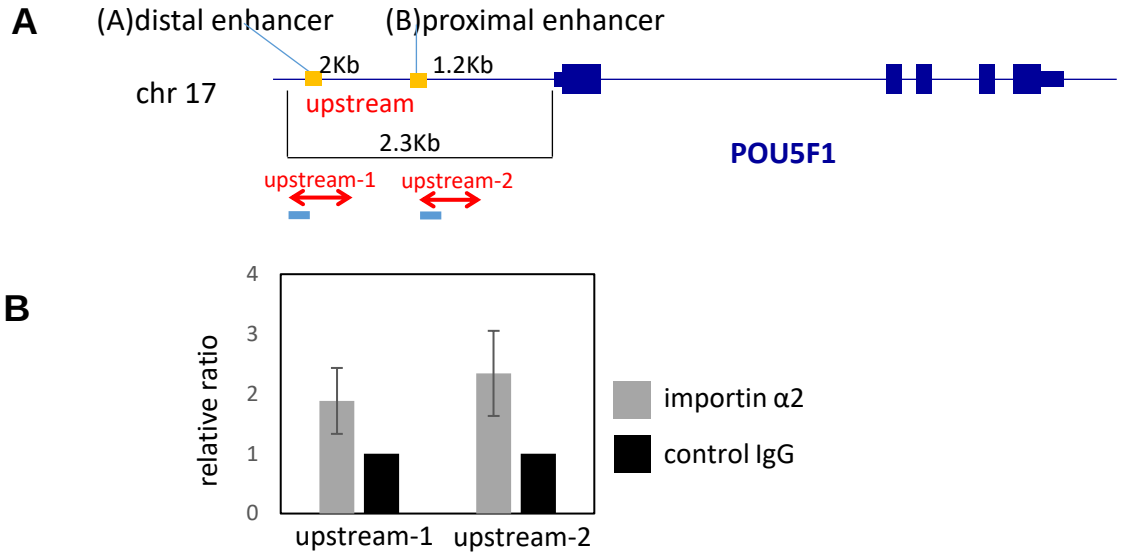
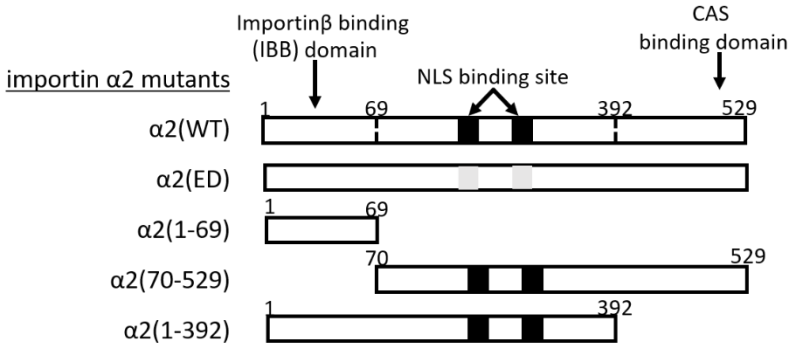
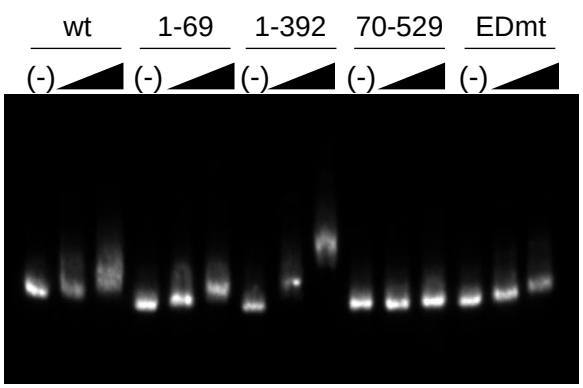

A
(A)distal enhancer
(B)proximal enhancer
2Kb
1.2Kb
chr 17
upstream
2.3Kb
POU5F1
upstream-1
upstream-2
B
relative ratio
4
3
2
1
0
upstream-1
upstream-2
importin α2
control IgG

Figure 3

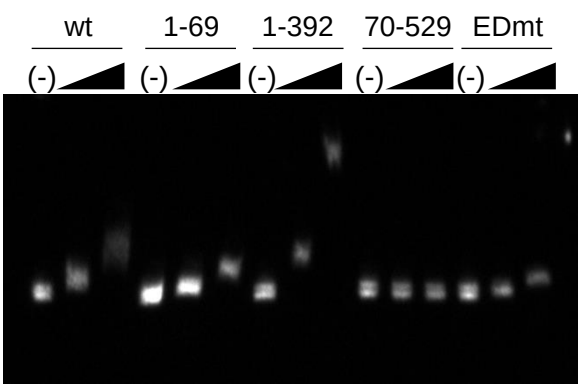
A



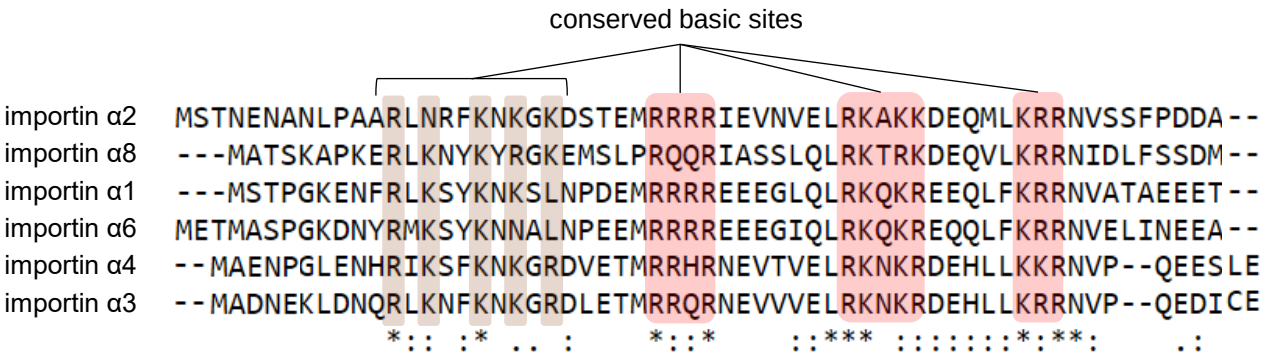
B



C



D



E

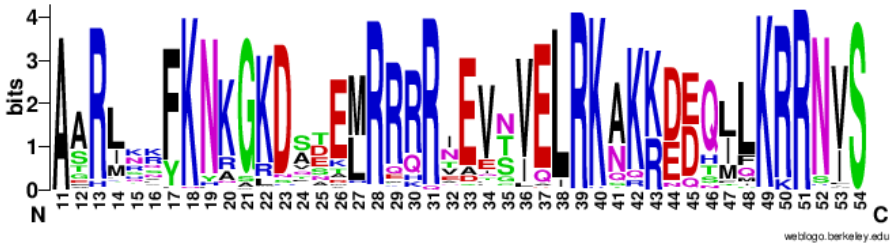
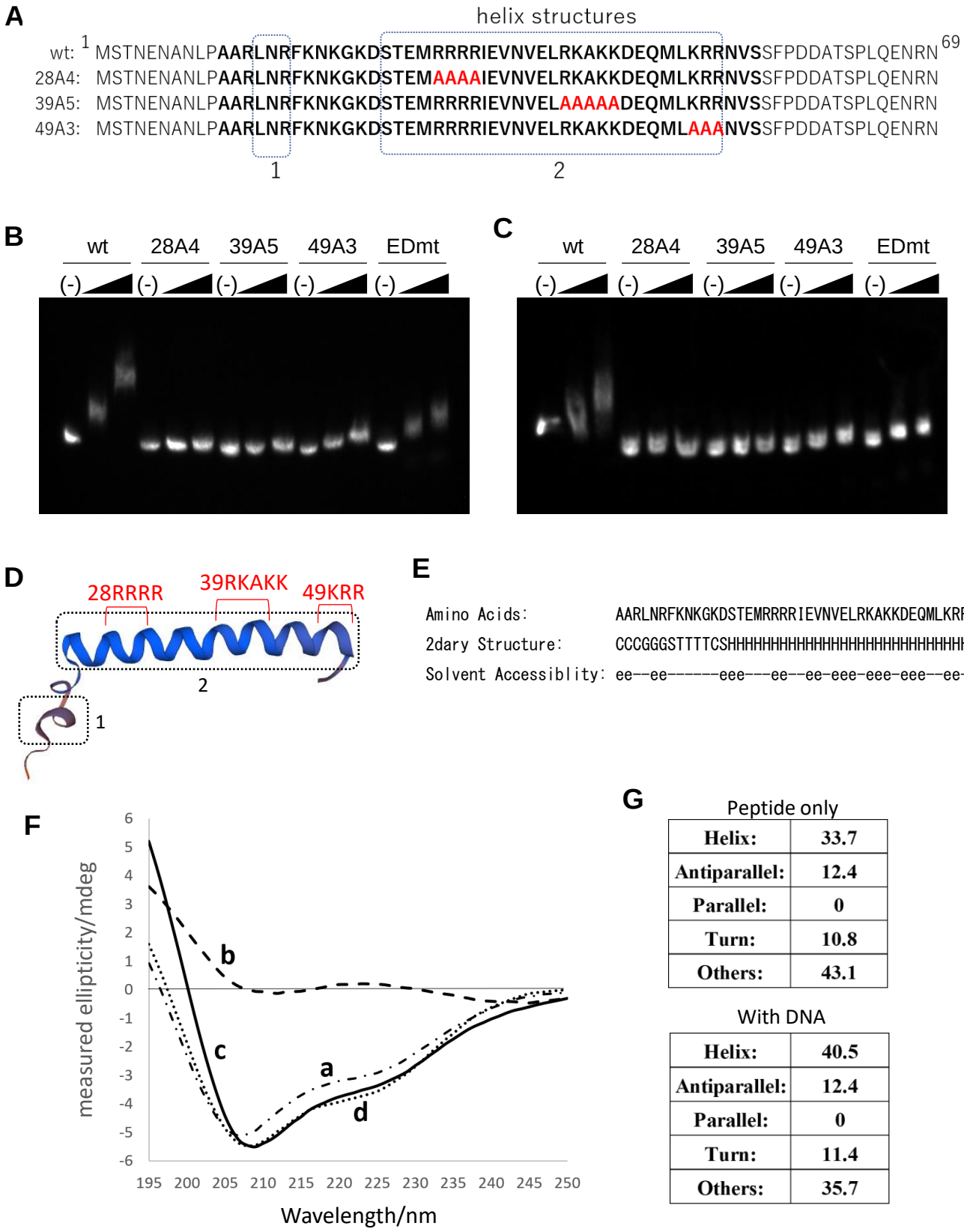


Figure 4



G

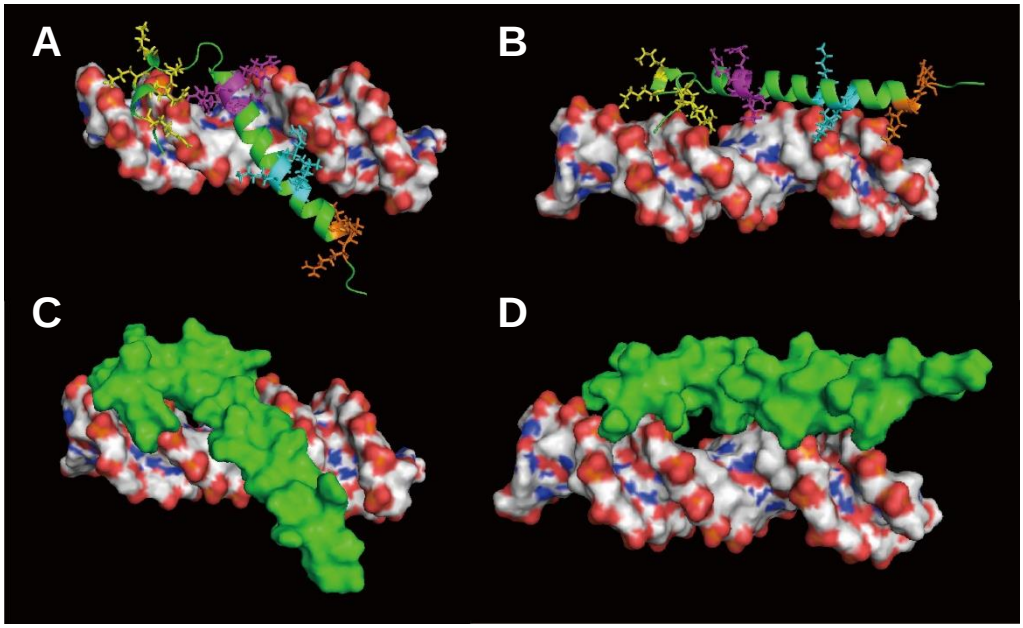
Peptide only

Helix:	**33.7**
Antiparallel:	**12.4**
Parallel:	**0**
Turn:	**10.8**
Others:	**43.1**

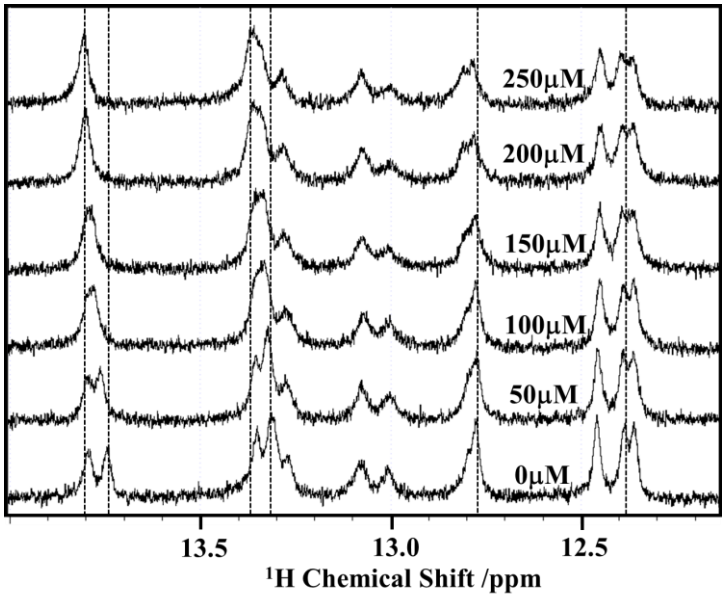
With DNA

Helix:	**40.5**
Antiparallel:	**12.4**
Parallel:	**0**
Turn:	**11.4**
Others:	**35.7**

Figure 5


A
B
C
D

E


250μM
200μM
150μM
100μM
50μM
0μM
13.5
13.0
12.5
¹H Chemical Shift /ppm

F

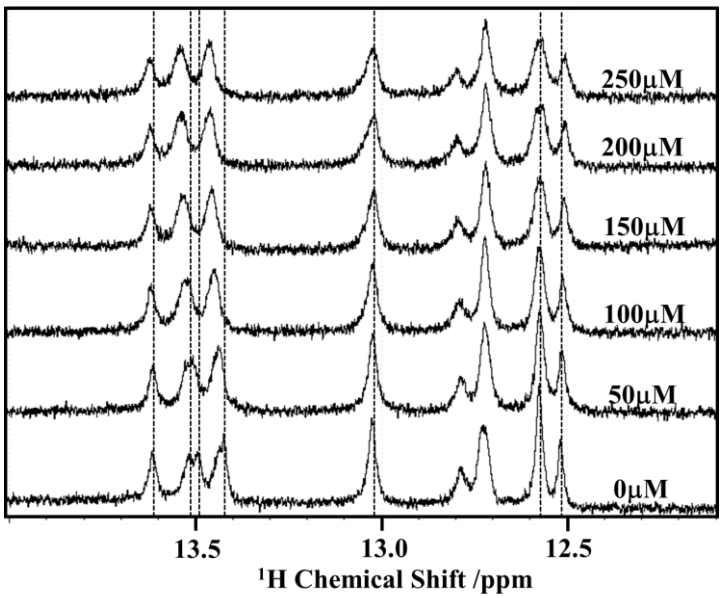

250μM
200μM
150μM
100μM
50μM
0μM
13.5
13.0
12.5
¹H Chemical Shift /ppm

Figure 6

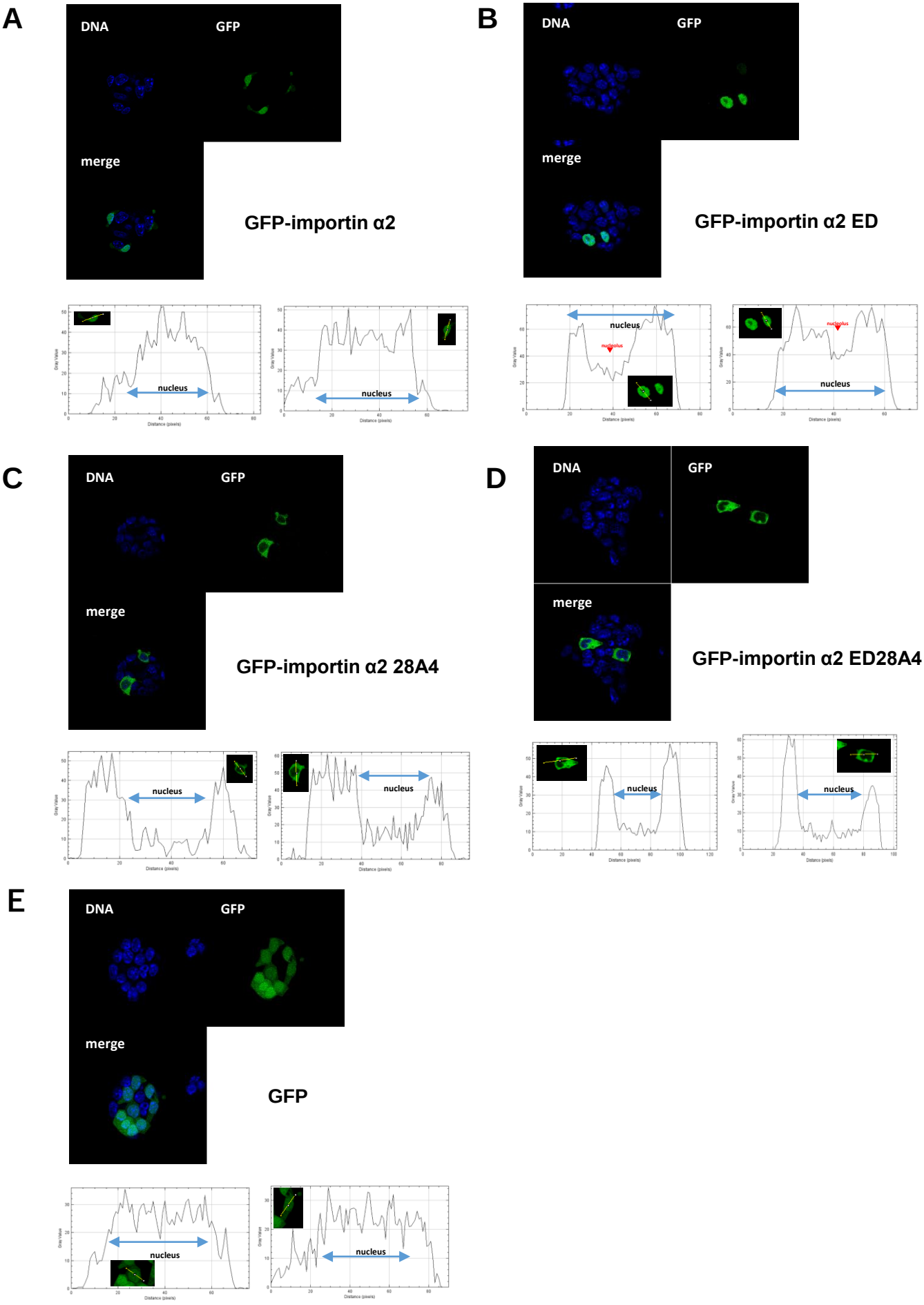

A
DNA
GFP
merge
GFP-importin α2
nucleus
nucleus
B
DNA
GFP
merge
GFP-importin α2 ED
nucleus
nucleolus
nucleolus
nucleus
C
DNA
GFP
merge
GFP-importin α2 28A4
nucleus
nucleus
D
DNA
GFP
merge
GFP-importin α2 ED28A4
nucleus
nucleus
E
DNA
GFP
merge
GFP
nucleus
nucleus

Figure 7

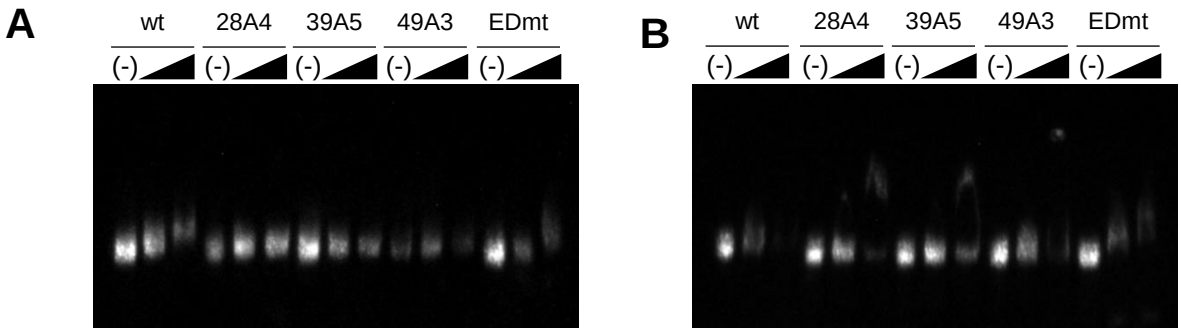
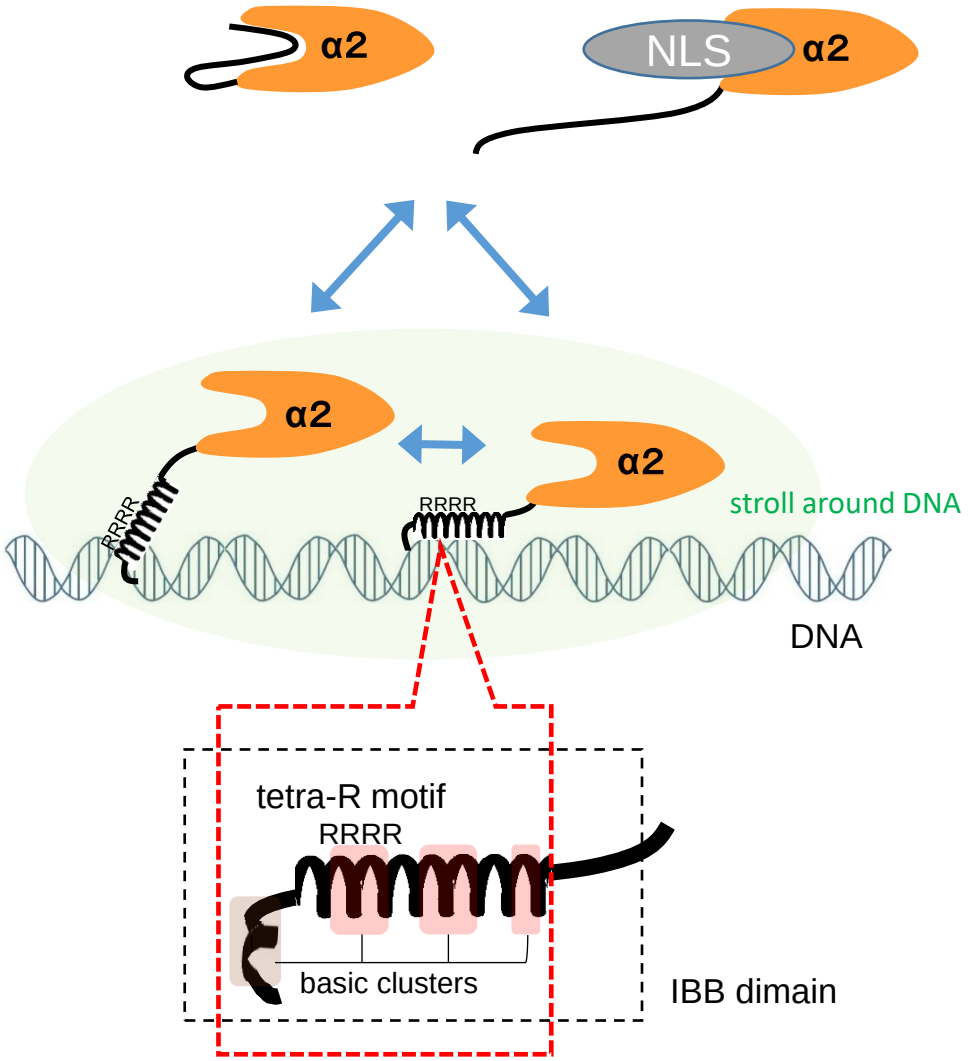


Figure 7



Nucleic Acid Associating Trolley pole domain (NAAT domain)