



ELSEVIER

Contents lists available at ScienceDirect

## Biochemistry and Biophysics Reports

journal homepage: [www.elsevier.com/locate/bbrep](http://www.elsevier.com/locate/bbrep)

## Comprehensive analysis of the dynamic structure of nuclear localization signals

Ryosuke Yamagishi<sup>a</sup>, Takahide Okuyama<sup>b</sup>, Shuntaro Oba<sup>a</sup>, Jiro Shimada<sup>a</sup>, Shigeru Chaen<sup>a</sup>, Hiroki Kaneko<sup>a,b,c,\*</sup><sup>a</sup> Department of Integrated Sciences in Physics and Biology, College of Humanities and Sciences, Nihon University, 3-25-40 Sakurajousui, Setagaya, Tokyo 156-8550, Japan<sup>b</sup> Graduate School of Integrated Basic Sciences, Nihon University, 3-25-40 Sakurajousui, Setagaya-ku, Tokyo 156-8550, Japan<sup>c</sup> National Institute of Biomedical Innovation, 7-6-8 Saito-Asagi, Ibaraki, Osaka 675-0085, Japan

## ARTICLE INFO

## Article history:

Received 30 July 2015

Received in revised form

2 November 2015

Accepted 2 November 2015

Available online 9 November 2015

## Keywords:

Nuclear localization signals (NLSs)

Dynamic structure

Statistical investigation

Intrinsically disordered regions (IDRs)

## ABSTRACT

Most transcription and epigenetic factors in eukaryotic cells have nuclear localization signals (NLSs) and are transported to the nucleus by nuclear transport proteins. Understanding the features of NLSs and the mechanisms of nuclear transport might help understand gene expression regulation, somatic cell reprogramming, thus leading to the treatment of diseases associated with abnormal gene expression. Although many studies analyzed the amino acid sequence of NLSs, few studies investigated their three-dimensional structure. Therefore, we conducted a statistical investigation of the dynamic structure of NLSs by extracting the conformation of these sequences from proteins examined by X-ray crystallography and using a quantity defined as conformational determination rate (a ratio between the number of amino acids determining the conformation and the number of all amino acids included in a certain region). We found that determining the conformation of NLSs is more difficult than determining the conformation of other regions and that NLSs may tend to form more heteropolymers than monomers. Therefore, these findings strongly suggest that NLSs are intrinsically disordered regions.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Understanding nuclear transport of proteins is essential to elucidate gene expression regulation, intracellular signaling, and intermolecular network. In addition, the control of nuclear transport of a variety of transcription factors and epigenetic factors can be applied to various areas including somatic cell reprogramming, treatment of diseases associated with abnormal gene expression such as cancer, and development of new antiviral drugs. The nuclei of eukaryotic cells are separated from the cytosol by the nuclear membrane, and most nuclear proteins cannot move through the nuclear pore alone. Nuclear proteins that can transfer into the nucleus often have nuclear localization signals (NLSs) that are recognized by nuclear transport proteins, which transport the protein into the nucleus. In the classical nuclear transport pathway, nuclear-targeted cargoes containing a NLS are imported by a heterodimeric import receptors consisting of importin  $\alpha$  and importin  $\beta$  [1,2]. Importin  $\beta$  mediates the interaction of the trimeric

complex with the nuclear pore while it translocates into the nucleus [3]. Importin  $\alpha$  is the adaptor protein, which directly binds to the NLS of the cargo. Currently, seven importin- $\alpha$  subtypes of human origin are known [4].

We have proved that importin  $\alpha 2$  (KPNA2, karyopherin alpha 2) has a novel binding site in its C-terminal region, completely different from the previously characterized classical nuclear localization signal (cNLS) binding site, and that importin  $\alpha 2$  maintains the undifferentiated state of ES cells by inhibiting nuclear import of particular transcription factors that induce differentiation [5]. It has been shown that two types of NLSs characterized by different amino acid sequences correspond to these two binding sites. There are several studies on the sequence and structure of NLSs of some proteins [5–8]. The structure of NLSs has been investigated using the partial structure of NLS peptides forming a complex by binding importin  $\alpha$  [9,10]. However, few studies have comprehensively investigated the conformation of NLSs in the entire protein. Therefore, we extracted the NLS region from proteins that were examined by X-ray crystallography analysis. Then, we investigated 1) the feasibility of determining NLS structure compared with other regions; 2) NLS dynamic structural changes; and 3) major differences in 1) determination of NLS structure and 2) NLS dynamic structural changes before and after complex formation with

\* Corresponding author at: Department of Integrated Sciences in Physics and Biology, College of Humanities and Sciences, Nihon University, 3-25-40 Sakurajousui, Setagaya, Tokyo 156-8550, Japan.

E-mail address: [kaneko@phys.chs.nihon-u.ac.jp](mailto:kaneko@phys.chs.nihon-u.ac.jp) (H. Kaneko).

**Table 1**

Conformational determination rate and normalized *B*-factor of the NLS and the entire polypeptide chain.

		<i>N</i>	<i>P</i> -value	Median	Mean
Conformational determination rate	NLS	542	1.900E−03	1.0000	0.7986
	Chain			0.9764	0.9455
Normalized <i>B</i> -factor	NLS	463	< 2.2E−16	0.60010	0.68552
	Chain			0	0

other proteins. Based on the statistical analyses, here we discuss the dynamic features of NLSs in nuclear transport, considering the possibility that NLSs are intrinsically disordered regions [11, 12].

## 2. Material and methods

### 2.1. Extraction of proteins with NLSs and acquisition of positional information of NLSs

For positional information of NLSs, Motif, which is an item of "Family & Domains" in UniProt (<http://www.uniprot.org/>), was used. We chose proteins whose "Nuclear localization signal" is described in the "Description" of the item, and that are categorized in Mammalia. Positional information of NLSs, amino acid sequence of the extracted protein, NLSs sequence information, and reference information of PDB were acquired from UniProt. See reference [13] for information on all mammalian NLSs that are included in UniProt regardless of PDB.

### 2.2. Acquisition of conformational information

PDB ID and chain ID were obtained for proteins analyzed by X-ray crystallography (excluding NMR), using the reference information of PDB acquired from UniProt. For each chain, the amino acid sequence was obtained from the category `pdbx_poly_seq_scheme` by referring mmCIF of PDB, and the sequence was saved in the FASTA format. Missing residues (indicated by "?" in mmCIF) were changed to lower-case letters [14]. Hereinafter, these sequences are referred to as "PDB sequences." Amino acids, whose conformation is determined, are indicated in capital letters, and missing residues are indicated in lower cases.

### 2.3. Numbering of sequences obtained from PDB

The sequence numbering in UniProt and in PDB are not always the same; therefore, new sequence numbers were set for the sequences from PDB in accordance with the sequence number from UniProt by aligning the amino acid sequences of the entire protein obtained from PDB and UniProt, and the sequences from PDB were renumbered. BioPython [15] Pairwise2 was utilized for alignment of sequences. The `blosum62` matrix was used, and gap open was set at −30, and gap extend was set at −0.5.

**Table 2**

Conformational determination rate of the NLS in monomers and heteropolymers.

Uniprot ID	Protein	Conformational determination rate	
		Monomer (PDB ID)	Heteropolymer (PDB ID)
O60566	Mitotic checkpoint serine/threonine-protein kinase BUB1 beta	1.00 (2WV1)	1.00 (3S15)
P05230	Fibroblast growth factor 1	0.75 (2Q9X,3B9U)	0.96 (1DJS,1E00,1EVT,1RY7,3CU1,3OJ2,3OJV,3OJM)
Q8IY67	Ribonucleoprotein PTB-binding 1	1.00 (3SMZ)	1.00 (3H2U,3H2V,3VF0)

### 2.4. Calculation of the conformational determination rate

The conformational determination rate was calculated using sequences from PDB. The conformational determination rate was defined as a ratio calculated by dividing the number of amino acids determining the conformation by the number of all amino acids included in a certain region. In other words, the rate would be a ratio of amino acids indicated in capital letters contained in the "PDB sequence." The conformational determination rate was calculated for both the NLS and the entire polypeptide chain.

### 2.5. Calculation of normalized *B*-factors

Normalized *B*-factors calculated by the following equation [16]:

$$\frac{B_i - \langle B \rangle}{\sigma(B)}$$

where  $B_i$  is *B*-factor of  $C\alpha$  of *i*-th amino acid,  $\langle B \rangle$  is the average value of all  $C\alpha$  atoms and  $\sigma(B)$  is the standard deviation of the *B*-value for the chosen protein.

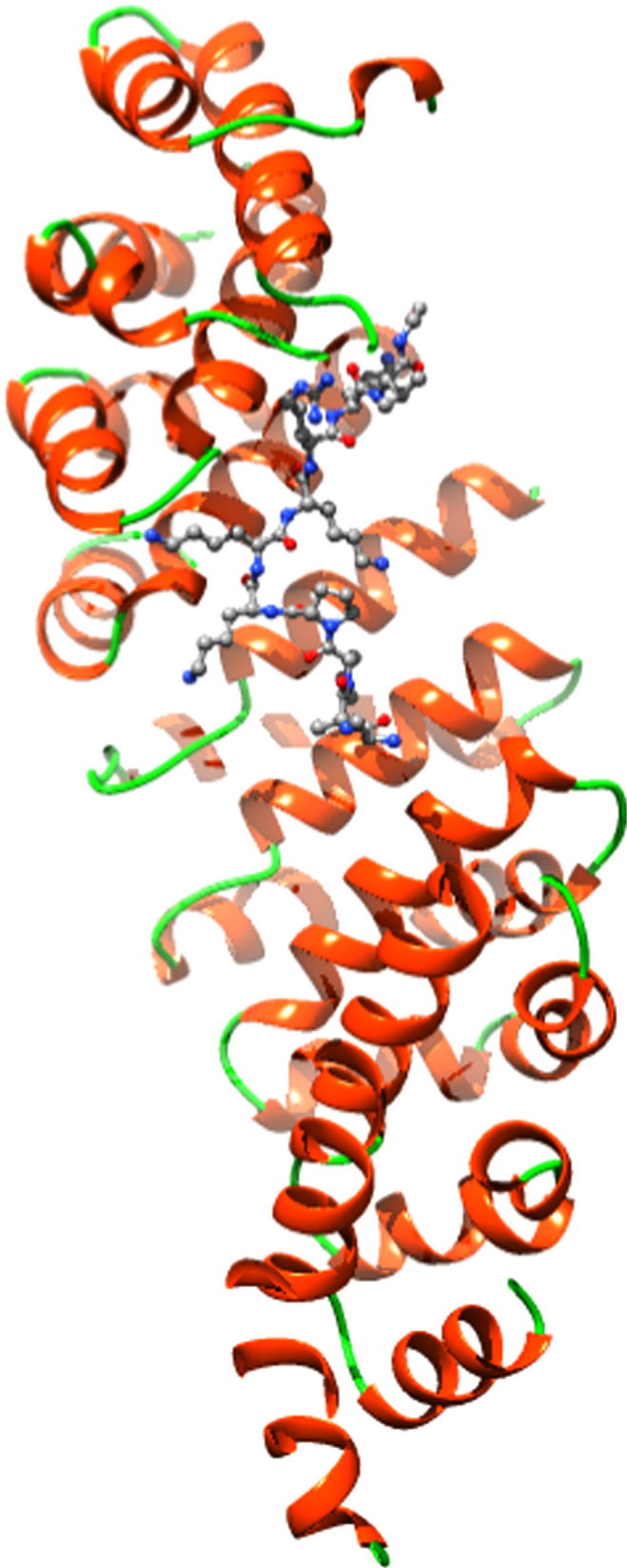
### 2.6. Statistical analyses

A test was conducted to examine whether differences in *B*-factor and in the conformational determination rate of the NLS and the entire polypeptide chain were statistically significant. Because the distribution of the conformational determination rate was not normally distributed, the Wilcoxon signed rank test (one-tailed test) was used for the test.

## 3. Results

### 3.1. Comparison of conformational determination rates of the NLS and the entire polypeptide chain

Here we defined the "conformational determination rate" to evaluate the stability of the conformation. The conformational determination rate was defined as a ratio calculated by dividing the number of amino acids determining the conformation by the number of all amino acids included in a certain region. This value is presented as a mean value per single amino acid. Higher conformational determination rates (apart from 0 and closer to 1) represent higher rate of structure in the region determined by X-ray crystallography. For calculation of the conformational determination rate of the NLS and the entire polypeptide chain, we extracted polypeptide chains that meet the following three conditions from the Protein Data Bank (PDB) [17]: 1) a protein with the NLS position defined by Universal Protein resource (UniProt) [18]; 2) a protein with the link to the PDB entry determined by X-ray crystallography analysis (excluding NMR) described in UniProt; 3) a protein with at least one NLS amino acid included in the SEQRES record of PDB extracted in 1) and 2). Using these



**Fig. 1.** Crystal structure of the complex importin  $\alpha$ -NLS. Simian-virus-40 large T-antigen NLS is indicated with a ball-and-stick model. Gray represents carbon, red represents oxygen, and blue represents nitrogen. Mouse importin  $\alpha$ 1 (Kpna1, karyopherin alpha 1) is indicated with a ribbon model. PDB used in this figure is PDB ID:1Q1T [19]. All figures in this article were drawn using Chimera [27].

conditions, 542 polypeptide chains including 55 NLSs were extracted from 54 proteins (based on UniProt codes). Among these 54 proteins, one (1.85%) was found to have more than one NLS. The conformational determination rate per single amino acid of the NLS and the conformational determination rate per single amino acid of the entire polypeptide chain were calculated for all the extracted chains. In addition, a comparison of the mean of the conformational determination rate per single amino acid of NLS with the mean of the conformational determination rate per single amino acid of the entire polypeptide chain was conducted. An overview of the results is shown in Table 1. Data of individual proteins are shown in Supplementary Table S1, S2 and Fig. S1. The conformational determination rate per single amino acid of the NLS was 0.7986, and the rate for the entire polypeptide was 0.9455. The conformational determination rate per single amino acid of the NLS was lower than the rate of the entire polypeptide. Moreover, the *P*-value indicated that the conformational determination rate of the NLS was significantly lower than the conformational determination rate of the entire polypeptide.

### 3.2. Comparison of normalized *B*-factors between the NLS and the entire polypeptide chain

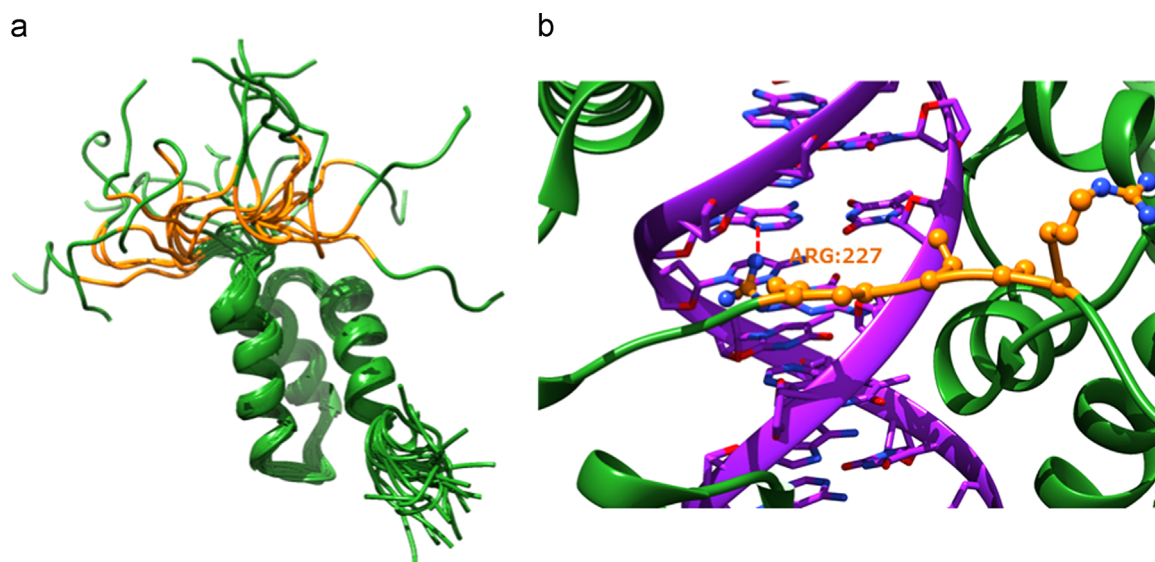
We extracted polypeptide chains meeting the following three conditions: 1) a protein with the NLS position defined by UniProt; 2) a protein with the link to the PDB entry determined by X-ray crystallography analysis (excluding NMR) described in UniProt; 3) a protein with the main-chain coordinates of at least one NLS amino acid included in PDB extracted in 1) and 2). Using these conditions, 463 polypeptide chains were extracted. Then, mean normalized *B*-factors per single amino acid of the NLS and the entire polypeptide chain were calculated for extracted polypeptide chains. The mean normalized *B*-factors per single amino acid were compared between the NLS and the entire polypeptide chain. The results are shown in Table 1. The mean normalized *B*-factor per single amino acid of the NLS was 0.6001, and the mean for the entire polypeptide chain was 0. The mean *B*-factor per single amino acid of the NLS was significantly higher than the mean of the entire polypeptide ( $P < 2.2E - 16$ ).

### 3.3. Conformational determination rate of monomers and heteropolymers

Monomers and heteropolymers (forming a complex with a nucleic acid or a different protein other than itself), whose conformational determination rates could be determined based on data of PDB and were comparable to each other, were extracted. A mean value of the conformational determination rates of the NLS was calculated for each monomer and heteropolymer. The results are shown in Table 2. Only three proteins fulfilled the conditions and were available for comparison. The conformational determination rates of NLS were equivalent before and after heteropolymer formation, or the rates increased after heteropolymer formation.

## 4. Discussion

According to the results in Table 1, the conformational determination rate of the NLS is lower than that of the entire protein; it is difficult to determine the conformation of the NLS regions because of the high mobility in crystals, which is also indicated by significantly high *B*-factors (Table 1).



**Fig. 2.** Structure of Oct3/4. (a) NMR structure of Oct3/4, a transcription factor containing a NLS. PDB used in this figure is PDB ID:1OCP [25]. Twenty-polypeptide chains were superimposed to create the structure by structural alignment. Oct3/4 is indicated with a ribbon model. Orange represents NLS regions, and green represents other regions. (b) Crystal structure of the complex Oct3/4-DNA. The DNA is in purple. The main chain structure of Oct3/4 is shown in the same colors as in (a). Only side chains of the NLS of Oct3/4 are shown with balls and sticks. ARG:227, a NLS, forms a hydrogen bonding with the DNA base. Hydrogen bond is indicated in red-dashed line. Drawn from chain A and DNA of PDB ID:3L1P [26].

We started this study because in the structure of the complex between importin  $\alpha$  and the cargo protein transported into the nucleus, there are no coordinate data showing the entire structure of the cargo protein through the NLS binding site of importin  $\alpha$  (major site). Only the information on the partial structure of a peptide containing several residues before and after the NLS of a protein with a major site of importin  $\alpha$  is available (Fig. 1) [10,19,20]. Furthermore, we have often observed high fluctuation in molecular dynamics calculation of transcription factors compared with other globular proteins. In addition, recently it was pointed out that many nuclear proteins, including transcription factors, epigenetic factors, and factors related to DNA replication and repair, are intrinsically disordered proteins (IDPs), characteristically found in eukaryotic organisms [21–23]. Therefore, we started to consider that NLSs might be intrinsically disordered regions (IDRs), characterized by the so-called coupled folding and binding mechanism [24], in which a polypeptide forms a particular structure through interaction with a target molecule and folds—although NLSs hardly form a particular structure by themselves. For example, the NLS of the transcription factor Oct3/4 indicated various conformations when the conformation was determined as a monomer using NMR (Fig. 2(a)) [25]. However, when it forms a complex by binding with the DNA, the NLS not only forms a particular conformation but is also involved in the binding and recognition of DNA through the interaction with DNA bases (Fig. 2(b)) [26]. We predicted that when Oct3/4 is transported to the nucleus the NLS may form a conformation similar to that shown in Fig. 1 by binding with importin  $\alpha$ , another target molecule.

The determination of intrinsic disorder is essentially based on the missing electron density in X-ray crystal structure and on the negative values of  $^{15}\text{N}$ - $^1\text{H}$  heteronuclear NOE measured by NMR and/or the data obtained by other experimental methods, such as circular dichroism, size exclusion chromatography, SAXS, protease digestion and so on [28]. Based on the above information and on the physical-chemical properties of the amino acids, the disorder prediction algorithms and software are also developed. Disordered regions were often defined as those regions with missing coordinates in X-ray structures during the development of the software. In fact, the training set for DISOPRED (a popular tools in the field) was created by identifying the missing coordinate residues

in X-ray crystal structure as natively disordered residues [29]. Thus our conformational determination rate is faithful to the IDP concept. Therefore we consider the rate as a new parameter to compare intrinsic disorder propensity between different regions. Lower conformational determination rates (apart from 1 and closer to 0) represent higher disorder. From these reasons, the conformational determination rate was compared between monomers and heteropolymers to confirm the relationship between NLSs and IDRs. As shown in Table 2, NLSs tend to form more easily heteropolymers than monomers, thus suggesting that NLSs may be IDRs, whose conformation is stabilized by a target molecule. Similar examples can be found in the previous experimental studies as follows. The NLS of NF- $\kappa\text{B}$  p65 (RelA) adopts a stable helical conformation in the crystal structure of NF- $\kappa\text{B}$  heterodimer (p65/p50) bound to I $\kappa\text{B}\alpha$  [30], although the NMR data show that the RelA NLS in the free state is disordered and highly dynamic [31]. In addition, the RelA NLS is thought to form extended-chain conformation when bound to importin  $\alpha$  [30]. In another instance, fluorescence spectroscopy data show that the NLS of forkhead transcription factor FoxO4 reduces its flexibility when bound to the 14-3-3 protein [32]. Our recent study showed that when transcription factors such as Oct6 and Brn2 form complexes via the classical binding site or the novel C-terminal binding site of importin  $\alpha$ 2, the two NLS conformations are different at the main-chain level [5] as is the case with the above shown NF- $\kappa\text{B}$  p65 (RelA), thus revealing the importance of target proteins in the stabilization of NLS structures. Moreover, the previous studies [33–35] support that NLSs may be IDRs. However, our data were obtained from only three examples, therefore not statistically significant.

Future studies on IDPs will help understand gene networks in eukaryotic organisms. However, it is difficult to precisely define IDRs for proteins, because the structures cannot be determined *per se* alone. We believe that this study and the idea of conformational determination rate defined here will be valuable to comprehend the intrinsic disorder of a protein. In addition, it is required to comprehensively investigate changes in the conformational determination rate caused by post-translational modifications, such as phosphorylation in or near NLSs, which were not examined in this study. In fact, it is important for the

nuclear transport in signal transduction pathway that the NLSs are exposed on the surface and stabilized through a conformational change occurred by phosphorylation, by hetero-multimer formation, or by homo-multimer formation and then recognized by transport proteins. We believe that understanding dynamic conformational characteristics of NLSs will lead to the development of new methods for somatic cell reprogramming through the regulation of nuclear transport and for the development of therapeutic and antiviral agents for diseases associated with abnormal gene expression.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant number 26280108.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.bbrep.2015.11.001>.

## References

- [1] N. Mosammaparast, L.F. Pemberton, Karyopherins: from nuclear-transport mediators to nuclear-function regulators, *Trends Cell Biol.* 14 (2004) 547–556, <http://dx.doi.org/10.1016/j.tcb.2004.09.004>.
- [2] D.S. Goldfarb, A.H. Corbett, D.A. Mason, M.T. Harreman, S.A. Adam, Importin alpha: a multipurpose nuclear-transport receptor, *Trends Cell Biol.* 14 (2004) 505–514, <http://dx.doi.org/10.1016/j.tcb.2004.07.016>.
- [3] R. Bayliss, T. Littlewood, M. Stewart, Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking, *Cell* 102 (2000) 99–108 (accessed 02.03.15) <http://www.ncbi.nlm.nih.gov/pubmed/10929717>.
- [4] R.A. Pumroy, G. Cingolani, Diversification of importin- $\alpha$  isoforms in cellular trafficking and disease states, *Biochem. J.* 466 (2015) 13–28, <http://dx.doi.org/10.1042/BJ20141186>.
- [5] N. Yasuhara, R. Yamagishi, Y. Arai, R. Mehmood, C. Kimoto, T. Fujita, et al., Importin alpha subtypes determine differential transcription factor localization in embryonic stem cells maintenance, *Dev. Cell* 26 (2013) 123–135.
- [6] M.R. Fontes, T. Teh, B. Kobe, Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin- $\alpha$ , *J. Mol. Biol.* 297 (2000) 1183–1194, <http://dx.doi.org/10.1006/jmbi.2000.3642>.
- [7] A. Lange, R.E. Mills, C.J. Lange, M. Stewart, S.E. Devine, A.H. Corbett, Classical nuclear localization signals: definition, function, and interaction with importin alpha, *J. Biol. Chem.* 282 (2007) 5101–5105, <http://dx.doi.org/10.1074/jbc.R600026200>.
- [8] S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, et al., Six classes of nuclear localization signals specific to different binding grooves of importin alpha, *J. Biol. Chem.* 284 (2009) 478–485, <http://dx.doi.org/10.1074/jbc.M807017200>.
- [9] K.E. Süel, A.E. Cansizoglu, Y.M. Chook, Atomic resolution structures in nuclear transport, *Methods* 39 (2006) 342–355, <http://dx.doi.org/10.1016/j.ymeth.2006.06.015>.
- [10] E. Conti, J. Kuriyan, Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha, *Structure* 8 (2000) 329–338 (accessed 02.03.15) <http://www.ncbi.nlm.nih.gov/pubmed/10745017>.
- [11] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol.* 6 (2005) 197–208, <http://dx.doi.org/10.1038/nrm1589>.
- [12] C.J. Oldfield, A.K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions, *Annu. Rev. Biochem.* 83 (2014) 553–584, <http://dx.doi.org/10.1146/annurev-biochem-072711-164947>.
- [13] R. Yamagishi, H. Kaneko, Data from Comprehensive Analysis of Nuclear Localization Signals, *Data Br.* in press, 2015.
- [14] B.W. Brandt, J. Heringa, J.A.M. Leunissen, SEQATOMS: a web tool for identifying missing regions in PDB in sequence context, *Nucleic Acids Res.* 36 (2008) W255–W259, <http://dx.doi.org/10.1093/nar/gkn237>.
- [15] P.J.A. Cock, T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke, et al., Biopython: freely available Python tools for computational molecular biology and bioinformatics, *Bioinformatics* 25 (2009) 1422–1423, <http://dx.doi.org/10.1093/bioinformatics/btp163>.
- [16] Z. Yuan, J. Zhao, Z.-X. Wang, Flexibility analysis of enzyme active sites by crystallographic temperature factors, *Protein Eng.* 16 (2003) 109–114 (accessed 09.01.15) (<http://www.ncbi.nlm.nih.gov/pubmed/12676979>).
- [17] H.M. Berman, The protein data bank, *Nucleic Acids Res.* 28 (2000) 235–242, <http://dx.doi.org/10.1093/nar/28.1.235>.
- [18] The UniProt Consortium, Activities at the Universal Protein Resource (UniProt), *Nucleic Acids Res.* 42 (2014) D191–D198, <http://dx.doi.org/10.1093/nar/gkt1140>.
- [19] M.R.M. Fontes, T. Teh, G. Toth, A. John, I. Pavo, D.A. Jans, et al., Role of flanking sequences and phosphorylation in the recognition of the simian virus-40 large T-antigen nuclear localization sequences by importin- $\alpha$ , *Biochem. J.* 375 (2003) 339–349, <http://dx.doi.org/10.1042/BJ20030510>.
- [20] M.-H. Chen, I. Ben-Efraim, G. Mitrousis, N. Walker-Kopp, P.J. Sims, G. Cingolani, Phospholipid scramblase 1 contains a nonclassical nuclear localization signal with unique binding site in importin alpha, *J. Biol. Chem.* 280 (2005) 10599–10606, <http://dx.doi.org/10.1074/jbc.M413194200>.
- [21] V.N. Uversky, A.K. Dunker, Understanding protein non-folding, *Biochim. Biophys. Acta* 1804 (2010) 1231–1264, <http://dx.doi.org/10.1016/j.bbapap.2010.01.017>.
- [22] K.S. Sandhu, Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins, *J. Mol. Recognit.* 22 (2009) 1–8, <http://dx.doi.org/10.1002/jmr.915>.
- [23] L.Y. Beh, L.J. Colwell, N.J. Francis, A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence, *Proc. Natl. Acad. Sci. USA* 109 (2012) E1063–E1071, <http://dx.doi.org/10.1073/pnas.1118678109>.
- [24] K. Sugase, H.J. Dyson, P.E. Wright, Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature* 447 (2007) 1021–1025, <http://dx.doi.org/10.1038/nature05858>.
- [25] E.H. Morita, M. Shirakawa, F. Hayashi, M. Imagawa, Y. Kyogoku, Structure of the Oct-3 POU-homeodomain in solution, as determined by triple resonance heteronuclear multidimensional NMR spectroscopy, *Protein Sci.* 4 (1995) 729–739, <http://dx.doi.org/10.1002/pro.5560040412>.
- [26] D. Esch, J. Vahokoski, M.R. Groves, V. Pogenberg, V. Cojocar, H. Vom Bruch, et al., A unique Oct4 interface is crucial for reprogramming to pluripotency, *Nat. Cell Biol.* 15 (2013) 295–301, <http://dx.doi.org/10.1038/ncb2680>.
- [27] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E. C. Meng, et al., UCSF Chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612, <http://dx.doi.org/10.1002/jcc.20084>.
- [28] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, et al., Intrinsically disordered protein, *J. Mol. Graph. Model.* 19 (2001) 26–59 (accessed 10.09.15) <http://www.ncbi.nlm.nih.gov/pubmed/11381529>.
- [29] J.J. Ward, J.S. Sodhi, L.J. McGuffin, B.F. Buxton, D.T. Jones, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J. Mol. Biol.* 337 (2004) 635–645, <http://dx.doi.org/10.1016/j.jmb.2004.02.002>.
- [30] M.D. Jacobs, S.C. Harrison, Structure of an I $\kappa$ B $\alpha$ /NF- $\kappa$ B complex, *Cell* 95 (1998) 749–758, [http://dx.doi.org/10.1016/S0092-8674\(00\)81698-0](http://dx.doi.org/10.1016/S0092-8674(00)81698-0).
- [31] C.F. Cervantes, S. Bergqvist, M. Kjaergaard, G. Kroon, S.-C. Sue, H.J. Dyson, et al., The RelA nuclear localization signal folds upon binding to I $\kappa$ B $\alpha$ , *J. Mol. Biol.* 405 (2011) 754–764, <http://dx.doi.org/10.1016/j.jmb.2010.10.055>.
- [32] V. Obsilova, J. Vecer, P. Herman, A. Pabianova, M. Sulc, J. Teisinger, et al., 14-3-3 Protein interacts with nuclear localization sequence of forkhead transcription factor FoxO4, *Biochemistry* 44 (2005) 11608–11617, <http://dx.doi.org/10.1021/bi050618r>.
- [33] Y. Zeng, Y. He, F. Yang, S.M. Mooney, R.H. Getzenberg, J. Orban, et al., The cancer/testis antigen prostate-associated gene 4 (PAGE4) is a highly intrinsically disordered protein, *J. Biol. Chem.* 286 (2011) 13985–13994, <http://dx.doi.org/10.1074/jbc.M110.210765>.
- [34] S. Costantini, A. Sharma, R. Rauti, M. Costantini, I. Autiero, G. Colonna, Genealogy of an ancient protein family: the Sirtuins, a family of disordered members, *BMC Evol. Biol.* 13 (2013) 60, <http://dx.doi.org/10.1186/1471-2148-13-60>.
- [35] Y. Minezaki, K. Homma, A.R. Kinjo, K. Nishikawa, Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation, *J. Mol. Biol.* 359 (2006) 1137–1149, <http://dx.doi.org/10.1016/j.jmb.2006.04.016>.