



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

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

## Data Article

## Data from comprehensive analysis of nuclear localization signals

Ryosuke Yamagishi<sup>a</sup>, Hiroki Kaneko<sup>a,b,\*</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> National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki, Osaka 675-0085, Japan

## ARTICLE INFO

## Article history:

Received 4 November 2015

Received in revised form

26 November 2015

Accepted 27 November 2015

Available online 12 December 2015

## Keywords:

Nuclear localization signal

Nuclear transport

Comprehensive analysis

UniProt

Distribution

## ABSTRACT

This article describes data related to a research article titled “Comprehensive analysis of the dynamic structure of nuclear localization signals” by Yamagishi et al. [1]. In this article, we provide the data covering wider range of the mammalian NLSs in UniProt (Universal Protein Resource) [2] regardless of their conformations. To be more specific as follows: We have extracted all NLSs which are clearly indicated as “NLS” with evidence type (a code from the Evidence Codes Ontology) [3] in UniProt. A total of 1364 NLSs in 1186 proteins were extracted from UniProt. The number of NLSs found in each protein (UniProt ID), the sequence length of NLSs and their distribution are shown.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	<i>Biology</i>
More specific subject area	<i>Molecular biology, Cell biology, Nuclear transport</i>
Type of data	<i>Table, graph</i>

DOI of original article: <http://dx.doi.org/10.1016/j.bbrep.2015.11.001>

\* Corresponding author.

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

<http://dx.doi.org/10.1016/j.dib.2015.11.064>

2352-3409/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

How data was acquired	<i>Database analysis</i>
Data format	<i>Analyzed</i>
Experimental factors	<i>Analysis of information on mammalian nuclear localization signals identified in UniProt</i>
Experimental features	<i>We have extracted all NLSs which are clearly indicated as “NLS” with evidence type (a code from the Evidence Codes Ontology) in UniProt. We analyzed the number of NLSs found in each protein (UniProt ID) and their distribution and the sequence length of NLSs and their distribution.</i>
Data source location	
Data accessibility	<i>Data are accessible in this article</i>

---

## 1. Value of the data

---

- A total of 1364 NLSs in 1186 proteins were extracted from UniProt. The data have value in providing only accurate NLS information that is clearly indicated as “NLS” with evidence type (a code from the Evidence Codes Ontology) in UniProt. Therefore the data can be used as a training set for the development of NLS prediction programs.
  - The data presented here are useful for researchers who study NLS and nuclear transport mechanism.
  - Future studies concerning development of new therapeutic agents for human diseases caused by deregulation of nuclear transport such as numerous cancers and developmental disorders would require the data included here and the features of mammalian NLSs shown by the data.
- 

## 2. Data

In summary, a total of 1364 NLSs in 1186 proteins were extracted from UniProt. Data of individual NLSs are shown in [Supplementary Table S1](#). The distribution of the length showed two peaks: one at 4–7 and one at 16–18, indicating the presence of monopartite and bipartite classical NLSs. The ratio of the NLSs consist of more than 30 residues was 1.10% (15/1364). We analyzed the sequence length of the NLSs and their distribution ([Fig. 1](#)). The numbers of NLSs found in one protein are also given in [Supplementary Table S2](#). We analyzed the distribution of the number of NLSs in one protein molecule ([Fig. 2](#)). The proteins having only one NLS site were in large part and the ratio of the part was 86.93% (1031/1186). The numbers of proteins having two, three, four and five NLS sites were 138, 12, 4 and 1 in 1186, respectively.

## 3. Experimental design, materials and methods

### 3.1. Extraction of proteins with NLSs

In order to obtain the proteins having NLSs, 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. In more detail, we extracted the proteins from UniProt that satisfy the following conditions: annotation: (type:motif AND “nuclear localization signal”) AND taxonomy: “Mammalia [40674]” AND reviewed: yes. The UniProt IDs of the proteins having NLSs were obtained by this means.

Download English Version:

<https://daneshyari.com/en/article/174974>

Download Persian Version:

<https://daneshyari.com/article/174974>

[Daneshyari.com](https://daneshyari.com)