© 2003 Oxford University Press

Nucleic Acids Research, 2003, Vol. 31, No. 1 **393–396**
DOI: 10.1093/nar/gkg101

NESbase version 1.0: a database of nuclear export signals

Tanja la Cour[1,2], Ramneek Gupta[1], Kristoffer Rapacki[1], Karen Skriver[2], Flemming M. Poulsen[2] and Søren Brunak[1,]*

[1]Center for Biological Sequence Analysis, Building-208, Technical University of Denmark, DK-2800 Lyngby, Denmark and [2]Department of Protein Chemistry, Institute of Molecular Biology, University of Copenhagen, DK-1353, Denmark

Received August 15, 2002; Revised and Accepted October 27, 2002

ABSTRACT

Protein export from the nucleus is often mediated by a Leucine-rich Nuclear Export Signal (NES). NESbase is a database of experimentally validated Leucine-rich NESs curated from literature. These signals are not annotated in databases such as SWISS-PROT, PIR or PROSITE. Each NESbase entry contains information of whether NES was shown to be necessary and/or sufficient for export, and whether the export was shown to be mediated by the export receptor CRM1. The compiled information was used to make a sequence logo of the Leucine-rich NESs, displaying the conservation of amino acids within a window of 25 residues. Surprisingly, only 36% of the sequences used for the logo fit the widely accepted NES consensus L-x(2,3)-[LIVFM]-x(2,3)-L-x-[LI]. The database is available online at http://www.cbs.dtu.dk/databases/NESbase/.

INTRODUCTION

Protein localization is a key feature which is often used to support functional hypotheses. Eukaryotic cells are characterized by having their genetic material confined by a nuclear envelope. This implies that transcriptional and translational events are physically separated, which creates a need for substantial transport across the nuclear envelope. This compartmentalization also provides a means of controlling the availability of regulatory proteins in the nucleus.

Transport across the nuclear envelope occurs through the evolutionary conserved Nuclear Pore Complex (NPC), a huge proteinaceous structure forming an aqueous channel with an internal diameter of ~9 nm for passive diffusion, and of ~25 nm for active transport. Active nucleocytoplasmic transport of proteins is dependent on a gradient of RanGTP across the nuclear envelope, and mediated by receptors belonging to the importin β superfamily (1–5 for review).

Most of the nucleocytoplasmic transport is active. This is a signal dependent process, requiring a sequence motif in the protein to be transported. The mechanism of nuclear import of proteins has been extensively studied (6,2), and Nuclear Localization Signals (NLSs) are annotated in SWISS-PROT and PIR (7,8). Both experimental and potential NLSs are retrievable from the NLSdb database at the PredictNLS prediction server (9).

Nuclear export of proteins is a more recent subject of investigation. A Leucine-rich Nuclear Export Signal (NES) was simultaneously identified in the HIV Rev protein (10) and in the Protein Kinase A inhibitor (PKI) (11). The evolutionary conserved CRM1 (also called exportin1/Xpo1) protein was identified as being the export receptor of proteins containing Leucine-rich NESs (12–17). The fungicide Leptomycin B (LMB) was shown to interact directly with CRM1 and to inhibit CRM1-mediated export (12), providing excellent experimental verification of this pathway. Only Crm1p of *Saccharomyces cerevisiae* is not sensitive to LMB, but a single amino acid substitution converts Crm1p into being LMB-sensitive (18). Other export receptors have been identified (19), but CRM1-mediated export of NES proteins remains the most extensively studied export pathway to date. Recently calreticulin was shown to mediate export of PKI in a NES-dependent, LMB-insenstive manner, indicating that CRM1 might not be the only export receptor recognizing the Leucine-rich NES (20).

Besides being important for a better understanding of eukaryotic gene function and regulation, insight into the mechanism and regulation of nuclear export might also be relevant from a therapeutic point of view: Many of the reported nucleocytoplasmic shuttle proteins are involved in signal transduction events and cell cycle regulation (21). In addition, export of unspliced and partially spliced HIV mRNA depends on the Leucine-rich NES of the HIV Rev protein (22).

To date, many Leucine-rich NESs have been identified and reported in the literature, but so far this information has not been compiled into a database form. Many of the identified Leucine-rich NESs deviate significantly from the generally accepted loose consensus L-x(2,3)-[LIVFM]-x(2,3)-L-x-[LI] proposed earlier (23). We have collected experimentally determined Leucine-rich NESs in the NESbase 1.0 database described in this paper.

*To whom correspondence should be addressed. Tel: +45 45252477; Fax: +45 45931585; Email: brunak@cbs.dtu.dk

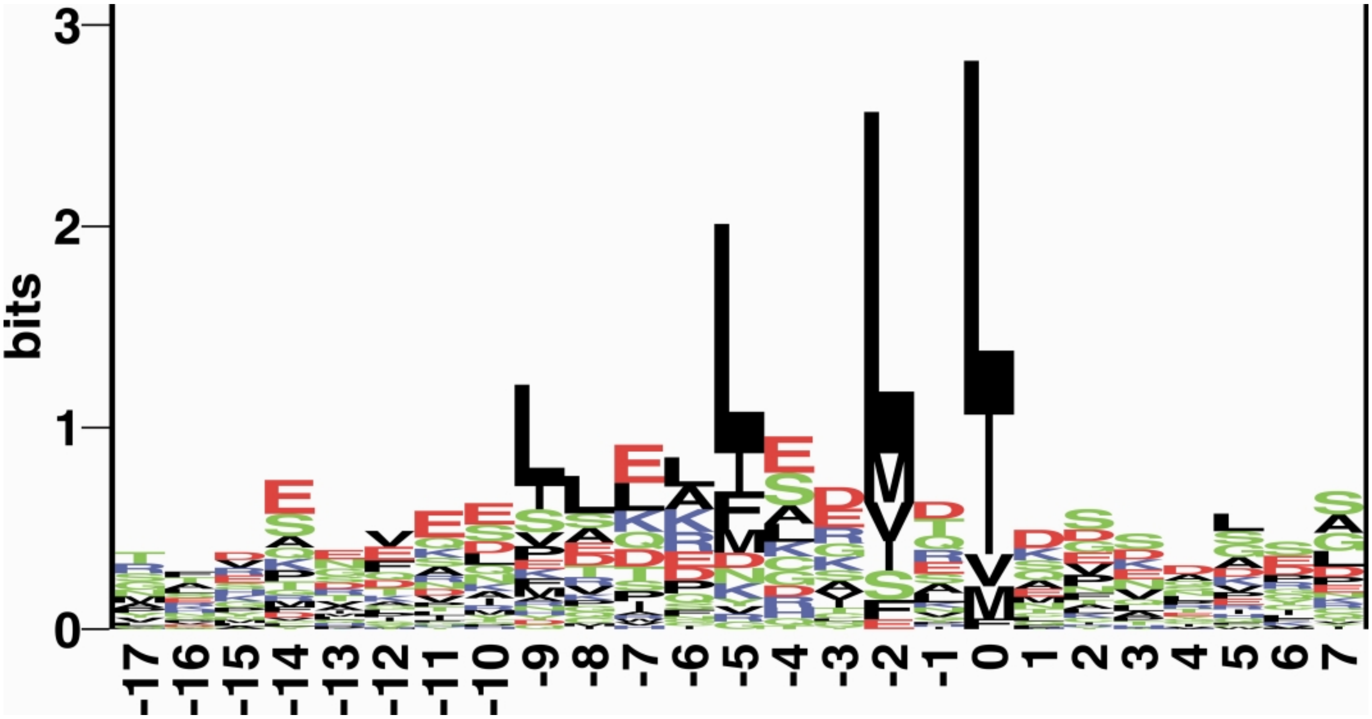


Figure 1. An alignment of 58 high-quality NESs is shown in the form of Shannon information content (25) represented as sequence logos (24). Only those signals, which were shown to be either necessary or sufficient, and for which the pathway was experimentally verified, were included. Alignment was performed by ClustalX (26) over a window of 25 amino acids. The height of each column reflects the non-random bias of particular residues at that position, the size of each residue letter reflecting its frequency at that position. The colour code is green for polar, black for hydrophobic, red for acidic and blue for basic amino acids.In the logo the hydrophobic residues Leucine, Isoleucine, Methionine, Valine and Phenylalanine dominates the aligned signal. Also prominent are the Glutamic acid and Aspartic acid residues. Only 36% of the sequences used for the logo fit the widely accepted NES consensus L-x(2,3)-[LIVFM]-x(2,3)-L-x-[LI].

LEUCINE-RICH NES SEQUENCE MOTIFS

Leucine-rich NES signals consist of 4–5 hydrophobic residues within a region of ~10 amino acids. These hydrophobic residues are predominantly Leucine, but may also be Isoleucine, Valine, Methionine and Phenylalanine. A mutational study of the PKI NES indicated that Leucines in the C-terminal end of the signal are more important for function, than the N-terminal Leucines (11). This is supported by a sequence logo (24), in which the C-terminal hydrophobic residues are seen clearly as the most conserved (Fig. 1). The logo indicates that residues favoured in the region of the signal, other than hydrophobic residues, are Glutamic acid and Aspartic acid. Comparison of this sequence logo with the aforementioned NES consensus L-x(2,3)-[LIVFM]-x(2,3)-L-x-[LI], indicates that the consensus is insufficient to describe many NESs: Only 36% of the sequences used for the logo, actually fit that consensus. This motivates the need to build a prediction method, which we currently are constructing.

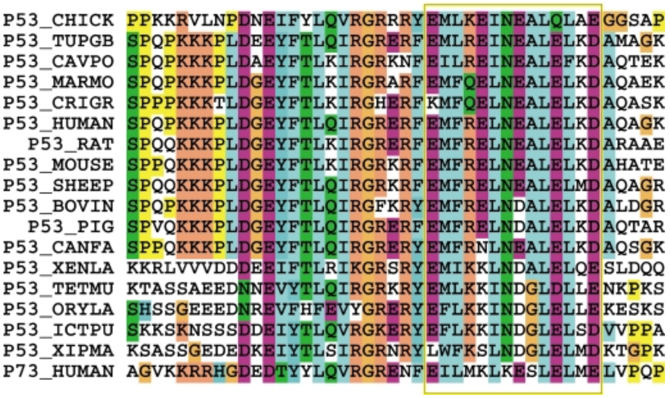
An example of the degree of conservation of a NES signal is shown in Figure 2, where the NES regions of the tumor suppressor p53 from 17 different species and the paralog p73 protein are aligned. The NES signal has been experimentally verified in both human p53 and p73. Within the NES region (as indicated by a box in the figure), besides a high degree of conservation of hydrophobic residues, also acidic residues are conserved both across species and between the p53 and p73 paralogs. In contrast, the conservation of a basic residue and an Asparagine in the region are conserved between the p53 orthologs while not in p73, indicating this feature only to be of importance to p53 specific function. The conservation of acidic residues in the region is in good agreement with acidic residues being slightly favoured in the NES region as illustrated in the sequence logo (Fig. 1) and suggest a role for acidic residues in nuclear export, although the position of these are not generally conserved among NESs. A complete sequence analysis of the NES conservation, will be published elsewhere.

DATA SOURCES

Since (NESs) are not annotated in the protein databases PIR and SWISS-PROT, we collected the information entirely from published literature. We screened over 200 published articles.

DATABASE FORMAT—VERSION 1.0

Version 1.0 of NESbase contains 75 entries with 80 experimentally determined NESs. An example of an entry is shown in Figure 3. For each entry there is a description of whether the NES was shown to be necessary and/or sufficient for export, and whether the export pathway was shown to be mediated by the CRM1 receptor. Information on steady-state localization of the protein and regulation of export was included, where information was readily available. Using SWISS-PROT identifiers, we screened the database of NLSs,



```
P53_CHICK PPKKRVLNPDNEIFYLQVRGRRRYEMLKEINEALQLAEGGSAP
P53_TUPGB SPQPKKKPLDEEYFTLQIRGRERFEMLREINEALELKDAMAGK
P53_CAVPO SPQPKKKPLDAEYFTLKIRGRKNFEILREINEALEFKDAQTEK
P53_MARMO SPQPKKKPLDGEYFTLKIRGRARFEMFQELNEALELKDAQAEK
P53_CRIGR SPPPKKKTLDGEYFTLKIRGHERFKMFQELNEALELKDAQASK
P53_HUMAN SPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGK
  P53_RAT SPQQKKKPLDGEYFTLKIRGRERFEMFRELNEALELKDARAAE
P53_MOUSE SPPQKKKPLDGEYFTLKIRGRKRFEMFRELNEALELKDAHATE
P53_SHEEP SPQQKKKPLDGEYFTLQIRGRKRFEMFRELNEALELMDAQAGR
P53_BOVIN SPQPKKKPLDGEYFTLQIRGFKRYEMFRELNDALELKDALDGR
  P53_PIG SPVQKKKPLDGEYFTLQIRGRERFEMFRELNDALELKDAQTAR
P53_CANFA SPPQKKKPLDGEYFTLQIRGRERYEMFRNLNEALELKDAQSGK
P53_XENLA KKRLVVVDDDEEIFTLRIKGRSRYEMIKKLNDALELQESLDQQ
P53_TETMU KTASSAEEDNNEVYTLQIRGRKRYEMLKKINDGLDLLENKPKS
P53_ORYLA SHSSGEEEDNREVFHFEVYGRERYEFLKKINDGLELLEKESKS
P53_ICTPU SKKSKNSSSDDEIYTLQVRGKERYEFLKKINDGLELSDVVPPA
P53_XIPMA KSASSGEDEDKEIYTLSIRGRNRYLWFKSLNDGLELMDKTGPK
P73_HUMAN AGVKKRRHGDEDTYYLQVRGRENFEILMKLKESLELMELVPQP
```

Figure 2. Evolutionary conservation of p53 NES. A Blastp search against SWISS-PROT was performed with all entries of NESbase.The result for human p53 was thirty ortholog p53 sequences and two p73 proteins, which are p53 paralogs. The human p73 paralog also have an experimental verified NES. This makes p53 the most illustrative example when studying evolutionary conservation of NESs. In this figure the seventeen p53 orthologs having unique NESs are aligned with the human p73 paralog. The region of conserved hydrophobic residues, constituting the experimentally verified NES, is boxed.

NLSdb, at http://cubic.bioc.columbia.edu/db/NLSdb/ for NESbase entries. Surprisingly only 7 NESbase entries were found to also contain an experimentally verified NLS, and 13 were found to contain what is described as a potential NLS. This means that currently the overlap between NESbase and NLSdb is small. NESbase 1.0 is cross-referenced to SWISS-PROT (or TrEMBL), NLSdb and MEDLINE abstracts. The Database features are described below:

NES-ACCESSION: NESbase accession code (e.g. NES-0001).

DATE: Dates for creation and update of entry.

PROTEIN: Protein description.

ORGANISM: Species name.

DB_REFERENCE: Sequence cross-reference to SWISS-PROT, PIR, GenBank and NLSdb.

NECESSITY: Experimental evidence (if any), indicating that the signal is necessary for export. Usually shown by deletion of or mutations in the signal that greatly impairs or abrogates export.

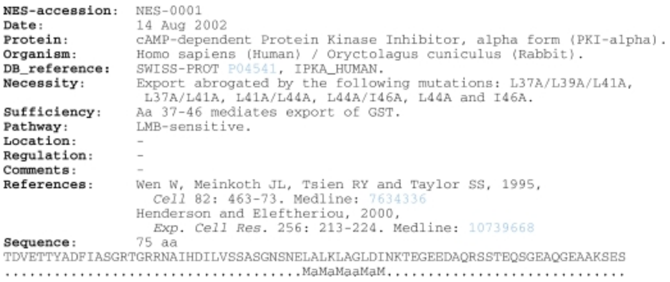
SUFFICIENCY: Experimental evidence (if any), indicating that the signal is sufficient for export. Usually shown by the ability of a peptide containing the NES, to mediate export of a reporter protein.

PATHWAY: LMB-sensitivity and/or CRM1-dependency if described.

STEADY STATE LOCATION: Nucleocytoplasmic shuttling proteins are often predominantly localized to either nucleus or cytoplasm at steady state. When readily available, information of steady state localization is given in this field.

REGULATION: Export can be regulated, for instance, by a change in phosphorylation state of the proteins or by protein–protein interactions masking the NES signal. When readily available, information about regulation is given in this field.

COMMENTS: Overall comments to indicate any other important details.



```
NES-accession:   NES-0001
Date:            14 Aug 2002
Protein:         cAMP-dependent Protein Kinase Inhibitor, alpha form (PKI-alpha).
Organism:        Homo sapiens (Human) / Oryctolagus cuniculus (Rabbit).
DB_reference:    SWISS-PROT P04541, IPKA_HUMAN.
Necessity:       Export abrogated by the following mutations: L37A/L39A/L41A,
                  L37A/L41A, L41A/L44A, L44A/I46A, L44A and I46A.
Sufficiency:     Aa 37-46 mediates export of GST.
Pathway:         LMB-sensitive.
Location:        -
Regulation:      -
Comments:        -
References:      Wen W, Meinkoth JL, Tsien RY and Taylor SS, 1995,
                   Cell 82: 463-73. Medline: 7634336
                 Henderson and Eleftheriou, 2000,
                   Exp. Cell Res. 256: 213-224. Medline: 10739668
Sequence:        75 aa
TDVETTYADFIASGRTGRRNAIHDILVSSASGNSNELALKLAGLDINKTEGEEDAQRSSTEQSGEAQGEAAKSES
....................................MaMaMaaMaM..............................
```

Figure 3. A typical database entry from NESbase 1.0. Underlining indicates html-links. The NESbase WWW homepage is at http://www.cbs.dtu.dk/databases/NESbase/.

REFERENCE(S): Literature reference(s) and MEDLINE links.

SEQUENCE: Sequence followed by an assignment field. The assignment field reflects the experimental data described in the necessity and sufficiency fields: (1) Point-mutations that greatly impair export either alone or together are designated ‘M’. (2) Sequence segments shown to mediate export of a truncated parent protein or a reporter protein are designated ‘a’.

DATABASE ACCESS

NESbase is made available on the WWW at http://www.cbs.dtu.dk/databases/NESbase/.

We encourage users to provide updates, corrections and new information to the database, which will be accordingly updated in order to provide as good data as possible. The WWW site has a submission page. Unpublished information will be held in confidence on request of the authors. We encourage users of NESbase to cite this paper.

ACKNOWLEDGEMENTS

We acknowledge support from the Danish National Research Foundation and the John and Birthe Meyer Foundation.

REFERENCES

1. Rout,M. and Aitchison,J. (2001) The nuclear pore complex as a transport machine. *J. Biol. Chem.*, **276**, 16593–16596.
2. Mattaj,I. and Englmeier,L. (1998) Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.*, **67**, 265–306.
3. Gorlich,D. and Kutay,U. (1999) Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.*, **15**, 607–660.
4. Nakielny,S. and Dreyfuss,G. (1999) Transport of proteins and RNAs in and out of the nucleus. *Cell*, **99**, 677–690.
5. Kuersten,S., Ohno,M. and Mattaj,I. (2001) Nucleocytoplasmic transport: Ran, beta and beyond. *Trends Cell Biol.*, **11**, 497–503.
6. Dingwall,C. and Laskey,R. (1991) Nuclear targeting sequences-a consensus? *Trends Biochem. Sci.*, **16**, 478–481.
7. Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.
8. Wu,C., Huang,H., Arminski,L., Castro-Alvear,J., Chen,Y., Hu,Z., Ledley,R., Lewis,K., Mewes,H., Orcutt,B., Suzek,B., Tsugita,A., Vinayaka,C., Yeh,L., Zhang,J. and Barker,W. (2002) The Protein Information Resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res.*, **30**, 35–37.
9. Cokol,M., Nair,R. and Rost,B. (2000) Finding nuclear localization signals. *EMBO Rep.*, **1**, 411–415.

10. Fischer,U., Huber,J., Boelens,W., Mattaj,I. and Luhrmann,R. (1995) The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*, **82**, 475–483.
11. Wen,W., Meinkoth,J., Tsien,R. and Taylor,S. (1995) Identification of a signal for rapid export of proteins from the nucleus. *Cell*, **82**, 463–473.
12. Fornerod,M., Ohno,M., Yoshida,M. and Mattaj,I. (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, **90**, 1051–1060.
13. Stade,K., Ford,C., Guthrie,C. and Weis,K. (1997) Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell*, **90**, 1041–1050.
14. Neville,M., Stutz,F., Lee,L., Davis,L.I. and Rosbash,M. (1997) The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr. Biol.*, **7**, 767–775.
15. Ossareh-Nazari,B., Bachelerie,F. and Dargemont,C. (1997) Evidence for a role of CRM1 in signal-mediated nuclear protein export. *Science*, **278**, 141–144.
16. Fukuda,M., Asano,S., Nakamura,T., Adachi,M., Yoshida,M., Yanagida,M. and Nishida,E. (1997) CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature*, **390**, 308–311.
17. Haasen,D., Kohler,C., Neuhaus,G. and Merkle,T. (1999) Nuclear export of proteins in plants: AtXPO1 is the export receptor for leucine-rich nuclear export signals in *Arabidopsis thaliana*. *Plant J.*, **20**, 695–705.
18. Neville,M. and Rosbash,M. (1999) The NES-Crm1p export pathway is not a major mRNA export route in *Saccharomyces cerevisiae*. *EMBO J.*, **18**, 3746–3756.
19. Ossareh-Nazari,B., Gwizdek,C. and Dargemont,C. (2001) Protein export from the nucleus. *Traffic*, **2**, 684–689.
20. Holaska,J., Black,B., Love,D., Hanover,J., Leszyk,J. and Paschal,B. (2001) Calreticulin is a receptor for nuclear export. *J. Cell Biol.*, **152**, 127–140.
21. Gama-Carvalho,M. and Carmo-Fonseca,M. (2001) The rules and roles of nucleocytoplasmic shuttling proteins. *FEBS Lett.*, **498**, 157–163.
22. Hope,T. (1999) The ins and outs of HIV. *Rev. Arch. Biochem. Biophys.*, **365**, 186–191.
23. Bogerd,H., Fridell,R., Benson,R., Hua,J. and Cullen,B. (1996) Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel *in vivo* randomization-selection assay. *Mol. Cell. Biol.*, **16**, 4207–4214.
24. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
25. Shannon,C.E. (1948) A mathematical theory of communication. *Bell System Tech. J.*, **27**, 379–423, 623–656.
26. Thompson,J., Higgins,D. and Gibson,T. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.