This document is the Accepted Manuscript version of a Published Work that appeared in final form in **Biochemical Journal** 441(1): 209-217 (2012), copyright © The Authors Journal compilation © 2012 Biochemical Society. To access the final edited and published work see https://doi.org/10.1042/BJ20111300

A GLOBAL SURVEY OF CRM1-DEPENDENT NUCLEAR EXPORT SEQUENCES IN THE HUMAN DEUBIQUITINASE FAMILY.

Iraia García-Santisteban[*], Sonia Bañuelos[†‡] Jose A. Rodríguez[*1]
*Department of Genetics, Physical Anthropology and Animal Physiology; †Department of Biochemistry and Molecular Biology, University of the Basque Country UPV/EHU; ‡Unidad de Biofísica CSIC-UPV/EHU.

[1]To whom correspondence should be addressed at the Dept. of Genetics, Physical Anthropology and Animal Physiology, School of Medicine, University of the Basque Country UPV/EHU, Bº Sarriena s/n, 48940 Leioa, Spain. Fax: +34 94 601 3400. E-mail josean.rodriguez@ehu.es

SYNOPSIS

The mechanisms that regulate the nucleocytoplasmic localization of human deubiquitinases remain largely unknown. The nuclear export recepor CRM1 binds to specific amino acid motifs termed nuclear export sequences (NESs). By using *in silico* prediction and experimental validation of candidate sequences, we identified 32 active NESs and 78 inactive NES-like motifs in human deubiquitinases. These data allowed us to evaluate the performance of three programs widely used for NES prediction, and to add novel information to the recently redefined NES consensus. The novel NESs identified here reveal a subset of 22 deubiquitinases bearing motifs that might mediate their binding to CRM1. We tested the effect of the CRM1 inhibitor leptomycin B (LMB) on the localization of YFP- or GFP-tagged versions of six NES-bearing deubiquitinases (USP1, USP3, USP7, USP21, CYLD and OTUD7B). YFP-USP21 and, to a lesser extent, GFP-OTUD7B relocated from the cytoplasm to the nucleus in the presence of LMB, revealing their nucleocytoplasmic shuttling capability. Two sequence motifs in USP21 had been identified during our survey as active NESs in the export assay. Using site-directed mutagenesis, we show that one of these motifs mediates USP21 nuclear export, whereas the second motif is not functional in the context of full length USP21.

Short title: Global survey of NES motifs in human deubiquitinases

Keywords: nuclear transport, CRM1, nuclear export, NES prediction, DUB, USP21

Abbreviations footnote: DUB, deubiquitinase; NES, nuclear export sequence; cNES, candidate nuclear export sequence; NLS, nuclear localization signal; GFP, green fluorescent protein; YFP, yellow fluorescent protein; LMB, leptomycin B; CHX, cycloheximide; ActD, actinomycin D; TP, true positive; TN, true negative; FP, false positive; FN, false negative; PPV, positive predictive value;

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

INTRODUCTION

Ubiquitylation, the covalent attachment of one or more units of the 76-amino acid peptide ubiquitin, is an important postranslational modification that regulates the levels, activity and localization of many cellular proteins [1]. Ubiquitylation is a reversible modification, and the removal of ubiquitin from substrates is catalyzed by a group of enzymes, termed deubiquitylating enzymes or deubiquitinases (DUBs) [2]. By specifically removing ubiquitin chains, DUBs modulate the ubiquitylation status of a variety of proteins, and thereby contribute to regulate important cellular processes, such as gene expression [3], proliferation [4], repair of DNA damage [5] and apoptosis [6].

The human genome encodes approximately 90 deubiquitinases, which can be classified into five different subclasses according to the structure of their catalytic domain [7]. A minority of human DUBs have been relatively well characterized in recent years. For example, USP7 has been shown to regulate the activity of the tumor suppressor proteins p53 [8], FOXO [9] and PTEN [10]. The involvement of USP7 and other deubiquitinases in processes and pathways related to tumor development has raised the possibility that these proteins might represent interesting targets for anticancer therapy [11, 12]. However, a significant hurdle towards the potential development of DUBs as therapeutic targets is the limited information still available on the basic biology of this family of proteins. In fact, despite recent advances, the physiological role and the mechanisms that regulate the function of most human DUBs are yet to be identified.

One of the aspects of DUB biology that remains poorly characterized is the regulation of their nucleocytoplasmic distribution, which, in turn, might critically regulate the accessibility of these enzymes to their substrates. In this regard, there is evidence that several DUBs may act on substrates located in both the nucleus and the cytoplasm. For example, USP21 catalyzes the deubiquitylation of the nuclear histone H2A [13] and the cytoplasmic protein Rip1 [14], but how the access of USP21 to its targets may be regulated is still unknown. Furthermore, it has been shown that DUB-mediated deubiquitylation may regulate the translocation of some substrates between the nucleus and cytoplasm [15]. However, very little is known about the mechanisms that may control the nucleocytoplasmic distribution of DUBs themselves. A systematic localization analysis of the 20 DUBs encoded by the fission yeast *Schizosaccharomyces pombe* has been recently reported [16], but few studies have specifically addressed the localization of DUBs in higher organisms [17-19].

Transport of proteins between the nucleus and the cytoplasm is carried out by specialized receptors that bind their cargoes in one compartment and escort them through the nuclear pore complex before releasing them in the other compartment. Transport receptors recognize specific amino acid motifs that function as targeting signals in the cargo proteins [20]. Thus, import receptors recognize and bind nuclear localization signals (NLSs), whereas export receptors bind to nuclear export sequences (NESs). “Classical” nuclear localization signals consist of one or two short stretches of basic amino acids [21], as exemplified by the SV40 large T antigen NLS (PKKKRKV). On the other hand, the best-characterized nuclear export sequences, termed “leucine-rich” NESs, generally consist of a stretch of hydrophobic amino acids with a poorly conserved “consensus” pattern [22]. This type of NES bridges the interaction of the cargo proteins with the export receptor CRM1 [23].

Some proteins can shuttle back and forth between nucleus and cytoplasm, either constitutively or in a regulated manner. Nucleocytoplasmic shuttling has long been recognized as an essential regulatory mechanism for many cellular proteins [24]. Most shuttling proteins identified to date exit the nucleus by interacting with the export receptor CRM1. Therefore, we reasoned that a comprehensive survey of CRM1-binding motifs (i.e. leucine-rich NESs) in human DUBs might represent a feasible strategy to begin to elucidate, on a whole-family scale, the mechanisms that control the nucleocytoplasmic distribution of these proteins.

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

A common approach to NES identification begins by examining the primary sequence of the protein(s) of interest using informatics programs, such as NetNES, ELM or NES Finder [25, 26], to pinpoint amino acid motifs that might constitute a nuclear export sequence. However, it has been shown that the NES-binding domain of CRM1 can accommodate a large variety of sequences, differing in both sequence and size [27-30] and, as a result, *in silico* prediction of candidate NESs (cNESs) is a challenging task. In this context, it is important to note that, as far as we know, the accuracy of the different programs available for NES prediction has not yet been directly compared in a prospective manner.

To identify sequence motifs in human deubiquitinases that might mediate their binding to CRM1, we used *in silico* prediction of cNESs in 85 human DUBs, followed by experimental testing of the predicted cNES in a nuclear export assay. With this two-step approach, we identified 32 novel functional NESs and 78 non-functional NES-like motifs in this protein family. This relatively large set of experimentally validated data allowed us to prospectively compare the accuracy of commonly used NES prediction programs, and to gain further insight into the amino acid features of nuclear export sequences.

On the other hand, our study reveals a subset of 22 human DUBs bearing one or more sequence motifs that might mediate their binding to the export receptor CRM1, raising the possibility that the function of some of these enzymes may be regulated by active nucleocytoplasmic shuttling. In fact, using fluorescently tagged proteins, we demonstrate the ability of YFP-USP21 and GFP-OTUD7B to shuttle between the nucleus and the cytoplasm. Finally, by using site-directed mutagenesis, we show that CRM1-dependent nucleocytoplasmic shuttling of YFP-USP21 is mediated by one of the novel NESs identified in our global survey. These findings reveal a mechanism that may regulate the dynamic access of this enzyme to its previously described nuclear and cytoplasmic substrates [13, 14].

EXPERIMENTAL

NES prediction using bioinformatic analysis

The amino acid sequence of each of 85 human deubiquitinases [7] was retrieved from NCBI (http://www.ncbi.nlm.nih.gov/protein), and submitted for analysis to three different web-based NES prediction servers: NetNES (http://www.cbs.dtu.dk/services/NetNES/), ELM (http://elm.eu.org/links.html) and NES Finder (http://research.nki.nl/fornerodlab/). Both ELM and NES Finder identify candidate NESs as linear sequence motifs resembling the “consensus” NES, although the specific amino acid pattern defining a candidate NES differs between the two programs. The output of these programs are short amino acid sequences predicted to be cNESs. The NetNES program, on the other hand, combines neural networks and hidden Markov models to calculate a “NES score” for each amino acid in the protein under analysis. The output of the NetNES program is a graphical representation of the NES score for each residue. Residues with a score exceeding the threshold value pre-determined by the program (0.5) are predicted to be part of an export sequence [25].

Plasmids, cloning procedures and site-directed mutagenesis

In order to evaluate the activity of the predicted DUB cNESs, a series of 110 expression plasmids based on the pRev(1.4)-GFP export assay reporter [31] were constructed. To this end, double-stranded DNA fragments encoding a 19-amino acid sequence, encompassing the amino acids predicted as cNES and several flanking residues, were generated by annealing of two complementary oligonucleotides and extension with Klenow enzyme (Fermentas). These DNA fragments were cloned into the pRev(1.4)-GFP vector (a gift from B. Henderson) using the BamHI/PinAI restriction sites, and confirmed by DNA sequencing. On the other hand, plasmids encoding several full length DUBs were generously provided by R. Bernards (USP1), PP. di Fiore (USP3), R. Everet (USP7), D. Barford (CYLD), and PC. Evans (OTUD7B), or obtained

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

from JW. Harper laboratory through Addgene (USP21). USP3, USP7 and USP21 cDNAs were amplified by PCR using high fidelity Pfu UltraII fusion HS DNA polymerase (Stratagene) and subcloned into pEYFP-C1 (Clontech) as HindIII/BamHI (USP3 and USP7) or XhoI/EcoRI (USP21) fragments.

The QuickChange Lightning Site-Directed Mutagenesis Kit (Stratagene) was used to introduce NES-inactivating mutations into full length YFP-USP21. The presence of the mutations was confirmed by DNA sequencing.

Cell culture, transfection and leptomycin B treatment

HeLa cells were grown in Dulbecco's modified Eagle's medium (Invitrogen), supplemented with 10% fetal bovine serum (Invitrogen), 100 U/ml penicillin and 100 µg/ml streptomycin (Invitrogen). Twenty four hours before transfection cells were seeded onto sterile glass coverslips in 12-well trays. Transfections were carried out with FuGene6 (Roche Diagnostics) following manufacturer´s protocol. Leptomycin B (Apollo Scientific) was added to the culture medium to a final concentration of 6 ng/ml for the indicated period of time.

Fluorescence microscopy analysis

Twenty four hours after transfection, cells expressing proteins tagged with GFP or YFP were fixed with 3.7% formaldehyde in phosphate-buffered saline (PBS) for 30 min, incubated with Hoechst 33285 (Sigma) to visualize the nuclei, washed with PBS, and mounted onto microscope slides using Vectashield (Vector).

Slides were examined using a Zeiss Axioskop fluorescence microscope, and images were taken with a Nikon Ds-Qi1Mc digital camera and the NIS-Elements F software.

In vivo nuclear export assay

The nuclear export assay was carried out essentially as reported [31]. Briefly, pRev(1.4)-GFP-based plasmids containing the candidate DUB NESs were transfected into HeLa cells. Survivin NES [32] and the empty pRev(1.4)-GFP were used as positive and negative controls, respectively. At 24 h post-transfection, cells were treated for 3 h with either 10 *µ*g/ml cycloheximide (CHX, Sigma) plus 5 *µ*g/ml actinomycin D (ActD, Sigma) or with CHX alone. CHX is added to ensure that cytoplasmic GFP arises from nuclear export and not from new protein synthesis, whereas ActD allows detection of weak NESs by preventing nuclear import mediated by Rev. The subcellular localization of the fluorescent proteins was determined in at least 200 cells per sample. Using the scoring system described in the original report [31], the activity of the functional NESs was rated between 1+ and 9+. The CRM1-dependence of the functional NESs identified in the nuclear export assay was confirmed by adding LMB (6 ng/ml) to the transfected cells.

Sequence aligment and modeling

Multiple sequence alignment of DUB cNESs was carried out with ClustalW2 [33], and edited to account for conserved hydrophobic positions.

The figures showing the structure of USP21 bound to ubiquitin were prepared with PyMol (The PyMol Molecular Graphics System; Schrödinger, LLC).

RESULTS

***In silico* prediction of candidate nuclear export sequences in human deubiquitinases**

In order to identify sequence motifs that might mediate the binding of human DUBs to the export receptor CRM1 we took the two-step approach illustrated in Figure 1A: *in silico* prediction of

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

candidate NES sequences (cNESs), followed by experimental testing of a subset of the predicted cNESs.

In the first step, the primary amino acid sequence of 85 human DUBs was analysed using three web-based NES prediction programs: NetNES [25], ELM [26] and NES Finder. As summarized in Figure 1B, and detailed in Supplementary Table 1, 75 candidate NESs were identified in the amino acid sequence of human DUBs using the NetNES program, 71 using the ELM program, and 351 using the NES Finder program. In total, 428 different candidate NESs were identified in the analysis. Even considering that these programs use different approaches for NES prediction (see *Experimental* section), the limited degree of coincidence in the cNESs predicted in human DUBs by NetNES, ELM and NES Finder was remarkable. As an example of this lack of agreement, Figure 1C shows the six different cNESs predicted in USP1. The amino acid motifs within this protein identified as candidate NESs by the different programs did not overlap. Only 63 of the 428 cNESs motifs identified were predicted by more than one program, and only 6 were predicted by the three programs.

Experimental testing of candidate DUB NESs

In order to evaluate the activity of candidate NESs, we used the pRev(1.4)-GFP-based nuclear export assay [31]. Rev(1.4)-GFP is a chimeric protein resulting from the fusion of an export deficient mutant of the HIV Rev protein and GFP, which localizes to nucleoli. In the assay, active NESs are identified on the basis of their ability to mediate export of Rev(1.4)-GFP to the cytoplasm. Actinomycin D (ActD), which disrupts nucleoli and blocks nuclear import mediated by Rev NLS, can be added to reveal the activity of weaker NESs. Besides identifying active export sequences, the assay allows for the comparison of the level of activity of the different NESs, based on the degree of cytoplasmic accumulation of the Rev(1.4)-[NES]-GFP chimeric proteins in the presence/absence of ActD.

The high number of candidate DUB NESs predicted *in silico* made it impractical to attempt the functional testing of all of them. Thus, we aimed to test the 63 candidate NES predicted by more than one program (Figure 1B), and a subset of the cNES predicted by single programs. This subset included, on one hand, the 16 sequences predicted only by NetNES with a "NES score" higher than 0.7. On the other hand, we selected the first 16 sequences predicted only by ELM, and the first 16 sequences predicted only by NES Finder from the list of predicted DUB cNESs (Supplementary Table 1). In addition, the recent report that USP10 may undergo CRM1-dependent nuclear export [34], prompted us to test three cNESs predicted by the NES Finder program in this protein.

Despite repeated attempts, four sequences could not be cloned in the pRev(1.4)-GFP vector and thus, the final number of DUB cNESs tested was 110. Figure 2 shows representative examples of the results obtained in the export assay. The pRev(1.4)-GFP empty vector was used as negative control, and a previously identified NES in Survivin [32] was included as a positive control. As shown in Figure 2, the DUB cNESs FIN6, ELM33 and ELM3, but not FIN1, were capable of relocating Rev(1.4)-GFP to the cytoplasm. This cytoplasmic relocation was efficiently reverted by the specific CRM1 inhibitor leptomycin B (LMB). These results, therefore, validated the sequences FIN6, ELM33 and ELM3 as active CRM1-dependent NESs. Using the nuclear export assay scoring system, ELM3 was assigned the highest export activity (9+), whereas FIN6 (3+) and ELM33 (7+) were weaker NESs in this experimental setting. Supplementary Table 2 provides a detailed account of the results obtained with the 110 DUB cNESs tested, including the export activity score of those sequences that tested positive in the assay. In total, 32 of the sequences tested were validated as functional NESs, whereas 78 sequences (including the three USP10 cNESs) were inactive and therefore, represent non-functional NES-like motifs.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accuracy of NES prediction programs

Table 1 summarizes the results of the export assay in relation to the program(s) that predicted each cNES. Of note, the percentage of positive hits among the 110 cNESs tested was considerably higher for those cNESs predicted by multiple programs than for those predicted by a single program (39.3% *versus* 16.3%).

In order to use these data to compare the accuracy of the prediction programs, we first extrapolated the percentage of experimentally validated NESs in each category to estimate the number of positive hits expected if the 428 predicted cNESs had been tested ("Estimated total positive" column in Table 1). Next, a two-by-two contingency matrix was generated for each program, with four possible outcomes for a given sequence (Figure 3A). A sequence predicted as cNES by the program constitutes a "true positive" (TP) if it was positive in the export assay, and a "false positive" (FP) if it tested negative. A sequence not predicted as cNES by the program (but predicted as cNES by at least one of the other two programs) constitutes a "true negative" (TN) if it was negative in the export assay and a "false negative" (FN) if it tested positive. The sensitivity, specificity and positive predictive value (PPV) of each program were calculated as indicated in Figure 3A. It must be noted that, in order to calculate these accuracy metrics, we needed to make the assumption that every potential NES in the amino acid sequence of human DUBs was predicted as cNES by at least one of the programs (i.e. that all the possible "true negatives" and "false negatives" were taken into account).

The contingency matrices for the NetNES, ELM and NES Finder programs are shown in Figure 3B. The data included in these matrices are derived from the "cNESs predicted" and "Estimated total positive" columns of Table 1. The accuracy metrics of the three programs calculated from these matrices are shown in Table 2. According to our estimation, the ELM program demonstrated the highest positive predictive value (38%), whereas the lowest PPV (10.8%) corresponded to the NES Finder program (Table 2).

Amino acid sequence features of DUB cNESs

In an attempt to gain further insight into the sequence determinants of CRM1-dependent NESs, we used sequence alignment methodology to examine the amino acid features of the tested 110 DUB cNESs in the light of the NES consensus redefined recently on the basis of novel structural information [30].

Alignment of those sequences that tested positive in the nuclear export assay (Figure 4A) highlighted some of the previously described common features of CRM1-binding motifs [27, 30, 35], including the presence of hydrophobic (frequently leucine) residues at four or five positions (Φ^0-Φ^4), and a prevalence of acidic residues in the vicinity of Φ^0. Interestingly, our analysis revealed a further common characteristic of many DUB NESs: the hydrophobic nature of the residue previous to Φ^2 in several of these sequences. We also noted that, although the presence of 2 residues between the Φ^2 and Φ^3 hydrophobic residues (Φ^2-X_2-Φ^3 spacing) has been reported to be preferred [30], a significant proportion of validated DUB NESs showed a Φ^2-X_3-Φ^3 spacing. In spite of sharing some common characteristics, five different subsets of positive DUB NESs with distinguishing features could be established. Thus, 11 out of 32 positive NES sequences (P-I group in Figure 4A) followed strictly the Φ^0-Φ^4 consensus defined by Güttler et al., whereas a second group of 10 sequences, also with a Φ^0-Φ^4 pattern (P-II), was characterized by an alternative Φ^0-X_3-Φ^1 spacing instead of the reported Φ^0-X_2-Φ^1. Seven other sequences conformed a group of NESs lacking the leading hydrophobic residue Φ^0 and therefore, having a Φ^1-Φ^4 pattern (P-III). A single sequence, NET8, conformed to the Rev-type consensus (P-IV). Finally, two experimentally validated sequences, predicted only by the NetNES program, did not fit any of the consensus mentioned above (P-V).

Not surprisingly, many of the predicted cNESs sequences that tested negative in the export assay displayed several common NES features. In fact, 20 of the 78 non-functional NES-like

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Accepted Manuscript

motifs followed the Φ^0-Φ^4 consensus pattern, with either the Φ^0-X_2-Φ^1 or the alternative Φ^0-X_3-Φ^1 spacings (Figure 4B). Twelve of these sequences (N-I group) had proline residues between Φ^1 and Φ^3, which would prevent them from adopting the proper conformation for CRM1 binding [27, 30], and would therefore explain their lack of export activity. However, there is no such an obvious reason for the lack of activity of other 8 cNESs with a Φ^0-Φ^4 pattern (N-II).

A subset of human deubiquitinases bear active NESs

The 32 experimentally validated cNESs identified in our assay reveal a subset of 22 human DUBs bearing amino acid motifs with the potential to mediate their binding to the export receptor CRM1. Figure 5 shows the location of these motifs in the corresponding deubiquitinase. A single NES was identified in most DUBs, but two large proteins, USP9X and USP24, were found to bear four and five NESs, respectively. This observation is in line with previous findings that several independent NESs may contribute to the nuclear export of large proteins such as FANCA or pericentrin [36, 37].

Several DUB NESs were located near the amino-terminal (e.g. USP1, USP15) or carboxy-terminal (e.g. USP7, OTUD4) ends of the protein. Other NESs, such as those identified in OTUD7A and OTUD7B, were located in the middle of the protein. The location of these sequences within the full length protein may be a relevant feature, as it has been proposed that NESs located near the amino- or carboxy-terminal end of a protein are more likely to be operational in their physiological context [28, 30].

The presence of one or more active NESs in their primary sequence suggests that some of these DUBs may undergo active CRM1-mediated nuclear export. It is important to note, however, that the activity of the reported NESs, thus far validated in the context of the Rev(1.4)-GFP chimeric protein, needs to be assessed in the context of the full length DUB, to determine their physiological importance.

Effect of CRM1 inhibition on the localization of NES-bearing DUBs

In order to begin addressing the potential role of these sequences in the context of their cognate full length proteins, we investigated the effect that inhibition of the CRM1-mediated export pathway has on the nucleocytoplasmic distribution of several NES-bearing human DUBs.

Given the limited availability of DUB-specific commercial antibodies validated for immunofluorescence, we decided to use GFP or YFP-tagged versions of USP1, USP3, USP7, USP21, CYLD and OTUD7B to determine their nucleocytoplasmic distribution in transfected cells. As shown in Figure 6A, GFP-USP1, YFP-USP3 and YFP-USP7 showed a prominent nuclear localization, whereas YFP-USP21, GFP-CYLD and GFP-OTUD7B were predominantly located in the cytoplasm. These results are consistent with previous reports describing the steady-state localization of some of these proteins [38-42]. Incubation with the specific CRM1 inhibitor LMB for 3 hours did not alter the distribution of the nuclear DUBs or that of GFP-CYLD, which remained in the cytoplasm. In contrast, LMB induced a clear relocation of the other two cytoplasmic DUBs, YFP-USP21 and GFP-OTUD7B, to the nucleus, revealing their ability to shuttle between nucleus and cytoplasm, and indicating that CRM1-mediated export plays a role in maintaining the cytoplasmic localization of these two DUBs. A time-course, semi-quantitative analysis (Figure 6B) showed that YFP-USP21 displayed a more rapid and pronounced response to LMB than GFP-OTUD7B. This observation, and the previously reported findings that USP21 may deubiquitinate both nuclear and cytoplasmic substrates [13, 14] prompted us to select this protein for further analysis.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

CRM1-dependent nucleocytoplasmic shuttling of USP21 is mediated by its amino-terminal NES

We aimed to identify the sequence determinants of USP21 nuclear export, focusing on the two active NES motifs identified in this protein by our global survey. As shown in Figure 7A, one of these motifs (ELM20) is located close to the amino-terminal end of the protein, whereas the second one (NET20) is located in the middle of the protein, embedded within its catalytic USP domain. The activity score of these motifs in the nuclear export assay was 4+ and 9+, respectively. Mutant versions of these motifs were generated, where the Φ^3 and Φ^4 were changed to alanine, and tested in the export assay (Figure 7B). The results confirmed that these mutations effectively abrogate the export activity of ELM20 and NET20 motifs (Figure 7B). Next, the corresponding point mutations (L142/L144A to inactivate ELM20, and L313/L315A to inactivate NET20) were separately introduced into full length YFP-USP1 using site-directed mutagenesis (Figure 7C), and the nucleocytoplasmic localization of the mutant proteins was examined. As shown in Figure 7C, the YFP-USP21$^{L142/L144A}$ mutant displayed a predominantly nuclear localization, whereas the YFP-USP21$^{L313/L315A}$ mutant was located in the cytoplasm, like the wild type protein.

These results indicate that the amino-terminal NES motif of USP21 (134-ELGAALSRLALRPEPPTLR-152) is both necessary and sufficient to ensure CRM1-mediated export of the protein, and represents a physiologically relevant NES. In order to also seek potential determinants of its nuclear entry, we examined USP21 sequence for candidate nuclear localization signals (cNLSs) using the PSORTII [43] and cNLS Mapper [44] programs. Each program identified one cNLS near the amino terminal end of USP21 (supplementary Figure 1). However, deletion of a 70 amino acid region encompassing both cNLSs reduced, but did not abrogate, LMB-induced nuclear entry of YFP-USP21(71-565). This result suggests that, unlike its NES-mediated nuclear export, nuclear import of USP21 may be mediated by multiple sequence determinants. These may include N-terminal basic NLSs, but also potential binding domains for other NLS-containing proteins that may contribute to USP21 nuclear entry through a "piggyback" mechanism.

DISCUSSION

Human DUBs are emerging as critical regulators of several important cellular processes [3-6], and potential therapeutic targets [11, 12]. However, many aspects of DUB biology remain largely unknown, including the mechanisms that regulate their subcellular localization. As a first step towards characterizing these mechanisms, we report here the results of a family-wide survey of active CRM1-dependent NESs in human DUBs.

Using three web-based prediction programs to analyse the primary amino acid sequence of 85 human DUBs, we identified 428 sequence motifs as candidate NESs, 110 of which were subsequently tested using a nuclear export assay [31]. To our knowledge, this study represents the first functional analysis of a relatively large series of candidate NESs in a protein family. Thirty two of the 110 cNESs tested were active in the assay, whereas 78 sequences were inactive. Three of the candidate NES motifs found to be inactive (ELM4, FIN15, and NET37) encompassed sequences previously tested by other groups [17, 18]. Remarkably, whereas the lack of activity of ELM4 and NET37 reported here was consistent with previous findings, a short amino acid sequence (VEVYLLELKL) included within the FIN15 motif was previously reported to be a weak, but active NES [17].

Only 63 of the 428 cNES motifs identified *in silico* were predicted by more than one program, illustrating the serious challenge posed by bioinformatic NES identification. In this regard, the performance of currently available NES prediction programs (NetNES, ELM and NES Finder) has never been, as far as we know, prospectively compared. We carried out such

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

a comparison by calculating the sensitivity, specificity and positive predictive value (PPV) of these programs. It is important to note that these accuracy metrics, although useful, have intrinsic limitations [45], and also that several assumptions needed to be made in order to calculate them. Therefore, the obtained values should be regarded as an estimation. With these caveats in mind, our analysis points to ELM as the best-performing tool for NES prediction among the three tested here. Even when using ELM, however, our estimation is that less than 40% of the predicted cNES are active in a functional assay. Interestingly, we noticed a higher percentage of positive hits among the cNESs predicted by multiple programs than among those predicted by a single program. Thus, combining the results of different prediction programs might be a simple approach to prioritize for testing those cNESs most likely to be functional.

The difficulty to identify CRM1-dependent NESs *in silico* stems largely from the ability of the receptor to accomodate a large variety of sequences into its NES-binding hydrophobic groove. Recent data have led to revise the NES consensus to include 5 hydrophobic residues (Φ^0-Φ^4 pattern) [30]. To gain further understanding on the characteristics of CRM1-dependent NESs, we examined the amino acid features of the novel set of 32 active NESs and 78 inactive NES-like motifs identified in human DUBs,.

Consistent with the redefined consensus, most experimentally validated DUB NESs (21 out of 32) fit a Φ^0-Φ^4 pattern. Importantly, nearly half of the active NESs with this pattern showed a 3 amino acid spacing between Φ^0 and Φ^1. This Φ^0–X_3-Φ^1 spacing would be compatible with the conformation adopted by the NES sequences from snurportin [29] and PKI [30], and might even favor a more canonical geometry of the first turn of the α-helix. Therefore, we propose that the possibility of a Φ^0–X_3-Φ^1 spacing should be included in the NES consensus. Further illustrating the ability of CRM1 to bind a wide range of sequences, 11 experimentally validated NESs failed to fit the Φ^0-Φ^4 consensus, having instead a Φ^1-Φ^4 pattern, a Rev-type pattern, or failing to resemble any of these patterns. Conversely, several cNESs bearing a Φ^0-Φ^4 pattern tested negative. The inactivity of some of these motifs could be ascribed to the presence of proline residues between Φ^1 and Φ^3, as previously reported [27, 30]. We speculate that the lack of activity of other NES-like motifs could be due to a combination of several “unfavorable” features, such as having a suboptimal set of hydrophobic residues [30], and/or lacking acidic amino acids in the vicinity of Φ^0. The 78 experimentally confirmed inactive NES-like motifs may be useful as a “negative control group” in the future refining of NES prediction algorithms.

The identification of sequence motifs that might bridge their interaction with CRM1 raises the possibility that a subset of human DUBs may be actively exported to the cytoplasm by this receptor. To begin addressing this possibility, we evaluated the nucleocytoplasmic distribution of six GFP- or YFP-tagged human DUBs in the presence or absence of the specific CRM1 inhibitor LMB. USP1, USP3 and USP7 showed a prominently nuclear localization in both untreated and LMB-treated cells, whereas CYLD remained in the cytoplasm even in the presence of LMB. Although these observations do not necessarily rule out the possibility that these DUBs are exported by CRM1, they suggest that other mechanisms may be more relevant in determining their final nucleocytoplasmic distribution. These mechanisms may include rapid nuclear import of USP1, USP3 and USP7, and cytoplasmic retention of CYLD, probably mediated by its B box domain [41]. Since human DUBs are known to be part of multiprotein complexes [46], retention in the nucleus or cytoplasm by interacting partners might be a common factor modulating their localization.

Importantly, LMB induced the nuclear relocation of YFP-USP21 and, to a lesser extent, of GFP-OTUD7B, thus identifying them as novel CRM1-dependent shuttling proteins. In particular, the rapid and pronounced nuclear accumulation of YFP-USP21 suggests that CRM1-dependent shuttling may be a relevant regulatory mechanism for this protein, facilitating the dynamic access of this enzyme to its nuclear and cytoplasmic substrates [13, 14]. Two USP21 sequence motifs (ELM20: 134-ELGAALSRLALRPEPPTLR-152 and NET20: 301-

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

DAQEFLKLLMERLHLEINR-319) tested positive in the nuclear export assay. By using site-directed mutagenesis in the context of the full length protein, we show that the ELM20 motif constitutes a physiologically relevant NES, whose mutational inactivation disrupts USP21 nuclear export. In contrast, mutation of the NET20 motif, which is embedded within USP21 catalytical USP domain, does not alter USP21 localization. The finding that the NET20 motif appeared to be irrelevant for the export of full length USP21 was surprising, given the high activity score of this motif in the export assay. This experimental observation, however, can be easily reconciled with the available crystal structure of USP21 USP domain bound to ubiquitin [47]. As shown in Figure 7D, the region corresponding to NET20 (residues 301-319) is relatively buried in the structure and partly occluded by the ubiquitin moiety. More importantly, as illustrated in Figure 7E, several residues that should make contacts with CRM1 (such as M310, corresponding to the Φ^2 position of the NES consensus) are engaged in packing interactions within the hydrophobic core of the protein. Furthermore, interaction between USP21 E304 and ubiquitin R72 has recently been shown to be essential for USP21-mediated deubiquitylation [48]. These interactions would, therefore, render the NET20 motif unable to mediate USP21/CRM1 binding. Although there is no structural information available for the amino-terminal region of USP21 where the ELM20 sequence resides, our experimental data suggest that this NES is exposed and available for CRM1 binding. These observations underscore the need to experimentally validate, in a case-by-case basis, the role that the NES motifs identified by our survey play in the localization of each DUB. It is conceivable that some of these motifs, like NET20, may be masked or fail to acquire the proper conformation for CRM1 binding in their original protein context. Others, like ELM20, will constitute physiologically relevant NESs.

In summary, the results presented here provide novel basic information relevant to the field of NES definition and prediction, and provide a starting point for the analysis of the mechanisms that regulate the nucleocytoplasmic distribution of human DUBs.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

REFERENCES

1. Grabbe, C., Husnjak K. and Dikic, I. (2011) The spatial and temporal organization of ubiquitin networks. Nat. Rev. Mol. Cell Biol. **12,** 295-307
2. Komander, D., Clague, M. J. and Urbé, S. (2009) Breaking the chains: structure and function of the deubiquitinases. Nat. Rev. Mol. Cell Biol. **10,** 550-563
3. Frappier, L. and Verrijzer, C. P. (2011) Gene expression control by protein deubiquitinases. Curr Opin Genet Dev. **21,** 207-213
4. Song, L. and Rape, M. (2008) Reverse the curse--the role of deubiquitination in cell cycle control. Curr Opin Cell Biol. **20,** 156-163
5. Bergink, S. and Jentsch, S. (2009) Principles of ubiquitin and SUMO modifications in DNA repair. Nature **458,** 461-467
6. Ramakrishna, S., Suresh, B. and Baek, K. H. (2011) The role of deubiquitinating enzymes in apoptosis. Cell Mol Life Sci. **68,** 15-26
7. Nijman, S. M., Luna-Vargas, M. P., Velds, A., Brummelkamp, T. R., Dirac, A. M., Sixma, T. K. and Bernards, R. (2005) A genomic and functional inventory of deubiquitinating enzymes. Cell **123,** 773-786
8. Li, M., Chen, D., Shiloh, A., Luo, J., Nikolaev, A. Y., Qin, J. and Gu, W. (2002) Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. Nature **416**, 648-653
9. van der Horst, A., de Vries-Smits, A. M., Brenkman, A. B., van Triest, M. H., van den Broek, N., Colland, F., Maurice, M. M. and Burgering, B. M. (2006) FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP Nat. Cell Biol. **8**, 1063-1073
10. Song, M. S., Salmena, L., Carracedo, A., Egia, A., Lo-Coco, F., Teruya-Feldstein, J. and Pandolfi, P. P. (2008) The deubiquitinylation and localization of PTEN are regulated by a HAUSP-PML network. Nature **455**, 813-817
11. Nicholson, B., Marblestone, J. G., Butt ,T. R. and Mattern, M. R. (2007) Deubiquitinating enzymes as novel anticancer targets. Future Oncol. **3,** 191-199
12. Hoeller, D. and Dikic, I. (2009) Targeting the ubiquitin system in cancer therapy. Nature **458,** 438-444
13. Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M. and Ito, T. (2008) Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes Dev. **22,** 37-49
14. Xu, G., Tan, X., Wang, H., Sun, W., Shi, Y., Burlingame, S., Gu, X., Cao, G., Zhang, T., Qin, J. and Yang, J. (2010) Ubiquitin-specific peptidase 21 inhibits tumor necrosis factor alpha-induced nuclear factor kappaB activation via binding to and deubiquitinating receptor-interacting protein 1 J. Biol. Chem. **285,** 969-978
15. Salmena, L. and Pandolfi, P. P. (2007) Changing venues for tumour suppression: balancing destruction and localization by monoubiquitylation Nat. Rev. Cancer **7,** 409-413
16. Kouranti, I., McLean, J. R., Feoktistova, A., Liang, P., Johnson, A. E., Roberts-Galbraith, R. H. and Gould, K. L. (2010) A global census of fission yeast deubiquitinating enzyme localization and interaction networks reveals distinct compartmentalization profiles and overlapping functions in endocytosis and polarity. PLoS Biol. **8,** e1000471
17. Soboleva, T. A., Jans, D. A., Johnson-Saliba, M. and Baker, R. T. (2005) Nuclear-cytoplasmic shuttling of the oncogenic mouse UNP/USP4 deubiquitylating enzyme. J. Biol. Chem. **280**, 745-752
18. Macedo-Ribeiro, S., Cortes, L., Maciel, P. and Carvalho, A. L. (2009) Nucleocytoplasmic shuttling activity of ataxin-3. PloS One **4,** e5834

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

19. Pastori, V., Sangalli, E., Coccetti, P., Pozzi, C., Nonnis, S., Tedeschi, G. and Fusi P. (2010) CK2 and GSK3 phosphorylation on S29 controls wild-type ATXN3 nuclear uptake. Biochem. Biophys. Acta. **1802**, 583-592
20. Görlich, D. and Kutay U. (1999) Transport between the cell nucleus and the cytoplasm. Annu. Rev. Cell Dev. Biol. **15,** 607-660
21. Lange, A., Mills, R. E., Lange, C. J., Stewart, M., Devine, S. E. and Corbett, A. H. (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. J. Biol. Chem. **282,** 5101-5105
22. Mattaj, I. W. and Müller, C. W. (2010) Solving the NES problem. Nat. Struct. Mol. Biol. **17,** 1288-1289
23. Hutten, S. and Kehlenbach, R. H. (2007) CRM1-mediated nuclear export: to the pore and beyond. Trends Cell Biol. **17,** 193-201
24. Gama-Carvalho, M. and Carmo-Fonseca, M. (2001) The rules and roles of nucleocytoplasmic shuttling proteins. FEBS Lett. **498,** 157-163
25. la Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K. and Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. Protein Eng. Des. Sel.**17,** 527-536
26. Gould, C. M., Diella, F., Via, A., Puntervoll, P., Gemünd, C., Chabanis-Davidson, S., Michael, S., Sayadi, A., Bryne, J. C., Chica, C., Seiler, M., Davey, N. E., Haslam, N., Weatheritt, R. J., Budd, A., Hughes, T., Pas, J., Rychlewski, L., Travé, G., Aasland, R., Helmer-Citterich, M., Linding, R. and Gibson, T. J. (2010) ELM: the status of the 2010 eukaryotic linear motif resource. Nucleic Acids Res. **38,** D167-180
27. Kosugi, S., Hasebe, M., Tomita, M. and Yanagawa, H. (2008) Nuclear export signal consensus sequences defined using a localization-based yeast selection system Traffic **9,** 2053-2062
28. Dong, X., Biswas, A., Süel, K. E., Jackson, L. K., Martinez, R., Gu, H. and Chook, Y. M. (2009) Structural basis for leucine-rich nuclear export signal recognition by CRM1. Nature **458,** 1136-1141
29. Monecke, T., Güttler, T., Neumann, P., Dickmanns, A., Görlich, D. and Ficner, R. (2009) Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. Science **324,** 1087-1091
30. Güttler, T., Madl, T., Neumann, P., Deichsel, D., Corsini, L., Monecke, T., Ficner, R., Sattler, M. and Görlich, D. (2010) NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. Nat Struct Mol Biol. **17,** 1367-1376
31. Henderson, B. R. and Eleftheriou, A. (2000) A comparison of the activity, sequence specificity, and CRM1-dependence of different nuclear export signals. Exp Cell Res. **256,** 213-224
32. Engelsma, D., Rodríguez, J. A., Fish, A., Giaccone, G. and Fornerod, M. (2007) Homodimerization antagonizes nuclear export of survivin. Traffic **8,** 1495-1502
33. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23**, 2947-2948
34. Yuan, J., Luo, K., Zhang, L., Cheville, J. C. and Lou, Z. (2010) USP10 regulates p53 localization and stability by deubiquitinating p53. Cell **140,** 384-396
35. Kutay, U. and Güttinger, S. (2005) Leucine-rich nuclear-export signals: born to be weak. Trends. Cell Biol. **15**, 121-124.
36. Ferrer, M., Rodriguez, J. A., Spierings, E. A., de Winter, J. P., Giaccone, G. and Kruyt, F. A. E. (2005) Identification of multiple nuclear export sequences in

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Fanconi anemia group A protein that contribute to CRM1-dependent nuclear export. Hum Mol Genet **14**, 1271-1281

37. Liu, Q., Yu, J., Zhuo, X., Jiang, Q. and Zhang, C. (2010) Pericentrin contains five NESs and an NLS essential for its nucleocytoplasmic trafficking during the cell cycle. Cell Res. **20**, 948-962
38. Nijman, S. M., Huang, T. T., Dirac, A. M., Brummelkamp, T. R., Kerkhoven, R. M., D'Andrea, A. D. and Bernards, R. (2005) The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. Mol. Cell. **17,** 331-339
39. Nicassio, F., Corrado, N., Vissers, J. H., Areces, L. B., Bergink, S., Marteijn, J. A., Geverts, B., Houtsmuller, A. B., Vermeulen, W., Di Fiore, P. P. and Citterio, E. (2007) Human USP3 is a chromatin modifier required for S phase progression and genome stability. Curr. Biol. **17,** 1972-1977
40. Fernández-Montalván, A., Bouwmeester, T., Joberty, G., Mader, R., Mahnke, M., Pierrat, B., Schlaeppi, J. M., Worpenberg, S. and Gerhartz, B. (2007) Biochemical characterization of USP7 reveals post-translational modification sites and structural requirements for substrate processing and subcellular localization. FEBS J. **274,** 4256-4270
41. Komander, D., Lord, C. J., Scheel, H., Swift, S., Hofmann, K., Ashworth, A. and Barford, D. (2008) The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B box module. Mol. Cell **29**, 451-464
42. Enesa, K., Zakkar, M., Chaudhury, H., Luong, le A., Rawlinson, L., Mason, J. C., Haskard, D. O., Dean, J. L. and Evans, P. C. (2008) NF-kappaB suppression by the deubiquitinating enzyme Cezanne: a novel negative feedback loop in pro-inflammatory signaling. J. Biol. Chem. **283,** 7036-7045
43 Nakai, K. and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. **24**, 34-36.
44 Kosugi, S., Hasebe, M., Tomita, M. and Yanagawa, H. (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. U S A. **106**, 10171-10176.
45. Baldi, P., Brunak, S., Chauvin, Y., Andersen C. A. F. and Nielsen, H. (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics **16**, 412-424
46. Sowa, M. E., Bennett, E. J., Gygi, S. P. and Harper, J. W. (2009) Defining the human deubiquitinating enzyme interaction landscape. Cell. **138**, 389-403.
47. Neculai, D., Avvakumov, G. V., Walker, J. R., Xue, S., Butler-Cole, C., Weigelt, J., Bountra, C., Edwards, A. M., Arrowsmith, C. H., Bochkarev, A. and Dhe-Paganon, S., Structural Genomics Consortium (SGC), (2009) PDB ID 3I3T. DOI:10.2210/pdb3i3t/pdb
48. Ye, Y., Akutsu, M., Reyes-Turcu, F., Enchev, R. I., Wilkinson, K. D. and Komander, D. (2011) Polyubiquitin binding and cross-reactivity in the USP domain deubiquitinase USP21. EMBO Rep. **12,** 350-357

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

AUTHOR CONTRIBUTION

Iraia García-Santisteban participated in the design of the study and performed experiments. Sonia Bañuelos analysed and interpreted the data and contributed to write the manuscript. Jose Antonio Rodríguez participated in the design of the study, performed experiments, analysed and interpreted the data, and wrote the paper.

ACKNOWLEDGEMENTS

We are very grateful to Ana Zubiaga for her continuous support and encouragement, and to JL. Zugaza for critical comments on the manuscript. We appreciate the generous gift of plasmids from the following investigators: B. Henderson, R. Bernards, PP. di Fiore, R. Everet, D. Barford, PC. Evans and JW. Harper. We appreciate the technical support by the staff from the High Resolution Microscopy Facility (Sgiker-UPV/EHU), and the DNA Sequencing Facility (Sgiker-UPV/EHU).

FUNDING

This work was supported by grants from the Basque Government (SAIOTEK S-PE07UN17, S-PE09UN65 and ETORTEK BioGUNE2010) and the Spanish Government (BFU2009-13245) to J.A.R, and a fellowship from the Department of Education of the Basque Government to I.G-S.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

TABLES

Table 1. Results of the nuclear export assay on human DUB cNESs in relation to the program(s) that identified each sequence as a candidate NESs. * †

Program(s)*	cNESs predicted	Export assay results: positive/tested (%)	Estimated total positive†
Multiple programs			
N+E+F	6	5/6 (83.3)	5
N+E	1	1/1 (100)	1
N+F	25	8/24 (33.3)	8
E+F	31	10/30 (33.3)	10
Single program			
N	43	2/15 (13.3)	6
E	33	5/15 (33.3)	11
F	289	1/19 (5.2)	15
Total multiple programs	63	24/61 (39.3)	24
Total single program	365	8/49 (16.3)	32
Total	**428**	**32/110 (29)**	**56**

*N+E+F: cNESs predicted by the three programs; N+E: cNESs predicted by NetNES and ELM; N+F: cNESs predicted by NetNES and NES Finder; E+F: cNESs predicted by ELM and NES Finder; N: cNESs predicted only by NetNES; E: cNESs predicted only by ELM; F: cNESs predicted only by NES Finder.
†Positive hits expected if every predicted cNESs had been tested, estimated by extrapolating the percentage of experimentally validated NES in each category.

Table 2. Accuracy metrics calculated for the three NES prediction programs used in our study.

Program	Sensitivity	Specificity	PPV*
NetNES	35.7	85.2	26.6
ELM	48.2	88.1	38
NES Finder	67.8	15.8	10.8

*PPV: positive predictive value.

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

FIGURE LEGENDS

Figure 1. *In silico* prediction of candidate nuclear export sequences in human deubiquitinases.
(**A**). Flowchart of the two-step approach followed in our study to identify active CRM1-dependent NESs in human deubiquitinases. (**B**). Venn diagram illustrating the overlap between the candidate NESs (cNESs) predicted by the NetNES, ELM and NES Finder programs in the primary sequence of human DUBs. The total number of cNES predicted by each program is indicated between brackets, and the total number of different cNESs predicted is indicated on the lower right corner. (**C**). Schematic representation of human USP1 (left) detailing the location of six different amino acid motifs (white boxes) predicted to be cNES in this protein. The program responsible for the prediction, as well as the position of the first and last residues of each motif, is indicated. On the right, sequence of the amino acid motifs identified as USP1 cNES by the different programs.

Figure 2. Experimental testing of DUB cNESs using the Rev(1.4)-GFP nuclear export assay in HeLa cells.
Images show representative examples of HeLa cells transfected with pRev(1.4)-GFP-derived plasmids containing four different DUB cNESs. The empty pRev(1.4)-GFP vector was used as negative control, and a plasmid containing the active NES of Survivin was used as positive control. Cycloheximide (CHX) was used to prevent protein synthesis, ensuring that cytoplasmic GFP arises from export and not from newly synthesized protein. Actinomycin D (ActD) was used to disrupt nucleoli and block Rev NLS-mediated import, thus allowing detection of weaker NESs. Leptomycin B (LMB) was used to block CRM1-dependent nuclear export. The FIN1 cNES was unable to relocate Rev(1.4)-GFP to the cytoplasm, even in the presence of ActD, and was therefore classified as negative. The FIN6, ELM3 and ELM3 cNESs tested positive in the assay, leading to different degrees of Rev(1.4)-GFP cytoplasmic accumulation. Each positive NES was assigned an export activity score, indicated between brackets, by counting more than 200 cells per sample in the presence or absence of ActD. LMB treatment readily prevented the cytoplasmic relocation of Rev(1.4)-GFP induced by the FIN6, ELM33 and ELM3 sequences, demonstrating their dependence on the CRM1 export pathway.

Figure 3. Accuracy of the NetNES, ELM and NES Finder programs for NES prediction.
(**A**) Calculation of the program accuracy metrics. Example of a general two-by-two contingency matrix used to relate the results of the nuclear export assay to the cNES prediction produced by each program. Four outcomes are posible for each of the 428 predicted cNES sequences in human DUBs: true positive (TP) if predicted by the program and active in the assay; false positive (FP), if predicted by the program, but inactive in the assay; false negative (FN) if not predicted by the program, but active in the assay; and true negative (TN) if not predicted by the program and inactive in the assay. The TP, TN, FN, and FP values were used to calculate the sensitivity, specificity and positive predictive value (PPV) of the prediction program by applying the formulae indicated to the right of the matrix. (**B**) Contingency matrices for the NetNES, ELM and NES Finder programs derived from the data summarized in Table 1.

Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Figure 4. Amino acid sequence features of DUB cNESs. Multiple sequence alignment of DUB cNES motifs.
The name given to each motif (cNES ID) is indicated to the left of the sequence, and the hydrophobic residues that conform to the NES consensus are highlighted in grey. (**A**) DUB cNESs that tested positive in the nuclear export assay. Five different groups of active NESs could be established based on the pattern and spacing of hydrophobic residues. P-I: Φ^0-X_2-Φ^1-X_3-Φ^2-X_{2-3}-Φ^3-X-Φ^4; P-II: Φ^0-X_3-Φ^1-X_3-Φ^2-X_{2-3}-Φ^3-X-Φ^4; P-III: Φ^1-X_3-Φ^2-X_{2-3}-Φ^3-X-Φ^4; P-IV: Rev-type consensus; P-V: non-consensus. (**B**). DUB cNESs with a Φ^0-Φ^4 pattern that tested negative in the nuclear export assay. The presence of proline residues (highlighted in bold) between Φ^1 and Φ^3 may account for the lack of activity of several of these sequences (N-I group) by preventing them from adopting a proper CRM1-binding conformation. Other sequences (N-II group) do not have proline residues between Φ^1 and Φ^3, but are inactive in spite of fitting the Φ^0-Φ^4 consensus pattern.

Figure 5. Human deubiquitinases bearing active NESs.
Schematic representation of 22 human deubiquitinases showing the position of amino acid motifs (hatched boxes) that tested positive in the nuclear export assay. The position of the first and last residues of the NES motifs (as tested in the export assay) are indicated. The amino acid sequence and export activity score of each motif are detailed in Supplementary Table 2.

Figure 6. Effect of CRM1 inhibition on the localization of NES-bearing human DUBs.
(**A**). Representative examples of the nucleocytoplasmic localization of GFP- or YFP-tagged versions of the indicated deubiquitinases in HeLa cells untreated (left set of panels) or treated for 3 hours with 6 ng/ml of the CRM1 inhibitor LMB (rigth set of panels). GFP-USP1, YFP-USP3 and YFP-USP7 were located in the nucleus of untreated and LMB-treated cells. The localization of YFP-USP21, GFP-CYLD and GFP-OTUD7B was predominanlty cytoplasmic in untreated cells. After LMB treatment, GFP-CYLD remained in the cytoplasm, but YFP-USP21 and GFP-OTUD7B partially relocated to the nucleus. (**B**). Semiquantitative time-course analysis of the LMB-induced nuclear relocation of YFP-USP21 and GFP-OTUD7B. Graphs show the percentage of transfected cells showing predominantly nuclear (N), predominantly cytoplasmic (C) or equally nuclear and cytoplasmic (NC) distribution of these proteins. The average and standard deviation of three independent experiments is shown, and the total number of cells examined in each condition is indicated (n). YFP-USP21 showed a faster and more pronounced relocation to the nucleus than GFP-OTUD7B.

Figure 7. The N-terminal NES is necessary and sufficient for YFP-USP21 export.
(**A**). Schematic representation of USP21 protein structure showing the catalytic domain (orange) and the two active NES motifs, termed ELM20 and NET20 (hatched boxes), identified in this DUB by our global survey. The position and export activity score of each motif is indicated above. (**B**). The amino acid sequence of the wild type and mutant variants of ELM20 and NET20 are indicated, with the hydrophobic residues that conform to the consensus highlighted in blue. In the NES mutants, the Φ^3 and Φ^4 residues were replaced by alanines (red). Below, images of the nuclear export assay showing that the alanine replacements introduced render ELM20 and NET20 sequences inactive. (**C**). Schematic representation of YFP-USP21 wild type and mutant variants with NES-inactivating mutations. The sequence changes are indicated above. The L142/144A

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Accepted Manuscript

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

mutant bears the two ELM20-inactivating point mutations. The L313/315A mutant bears the two NET20-inactivating changes. Below, images show representative examples of HeLa cells expressing wild type and the NES-mutant versions of YFP-USP21. Wild type YFP-USP21 and the L313/315A mutant are located in the cytoplasm. In contrast, the L142/144A mutant shows a predominantly nuclear localization. Graphs indicate the percentage of transfected cells showing predominantly nuclear (N), predominantly cytoplasmic (C) or equally nuclear and cytoplasmic (NC) distribution of these proteins. (**D**). Full view of the structure of USP21 USP domain, encompassing residues 209-563, bound to ubiquitin (PDB code 3I3T). The NET20 motif (residues 301-319) is highlighted in magenta, the rest of USP21 sequence in green, and ubiquitin in blue. A ribbon representation (left panel) and a surface representation (right panel) are shown to illustrate the partial occlusion of NET20 by ubiquitin. (**E**). Close-up of the NET20 segment showing that the side chains of most hydrophobic residues in this motif (shown in stick representation) are oriented towards the inside of the protein.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

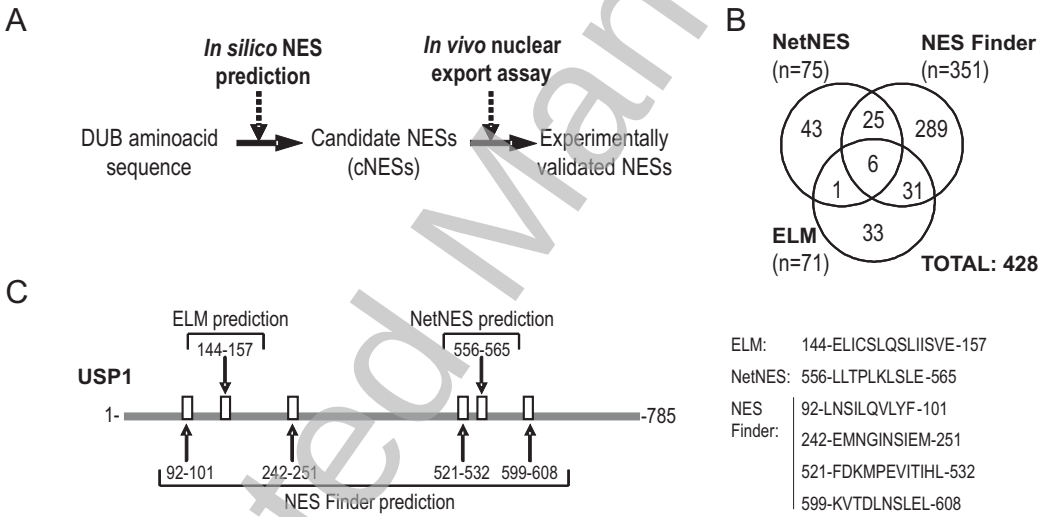
Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Figure 1



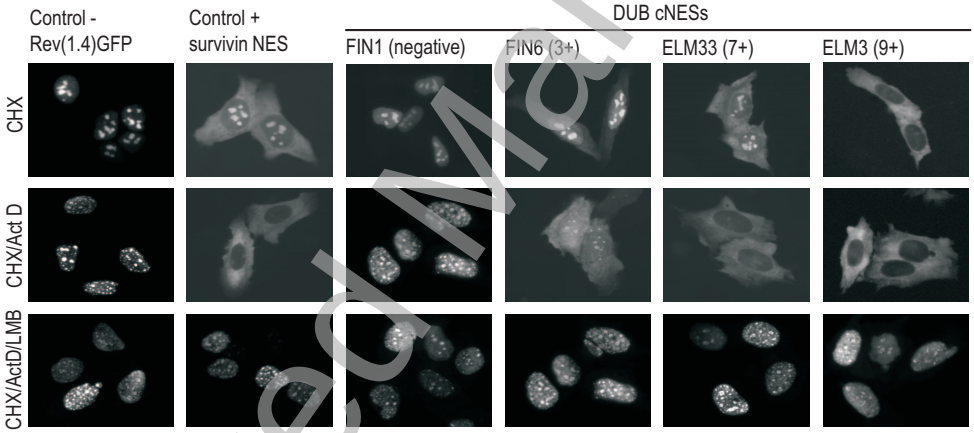
Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Figure 2



Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Figure 3

A

Prediction program		cNES predicted: YES	cNES predicted: NO
Export assay positive	YES	TP	FN
	NO	FP	TN

Sensitivity: 100xTP/(TP+FN)

Specificity: 100xTN/(FP+TN)

PPV: 100xTP/(TP+FP)

B

NetNES		cNES predicted: YES	cNES predicted: NO
E. a. positive	YES	20	36
	NO	55	317

ELM		cNES predicted: YES	cNES predicted: NO
E. a. positive	YES	27	29
	NO	44	328

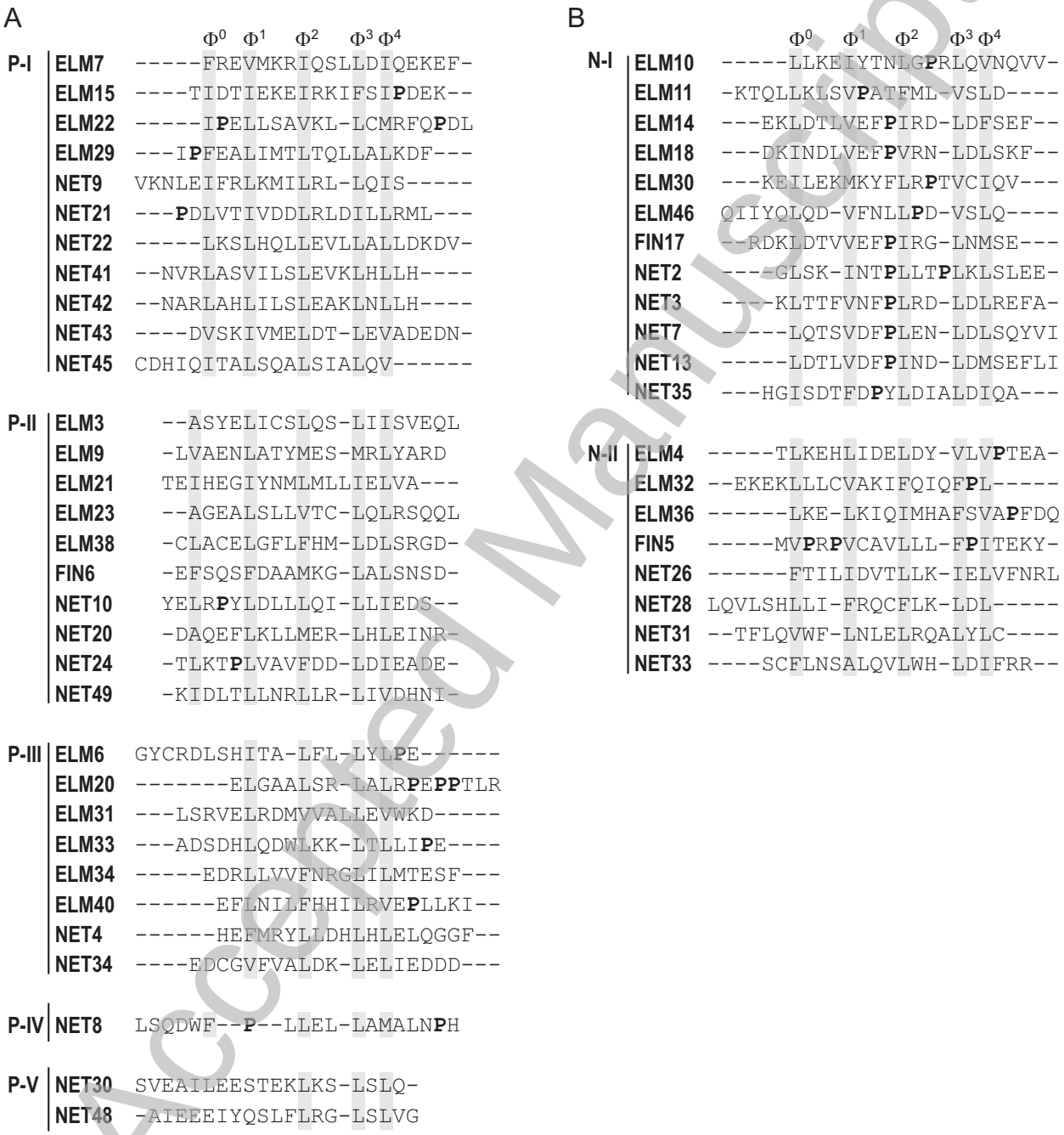
NES Finder		cNES predicted: YES	cNES predicted: NO
E. a. positive	YES	38	18
	NO	313	59

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Figure 4



Accepted Manuscript

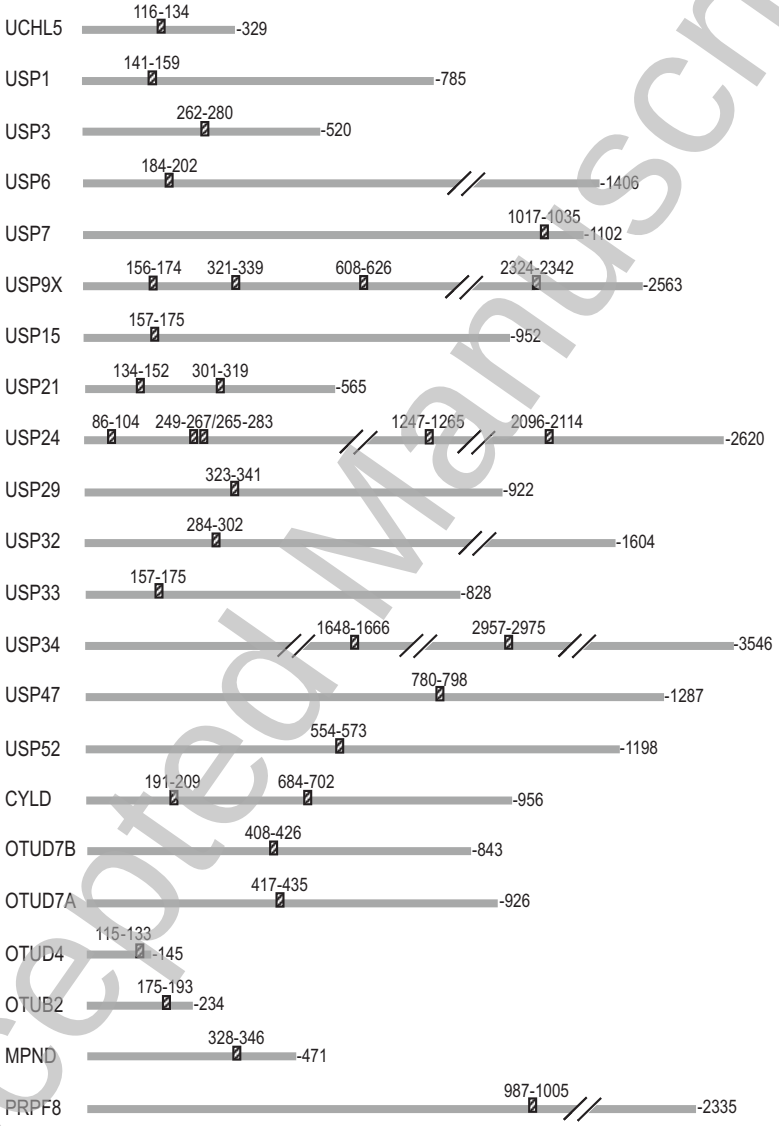
Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Figure 5

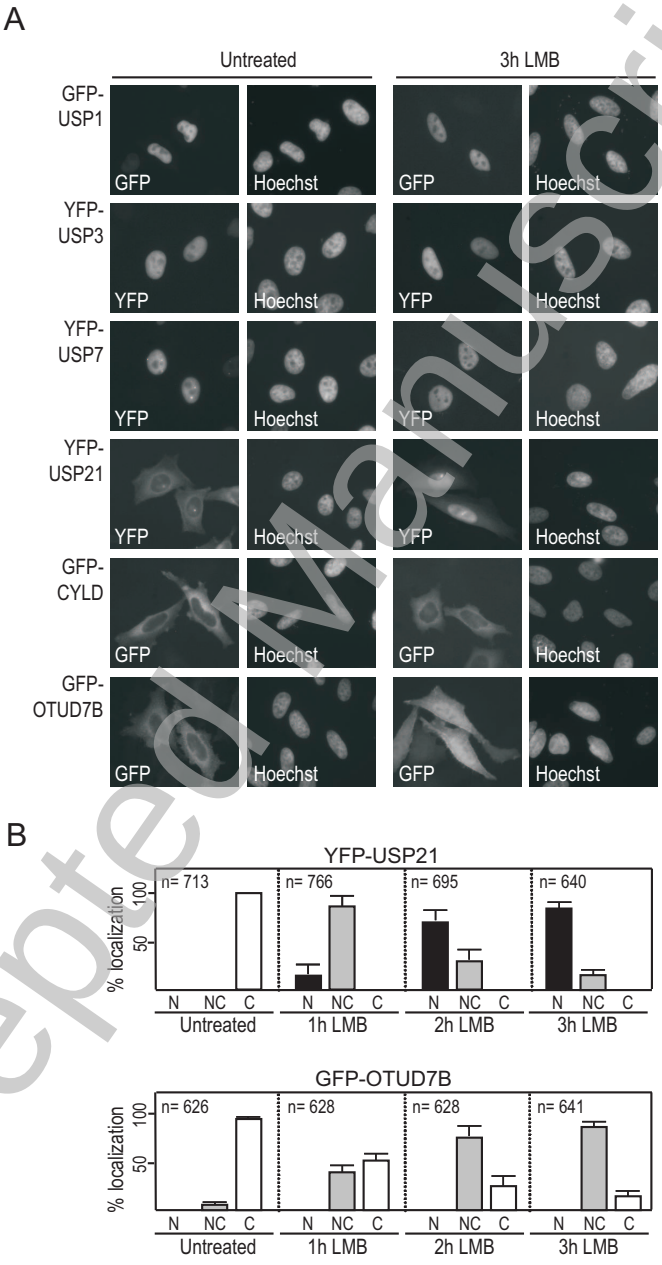


Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Accepted Manuscript

Figure 6

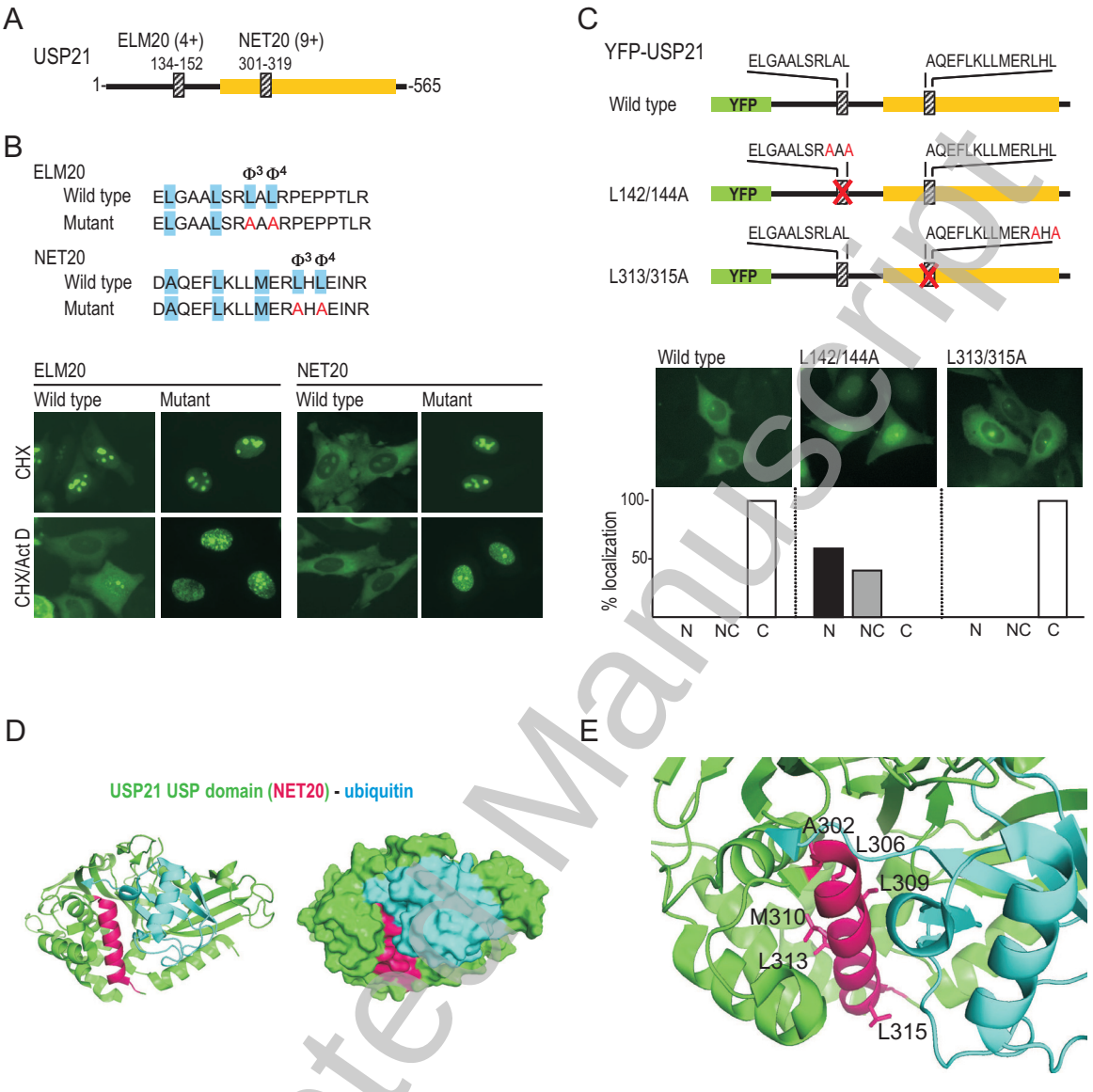


Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited

Biochemical Journal Immediate Publication. Published on 02 Sep 2011 as manuscript BJ20111300

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20111300

Figure 7



Accepted Manuscript

Licenced copy. Copying is not permitted, except with prior permission and as allowed by law.
© 2011 The Authors Journal compilation © 2011 Portland Press Limited