

Polyglutamylation is a posttranslational modification with a broad range of substrates.

Juliette van Dijk, Julie Miro, Jean-Marc Strub, Benjamin Lacroix, Alain van Dorsselaer, Bernard Edde, Carsten Janke

▶ **To cite this version:**

Juliette van Dijk, Julie Miro, Jean-Marc Strub, Benjamin Lacroix, Alain van Dorsselaer, et al.. Polyglutamylation is a posttranslational modification with a broad range of substrates.. Journal of Biological Chemistry, 2007, 283 (7), pp.3915-3922. 10.1074/jbc.M705813200 . hal-00217037

HAL Id: hal-00217037
https://hal.science/hal-00217037

Submitted on 25 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

POLYGLUTAMYLATION IS A POSTTRANSLATIONAL MODIFICATION WITH A BROAD RANGE OF SUBSTRATES*

Juliette van Dijk[1], Julie Miro[1], Jean-Marc Strub[2], Benjamin Lacroix[1], Alain van Dorsselaer[2], Bernard Edde[1,3] and Carsten Janke[1]

From [1]CRBM, Universités Montpellier 2 and 1, CNRS, 34293 Montpellier, France; [2]LSMBO, CNRS, 67087 Strasbourg, France and [3]Université Paris 6, 75252 Paris, France

Running head: New substrates of polyglutamylation

Address correspondence to: Carsten Janke, CRBM, CNRS, 1919 Route de Mende, 34293 Montpellier, France.

Tel: +33 4 67613335; Fax +33 4 67521559; E-mail: carsten.janke@crbm.cnrs.fr;

Polyglutamylation is a posttranslational modification that generates lateral acidic side chains on proteins by sequential addition of glutamate amino acids. This modification was first discovered on tubulins and it is important for several microtubule functions. Besides tubulins, only the nucleosome assembly proteins NAP1 and NAP2 have been shown to be polyglutamylated. Here, using a proteomic approach, we identify a large number of putative substrates for polyglutamylation in HeLa cells. By analyzing a selection of these putative substrates, we show that several of them can serve as in vitro substrates for two of the recently discovered polyglutamylases, TTLL4 and TTLL5. We further show that TTLL4 is the main polyglutamylase enzyme present in HeLa cells and that new substrates of polyglutamylation are indeed modified by TTLL4 in a cellular context. No clear consensus polyglutamylation site could be defined from the primary sequence of the here-identified new substrates of polyglutamylation. However, we demonstrate that glutamate-rich stretches are important for a protein to become polyglutamylated. Most of the newly identified substrates of polyglutamylation are nucleo-cytoplasmic shuttling proteins, including many chromatin binding proteins. Our work reveals that polyglutamylation is a much more widespread posttranslational modification than initially thought and thus that it might be a regulator of many cellular processes.

One fundamental aspect of proteomic complexity comes from the various processing events that many proteins undergo following their synthesis. Posttranslational modifications, such as phosphorylation, acetylation and methylation, are reversible monomodifications and are known to function as switches for the activity of many proteins. Polyglutamylation is a reversible polymodification generated by sequential covalent attachment of glutamic acids (up to 20 in some cases) to an internal glutamate residue of the target protein (1). The length of the resulting side chain is regulated by the balance between the enzymes that catalyze glutamylation, recently identified as members of the tubulin-tyrosine-ligase-like (TTLL) protein family (2-4), and yet unidentified deglutamylase enzymes (5). Thus, this modification does not only generate "on" and "off" states, but a range of signals that might allow for gradual regulation of protein functions.

The only known targets of polyglutamylation are α- and β-tubulins, the structural units of microtubules (MTs (1,6)), and the nucleosome assembly proteins, NAP1 and NAP2 (7). Tubulins are modified in their acidic, glutamate-rich C-terminus (1), which is the binding site for most MT-associated proteins (MAPs; reviewed in (8)). Tubulin polyglutamylation was therefore proposed to generate functionally divergent MTs by regulating the affinity between MAPs and MTs (4,9-11). It was also shown to be important for centriole stability (12), axonemal motility (2,13,14) and neurite outgrowth (15). The role of NAP polyglutamylation has not yet been addressed, but as for tubulins, it may regulate the affinity of NAPs for their binding partners.

Additional polyglutamylated proteins probably exist as the polyglutamylation-specific antibody GT335 (16) recognizes protein bands besides those of tubulins and NAPs on western blots of HeLa cell extracts (4,7). Here, using a proteomic approach, we identify new substrates for polyglutamylation and show that they are modified by two polyglutamylases from the TTLL protein

family. No clear "glutamylation motif" could be defined in their primary sequences, but all the here-identified new substrates contain glutamate-rich stretches that are most likely the acceptor sites for the modification. Our study opens the door to further investigations of the role of polyglutamylation as a general regulatory event of a broad range of cellular functions.

Experimental Procedures

Purification of GT335-reactive proteins from HeLa cells by immuno-affinity chromatography- 10 mg of GT335 antibody was coupled to an N-hydroxysuccinimide-activated Hi-Trap column (1 ml, GE Healthcare) according to manufacturer's instructions. HeLa cells were extracted in PBS containing 0.1% Triton X-100 and protease inhibitors (aprotinin, leupeptin, and 4-(2-aminoethyl)-benzenesulfonyl fluoride, each at 10 µg/ml). The high speed supernatant (15 mg of protein) was loaded onto the column and after extensive washing with PBS, bound proteins were eluted with PBS containing 0.7 M NaCl and concentrated to 1 mg/ml on an Amicon ultrafiltration device (4 ml, cut-off 10 kDa, Millipore).

Western blotting- Proteins fused to EYFP were immunodetected with rabbit anti-GFP antibody (1:5,000, Torrey Pines Biolabs) and HRP-conjugated anti-rabbit IgG (GE Healthcare) followed by detection with enhanced chemo-luminescence (GE Healthcare). GT335 antibody was biotinylated with NHS-LC-Biotin (Pierce) according to manufacturer's instructions and incubated simultaneously with HRP-labeled streptavidin (GE Healthcare). For producing polyclonal anti-TTLL4 and anti-TTLL5 anti-bodies, bacterially expressed and purified full-length mouse TTLL4 or fragment 692-1152 of mouse TTLL5 were injected into rabbits. Anti-TTLL antibodies were then purified from antisera against their respective targets.

Mass spectrometric analysis- To identify the protein content of the GT335-affinity-purified fraction, the Coomassie stained protein bands were excised from the gel and submitted to in-gel trypsin digestion (17). Peptides were then concentrated on a C18 pre-column, separated on a reversed-phase capillary column (Pepmap C18, 75 µm inner diameter, 15 cm length, LC Packings) and analyzed by MS/MS on a Q-Tof II mass spectrometer (Micromass Ltd., Manchester, UK). Data analysis was performed with Mascot (Matrix Science Ltd., London, UK) against the NCBI (The National Center for Biotechnology Information) Database. Double and triple charged peptides were used for the research in database and the peptide tolerance was set to 0.2 Da for MS and to 0.3 Da for MS/MS. One missed cleavage by trypsin was accepted and carbamidomethylated cysteine and oxidized methionine were set as variable modifications. All entries were selected for the taxonomy. The Mascot score cut-off value for a positive protein hit was set to 50 and peptides with a score below 40 were manually interpreted to validate or discard the identification. Identified proteins were submitted to a second research including the polyglutamylation as variable modification. Additions of 1 to 8 glutamic acids were considered and a new research has been done each time.

In vitro polyglutamylation of the GT335-affinity-purified fraction- The GT335-affinity-purified fraction (final protein concentration of 0.3 mg/ml) was incubated for 2 hours at 30°C in the presence of 50 mM Tris-HCl, pH 9.0, 0.5 mM ATP, 2.4 mM $MgCl_2$, 0.5 mM dithiothreitol, and 60 µM L-[^{3}H]-glutamate (45–55 Ci/mmol, GE Healthcare). Following SDS-PAGE, gels were exposed to MP films (GE Healthcare) for 40 days at -80°C after enhancement with Amplify (GE Healthcare).

Immuno-precipitation experiments- For the depletion of TTLLs, the GT335-affinity-purified fraction (10 µg of total protein) was incubated two hours at 4°C on a rotary agitator with protein G magnetic beads (Dynabeads) that were coupled to anti-TTLL4 or anti-TTLL5 antibodies (2 µg). After washing with PBS, the TTLL-depleted fraction was polyglutamylated in vitro as described above. For immuno-precipitation of the EYFP fusions of NAPs and their chimeras, the high-speed supernatant of HeLa cell extracts was incubated under similar conditions with beads linked to anti-GFP antibody. Following extensive washing, the beads were directly resuspended in Laemmli buffer.

In vitro polyglutamylation assay with recombinant proteins- Putative substrates of polyglutamylation were cloned from mouse testis or brain cDNA

libraries, expressed as polyhistidine tag-fusion proteins in bacteria and purified by nickel-affinity chromatography (except for EB1 which was a generous gift of S. Honnappa and M.O. Steinmetz). TTLL proteins were expressed as GST-fusions in bacteria and purified on glutathione sepharose as previously described (4). Since only TTLL5 (CAM84325) was soluble and active when produced in bacteria as its full-length version, we produced truncated version of TTLL4 (TTLL4_C639, last 639 residues of CAM84324) and TTLL6 (TTLL6_N705, first 705 residues of CAM84326) and a shorter, naturally occurring version of TTLL7 (TTLL7S, CAM84327) (3, 4). These shorter enzymes have been previously demonstrated to possess similar activities as compared to the full-length TTLLs (4). The polyglutamylation assay was performed by incubating 1 µg of each purified TTLL enzyme with 4 µg of each putative substrate in 20 µl of 50 mM Tris-HCl, pH 9.0, 0.5 mM ATP, 2.4 mM $MgCl_2$, 0.5 mM dithiothreitol, and 7.4 µM L-[^{3}H]-glutamate (45-55 Ci/mmol, GE Healthcare) for 2 hours at 30°C. Alternatively, 4 µg of taxotere-stabilized MTs, prepared from adult mouse brains (18), were used as a substrate. Quantification of the radioactivity incorporated into the different putative substrate proteins was done by scintillation counting of the protein bands after SDS-PAGE and electrotransfer onto nitrocellulose. Each polyglutamylation test was performed at least three times in independent experiments. The overall count (dpm) varied between each set of experiments; however, the relative values of glutamate incorporated into different substrates tested with the same enzyme dilution were similar in all experiments. Each figure is one set of a given experiment performed at the same time with all substrates presented. Background levels (no incorporation) are below 100 dpm.

In vivo polyglutamylation of the new substrates- Putative substrates of polyglutamylation were cloned as EYFP fusion proteins in mammalian expression vectors and transfected into HeLa cells together with an active or an enzymatically inactive (mutation E901G in the ATP-binding site) version of TTLL4_C639-EYFP. After 24 hours of transfection, cells were lysed in Laemmli buffer and subsequently analyzed by gel electrophoresis followed by western blot. In Fig. 5C, NAPs and their chimeras were expressed alone and immuno-precipitated with anti-GFP antibody prior to analysis.

Mutagenesis- In order to exchange the C-terminal tails between the different NAPs, we introduced the restriction site BstBI into a highly conserved region of the NAPs. This was done by mutagenesis of NAPs (NAP1: T972C, NAP2: T951C, NAP4: T1170C T1171G T1172A, introducing a L391E amino acid change) using the quick change PCR method. The resulting mutants were expressed and were shown to behave like the wild-type proteins in the polyglutamylase assay (data not shown). We then used BstBI in combination with other enzymes to exchange the NAP domains. The resulting chimeras were cloned into bacterial and mammalian expression vectors.

RESULTS

Purification and identification of putative substrates for polyglutamylation. Since the GT335 antibody detects protein bands besides those corresponding to the glutamylated forms of tubulins and NAPs on western blots of HeLa cell extracts, it is likely that this antibody specifically recognizes some other polyglutamylated proteins. We therefore performed a large-scale purification of GT335-reactive proteins from HeLa cell extracts by chromatography on a GT335-column. The affinity-purified fraction was submitted to SDS-PAGE and stained with Coomassie Blue (Fig. 1, lane 1) or transferred onto nitrocellulose and probed with GT335 antibody (Fig. 1, lane 2). We visualized 40 protein bands on the gel after Coomassie staining, and about half of them were detected with GT335 antibody on western blot. The remaining proteins might have been non-specifically retained on the column, for example by interacting with GT335-reactive proteins. However, it is also possible that some proteins are recognized by GT335 antibody only in their native form. We therefore excised all 40 protein bands from the gel and determined their protein content by nano-LC-MS/MS after trypsin digestion. This analysis gave rise to the identification of about 170 proteins (Table S1).

In order to identify true substrates of polyglutamylation, the masses obtained by nano-LC-MS/MS were re-analyzed to search for masses

that could be attributed to polyglutamylated tryptic peptides. However, we found no masses corresponding to modified peptides (mass increments of n times 129 Da, the mass of a glutamyl unit, with n ranging from 1 to 8). It has been previously reported that the polyglutamylated tryptic peptides of tubulins and NAPs are undetectable with the classical mass spectrometry method used here (1,7). The same seems to be true for the other substrates of polyglutamylation. Actually, because of their high acidity and/or branched structure, the modified peptides are lost during the reverse phase chromatography used to fractionate the peptides (7). An alternative procedure would be to fractionate the tryptic peptides by ion-exchange chromatography followed by a desalting step. This method has been used in the case of NAPs (7), but in order to follow the modified peptides during the purification steps, radioactive glutamate had to be incorporated into the proteins. Another problem, due to the acidic character of the modified peptides, is that their ionization in electrospray or MALDI is very hard even in the negative mode. Several procedures have been proposed to solve this problem (19-21), but the method remains far from straightforward. Taken together, we judged a mass spectrometry approach as not appropriate to identify new substrates of polyglutamylation.

To pre-select proteins with a high probability of being polyglutamylated, we took advantage of the fact that a polyglutamylase activity co-purifies with the GT335-reactive proteins (7). Incubation of the GT335-affinity-purified fraction with [^{3}H]-glutamate and MgATP resulted in the radioactive labeling of 13 protein bands (Fig. 1, lane 3). Since the new substrates of polyglutamylation might, as tubulins and NAPs, be modified within a glutamate-rich sequence we searched for proteins containing such motifs among the proteins identified within the 13 [^{3}H]-labeled bands. We found at least one protein with a glutamate-rich stretch in each of these bands (Table 1) and thus expect that these proteins are very likely targets of polyglutamylation.

TTLL4 and TTLL5 modify new substrates in vitro. To confirm whether the very likely targets of polyglutamylation, pre-selected in Table 1, can indeed be polyglutamylated, a selection of them was produced in bacteria and purified for analysis in an in vitro polyglutamylation assay. Not all proteins were tested because some could not be produced in bacteria due to their large size (nucleoporin, PELP1, Myb-binding protein 1A) and others shared a high homology with some of the tested proteins (PHAPII is a shorter version of SET that differs only in few amino acids at the N-terminus and ANP32E shares 57% identity with ANP32A (22)). To test the ability of the selected proteins to be polyglutamylated in vitro, we need a source of polyglutamylase activity. Presently, the only known mammalian polyglutamylase enzymes are members of the TTLL protein family (4). Following overexpression of the different TTLLs in HeLa cells, only TTLL4 and TTLL5 increased the GT335-reactivity of non-tubulin substrates (4). Thus, these two TTLLs are strong candidates for being the enzymes that polyglutamylate non-tubulin substrates. We therefore used recombinant TTLL4_C639 (a truncated version of TTLL4, see Experimental Procedures) and TTLL5 for in vitro polyglutamylation assays. To investigate the specificity of different enzymes for the new substrates, we also tested two other polyglutamylases, TTLL6_N705 and TTLL7S. The tested proteins, NAP1, NAP2, NF45, SET, B23, ANP32A and ANP32B (Table 1, proteins in bold) are all strongly modified by TTLL4_C639 and to a lesser extent by TTLL5 (Fig. 2A). The two other enzymes, TTLL6_N705 and TTLL7S, which are both highly active on MTs, did not modify the new substrates (Fig. 2A). Thus, in vitro, non-tubulin substrates are specifically polyglutamylated by TTLL4 and TTLL5. Moreover, since all pre-selected proteins tested are modified, it is likely that the presence of glutamate-rich sequences is indeed, as proposed earlier, a signature of polyglutamylated proteins.

Among the protein bands that were not labeled with [^{3}H]-glutamate by the polyglutamylase activity co-purified on the GT335-column, we identified several other proteins containing glutamate-rich stretches (Table S1, proteins in bold). These proteins might be polyglutamylated either to a lower extent or by a different enzyme, which was not co-purified on the GT335-column. To investigate this possibility, we repeated the in vitro polyglutamylation assays with the same set of TTLL enzymes on 6 of these proteins (GRP78, NASP, RANGAP, RNP-K, NCT and EB1; Fig. 2B). As before, TTLL4_C639 was the most efficient enzyme. It incorporated [^{3}H]-glutamate

into RNP-K and EB1, to a lesser extent into RANGAP and NCT, but poorly into GRP78 and NASP. Note that TTLL6_N705 and TTLL7S, which were inactive on the previously tested non-tubulin substrates, slightly modified EB1. This might be due to the close structural homology between EB1 and tubulins (23). Overall, the proteins modified in this test incorporated much less glutamate than the previously identified substrates, which might explain why they were not seen in the autoradiography (Fig. 1, lane 3). However, they have to be considered as potential substrates of polyglutamylation and their modification might be important in vivo.

TTLL4 is the major polyglutamylase co-purified with the new substrates. Since the GT335-affinity purified fraction possesses a polyglutamylase activity able to modify the new substrates (Fig. 1, lane 3), we wanted to determine if it contains either TTLL4 or TTLL5, the two enzymes that modify the substrates in vitro. We therefore raised polyclonal antibodies against these TTLLs. The epitopes of both, anti-TTLL4 and anti-TTLL5 antibodies are located outside the domains that are conserved between the different TTLLs (Fig. S1). Accordingly, when we probed extracts of HeLa cells overexpressing the different full-length TTLL polyglutamylases, anti-TTLL4 and anti-TTLL5 antibodies specifically recognized their targets and exhibited no cross-reactivity with any other TTLL polyglutamylase (Fig. 3A). We then used these antibodies to probe the GT335-affinity purified fraction, but none of the antibodies detected a protein of the expected size on western blot (not shown), most probably because of the absence or low concentration of the enzymes. Since both antibodies have been demonstrated to precipitate their target TTLL proteins (not shown), we used them to deplete the GT335-affinity purified fraction. Following incubation of the depleted fractions with [^{3}H]-glutamate and MgATP, we observed a strong decrease of glutamate incorporation when TTLL4, but not when TTLL5 was depleted (Fig. 3B). Thus TTLL4 is the major enzyme involved in the polyglutamylation of the new substrates identified in HeLa cells.

TTLL4 polyglutamylates the new substrates in HeLa cells. To show that the proteins that are polyglutamylated in vitro can also be modified by TTLL4 under in vivo conditions, we expressed them as EYFP-tagged versions in HeLa cells. In untransfected cells (not shown) or cells overexpressing an inactive (ATPase-dead) TTLL4_C639 (Fig. 4, first lane), only proteins in the 50 kDa region (tubulins, NAPs) and a 180 kDa protein are detected with GT335 antibody. This indicates that the endogenous polyglutamylase activity of HeLa cells might be too low to modify most substrates above detection levels of GT335 in western blot. Accordingly, in the absence of exogenous polyglutamylase activity, only substrates that were strongly modified in vitro (NAP1, NAP2, SET and the ANP32 proteins) could be detected by GT335 antibody following their overexpression (Fig. 4, lower panel, "D"). We then expressed the selected proteins together with an active version of TTLL4_C639 (Fig. 4; "A"). A significant increase in GT335-reactivity was observed on a range of endogenous proteins (Fig. 4, compare first and second lane), demonstrating that GT335 specifically detects polyglutamylation on a large range of proteins. Moreover, the GT335-reactivity of NAP1, NAP2, SET and the ANP32 proteins increased further, and other putative substrates (RANGAP and nucleolin) also became labeled with GT335 antibody (Fig. 4, lower panel, "A"). The expression levels of the substrates, as visualized with anti-GFP antibody, are similar in cells co-transfected with ATPase-dead and active TTLL4_C639 (Fig. 4, upper panel, "D", "A"). Thus, increase of GT335 detection levels clearly correlates with increased polyglutamylation of the respective proteins by the active TTLL4_C639. Moreover, a shift in migration was observed for most of the modified proteins, suggesting that the charge and/or size of these proteins have been changed (Fig. 4, upper panel, compare "A" with "D" for the modified proteins).

Requirements for polyglutamylation. We identified a number of proteins that are subjected to polyglutamylation. All these proteins possess glutamate-rich stretches within their primary sequence, but no obvious polyglutamylation consensus site could be defined. To better understand the determinants of polyglutamylation, we analyzed the primary sequences of several members of the NAP family (Fig. 5A and S2A) and their ability to be modified in vitro by TTLL4_C639 (Fig. 5B) or in HeLa cells by endogenous polyglutamylases (Fig. 5C). NAP1

and NAP2 are, in contrast to NAP3 and NAP4, strongly polyglutamylated in vitro and in vivo (Fig. 5). NAP1 was previously shown to possess a major polyglutamylation site in its C-terminus and a minor site in its N-terminus (7). Both sites consist of glutamate-rich stretches that are not well conserved among the members of the NAP family (Fig. S2A and Fig. 5A). NAP2 diverges from NAP1 in the N-terminal site. Since this is a minor polyglutamylation site, NAP2 is still strongly modified. On the contrary, the sequence of NAP4 diverges in the major C-terminal site and thus NAP4 is only lowly modified. NAP3 diverges in both sites and accordingly, it is not modified. Thus, we have shown a correlation between the presence of the two previously identified polyglutamylation sites and the extent of modification. In addition to these two sites, all NAPs possess a central domain which is rich in glutamates (Fig. S2A). This domain does not seem to be targeted by the polyglutamylation, as NAP3 is not modified although it has the most extended glutamate-rich central domain (Fig. 5A). The fact that this domain is not modified, although it is located at the outer surface of the protein according to the structure of NAP1 (24), suggests that the presence of a glutamate-rich sequence is not the only prerequisite for its polyglutamylation.

We then produced six chimeras of the NAP1, 2 and 4 proteins by swapping their C-terminal tails (Fig. 5A) and tested their ability to be modified by TTLL4_C639 in vitro (Fig. 5B) or by endogenous polyglutamylase activity in HeLa cells (Fig. 5C). Chimeras that possess the non-acidic C-terminal tail of NAP4 (NAP1_4 and NAP2_4), were poorly polyglutamylated both in vitro and in vivo, in good agreement with the lack of the major polyglutamylation site. On the other hand, chimeras that bear the highly acidic tail of NAP2 (NAP1_2 and NAP4_2) are, as NAP2, strongly modified. Concerning the chimeras that have the highly acidic tail of NAP1 (NAP2_1 and NAP4_1), NAP2_1 is modified to the same extent as NAP1 but NAP4_1 is less modified than one would expect according to the presence of the tail of NAP1. This again suggests that other, yet unknown determinants of polyglutamylation contribute to the modification.

To further demonstrate that the presence of a glutamate-rich stretch is not the only prerequisite for a protein to become polyglutamylated, we analyzed different members of the nucleophosmin / nucleoplasmin family. Nucleophosmin B23 has two putative acidic polyglutamylation sites (Fig. S2B). NPN2, an oocyte-specific homologue of B23, is more acidic in the first site but less in the second (Fig. S2B). Although, taken together, the total number of glutamate residues in two potential polyglutamylation sites is similar, NPN2 is modified to a lower extent than nucleophosmin B23 by TTLL4_C639 (Fig. 6). This suggests that the second site contributes more to the degree of polyglutamylation than the first one. Consistently, NPN3, which completely lacks the second site, is poorly modified by TTLL4_C639 (Fig. 6). Again, these results reveal that not all glutamate-rich stretches present in proteins are equally subjected to polyglutamylation.

We further analyzed the highly homologous proteins EB1 and EB3. Although both proteins carry a glutamate-rich C-terminus (Fig. S2C), only EB1 was modified by TTLL4_C639 in vitro (Fig. 6). Other TTLL enzymes that are active on EB1 (Fig. 2B) were inactive on EB3 as well (not shown).

Finally, we tested the protein HMG-B1, which has not been identified as a potential substrate, but has an extended glutamate-rich C-terminus (Fig. S2D). HMG-B1 was almost not modified in the in vitro assay with TTLL4_C639 (Fig. 6).

Taken together, these observations show that the presence of glutamate-rich sequences can be a prerequisite, but is not sufficient by itself for a protein to become a target of polyglutamylation, suggesting that other features are required.

DISCUSSION

Polyglutamylation was initially discovered on brain α-tubulin and subsequently also observed on β-tubulin (1,5,6). It was considered as a tubulin-specific modification until it was demonstrated to target also NAP1 and NAP2 (7). In the present study, we show that polyglutamylation is an even more general post translational modification.

New substrates for polyglutamylation. Using the affinity of polyglutamylated proteins for the GT335 antibody, we purified many potential polyglutamylated proteins (Table S1). We demonstrated that several of these proteins are indeed subjected to polyglutamylation as they are

modified by TTLL4 in vitro and/or in a cellular context. As we could not test all the putative substrates, a number of other targets of polyglutamylation might be present among the proteins listed in Table S1. Moreover, this list is by no means exhaustive as many of the substrates can be low-abundance proteins. It is also possible that not all polyglutamylated proteins interact with the GT335 antibody. This antibody has been raised against a synthetic peptide (EGEGEEEG) modified by the addition of two glutamyl units on the third glutamate (16) and it might, therefore, be biased towards proteins that are polyglutamylated within glutamate-rich sequences.

Specificity of the polyglutamylation reaction. All the here-identified substrates of polyglutamylation are modified within glutamate-rich stretches, suggesting that an acidic environment at the acceptor site is generally necessary for polyglutamylation to occur. Nevertheless, we demonstrate that one cannot simply conclude that a protein possessing glutamate-rich sequences is subjected to polyglutamylation. Substrate recognition by polyglutamylases is probably a complex mechanism. Solving the crystal structure of the polyglutamylase enzymes and thus knowing the properties of the surface residues of the enzyme that are in contact with the substrate, will certainly facilitate understanding of the prerequisites for the polyglutamylation reaction. Taken together, we suggest that proteins with glutamate-rich sequences can be generally considered as potential substrates for polyglutamylation, although this needs to be tested for each single case.

Possible new roles of polyglutamylation. Polyglutamylation generates acidic side chains of variable lengths on proteins. Such a modification might act as a fine-regulator of many cellular functions, for example by controlling the affinity of the here-identified target proteins for their binding partners, as earlier proposed for tubulins (reviewed in (25)). Many of the new substrates of polyglutamylation (ANP32 proteins, SET, nucleophosmin B23, nucleolin and RNP-K), as well as the NAP proteins, shuttle between the cytosol and the nucleus (22,26,27). Since nucleoporin and RANGAP, proteins involved in nucleo-cytoplasmic transport, are also modified, polyglutamylation could play an important regulatory role in nucleo-cytoplasmic shuttling. Moreover, most of these shuttling proteins are histone chaperones (NAP1, NAP2, B23, SET/PHAPII, nucleolin) and/or regulators of transcription (ANP32 proteins, SET/PHAPII, NF45, PELP1, RNP-K), suggesting that polyglutamylation might also regulate access to chromatin. Finally, EB1 and RANGAP are regulators of MT dynamics, suggesting that not only the polyglutamylation of tubulin, but also of its associated proteins could be important for the dynamics and functions of MTs. In cycling cells, the maximum yield of tubulin polyglutamylation is achieved during mitosis (18). It is thus tempting to speculate that, by targeting simultaneously MTs and chromatin binding proteins, polyglutamylation allows for a crosstalk between spindle dynamics and changes in chromatin structure accompanying mitosis.

REFERENCES

1. Edde, B., Rossier, J., Le Caer, J. P., Desbruyeres, E., Gros, F., and Denoulet, P. (1990) *Science* **247**(4938), 83-85
2. Janke, C., Rogowski, K., Wloga, D., Regnard, C., Kajava, A. V., Strub, J. M., Temurak, N., van Dijk, J., Boucher, D., van Dorsselaer, A., Suryavanshi, S., Gaertig, J., and Edde, B. (2005) *Science* **308**(5729), 1758-1762
3. Ikegami, K., Mukai, M., Tsuchida, J., Heier, R. L., Macgregor, G. R., and Setou, M. (2006) *J Biol Chem* **281**(41), 30707-30716
4. van Dijk, J., Rogowski, K., Miro, J., Lacroix, B., Edde, B., and Janke, C. (2007) *Mol Cell* **26**(3), 437-448
5. Audebert, S., Desbruyeres, E., Gruszczynski, C., Koulakoff, A., Gros, F., Denoulet, P., and Edde, B. (1993) *Mol Biol Cell* **4**(6), 615-626

6. Rudiger, M., Plessman, U., Kloppel, K. D., Wehland, J., and Weber, K. (1992) *FEBS Lett* **308**(1), 101-105
7. Regnard, C., Desbruyeres, E., Huet, J. C., Beauvallet, C., Pernollet, J. C., and Edde, B. (2000) *J Biol Chem* **275**(21), 15969-15976
8. Nogales, E. (2000) *Annu Rev Biochem* **69**, 277-302
9. Boucher, D., Larcher, J. C., Gros, F., and Denoulet, P. (1994) *Biochemistry* **33**(41), 12471-12477
10. Larcher, J. C., Boucher, D., Lazereg, S., Gros, F., and Denoulet, P. (1996) *J Biol Chem* **271**(36), 22117-22124
11. Bonnet, C., Boucher, D., Lazereg, S., Pedrotti, B., Islam, K., Denoulet, P., and Larcher, J. C. (2001) *J Biol Chem* **276**(16), 12839-12848
12. Bobinnec, Y., Khodjakov, A., Mir, L. M., Rieder, C. L., Edde, B., and Bornens, M. (1998) *J Cell Biol* **143**(6), 1575-1589
13. Gagnon, C., White, D., Cosson, J., Huitorel, P., Edde, B., Desbruyeres, E., Paturle-Lafanechere, L., Multigner, L., Job, D., and Cibert, C. (1996) *J Cell Sci* **109 (Pt 6)**, 1545-1553
14. Million, K., Larcher, J., Laoukili, J., Bourguignon, D., Marano, F., and Tournier, F. (1999) *J Cell Sci* **112 (Pt 23)**, 4357-4366
15. Ikegami, K., Heier, R. L., Taruishi, M., Takagi, H., Mukai, M., Shimma, S., Taira, S., Hatanaka, K., Morone, N., Yao, I., Campbell, P. K., Yuasa, S., Janke, C., Macgregor, G. R., and Setou, M. (2007) *Proc Natl Acad Sci U S A* **104**(9), 3213-3218
16. Wolff, A., de Nechaud, B., Chillet, D., Mazarguil, H., Desbruyeres, E., Audebert, S., Edde, B., Gros, F., and Denoulet, P. (1992) *Eur J Cell Biol* **59**(2), 425-432
17. Fraering, P. C., Ye, W., Strub, J. M., Dolios, G., LaVoie, M. J., Ostaszewski, B. L., van Dorsselaer, A., Wang, R., Selkoe, D. J., and Wolfe, M. S. (2004) *Biochemistry* **43**(30), 9774-9789
18. Regnard, C., Desbruyeres, E., Denoulet, P., and Edde, B. (1999) *J Cell Sci* **112 (Pt 23)**, 4281-4289
19. Juhasz, P., and Biemann, K. (1994) *Proc Natl Acad Sci U S A* **91**(10), 4333-4337
20. Kjellstrom, S., and Jensen, O. N. (2004) *Anal Chem* **76**(17), 5109-5117
21. Wilson, C. L., Monteith, W. B., Danell, A. S., and Burns, C. S. (2006) *J Pept Sci* **12**(11), 721-725
22. Santa-Coloma, T. A. (2003) *Cerebellum* **2**(4), 310-320
23. Honnappa, S., John, C. M., Kostrewa, D., Winkler, F. K., and Steinmetz, M. O. (2005) *Embo J* **24**(2), 261-269
24. Park, Y. J., and Luger, K. (2006) *Proc Natl Acad Sci U S A* **103**(5), 1248-1253
25. Verhey, K. J., and Gaertig, J. (2007) *Cell Cycle* **6**(17), 2152-2160
26. Borer, R. A., Lehner, C. F., Eppenberger, H. M., and Nigg, E. A. (1989) *Cell* **56**(3), 379-390
27. Michael, W. M., Eder, P. S., and Dreyfuss, G. (1997) *Embo J* **16**(12), 3587-3598

FOOTNOTES

*The authors are grateful to S. Honnappa and M.O. Steinmetz (Paul Scherrer Institut, Villigen, Switzerland) for providing essential reagents, to A. Bouaziz, J-M Donnay, J.C. Mazur, C. Rouveyrol and T. Zobel (CRBM, Montpellier, France) for technical assistance and to K. Rogowski, C. Bonne-Andrea, A. Burgess (CRBM, Montpellier, France) and M. Magiera (IGMM, Montpellier, France) for critical reading of the manuscript. This work was supported by the CNRS, the Universities Montpellier 1 and 2, the Association de la Recherche contre le Cancer awards CR504/7817 and 3140 (to C.J.), the French National Research Agency award JC05_42022 to C.J. and by financial support of La Ligue contre le Cancer to C.J. and B.L.

The abbreviations used are: MAP, Microtubule-associated protein; MT, microtubule; NAP, nucleosome assembly protein; TTLL, tubulin-tyrosine-ligase-like.

FIGURE LEGENDS

Fig. 1. Purification of GT335-reactive proteins from HeLa cell extracts. The high speed supernatant of HeLa cell extracts was submitted to GT335-affinity chromatography. The bound fraction was run on a 10% SDS polyacrylamide gel and stained with Coomassie blue (lane 1), or transferred onto nitrocellulose for detection with GT335 antibody (WB, lane 2). The incorporation of [^{3}H]-glutamate by the co-purified polyglutamylase activity was visualized by autoradiography (lane 3). The protein bands (whose content is listed in Table S1) are numbered and their number is in a box for the ones that appear radioactively labeled following in vitro polyglutamylation. Molecular mass markers are indicated.

Fig. 2. In vitro polyglutamylation of selected putative substrates with different TTLL enzymes. Proteins selected among the bands labeled by the co-purified polyglutamylase activity (A; Table 1) and proteins selected among the other bands according to the presence of glutamate-rich sequences (B; proteins in bold in Table S1) were expressed and purified from bacteria. They were then subjected to an in vitro polyglutamylation assay with recombinant TTLL4_C639, TTLL5, TTLL6_N705 and TTLL7S. The activity of each TTLL was verified by using MTs as a substrate. The amount of [^{3}H]-glutamate incorporated in the different substrates, was determined by scintillation counting.

Fig. 3. Identification of the polyglutamylase activity co-purified with the GT335-reactive proteins. The specificity of polyclonal anti-TTLL4 and anti-TTLL5 was demonstrated on extracts from HeLa cells overexpressing the different TTLL polyglutamylases in fusion with EYFP (A). These antibodies were then used to immuno-deplete the GT335-affinity-purified fraction (B). As a control, depletion was also performed with rabbit IgG. The depleted fractions were incubated with [^{3}H]-glutamate for 2 hours, run on an SDS-PAGE and stained with Coomassie Blue. [^{3}H]-glutamate incorporation was visualized by autoradiography. Only the depletion with anti-TTLL4 antibody reduced the polyglutamylase activity in the GT335-affinity-purified fraction. Molecular mass markers are indicated.

Fig. 4. In vivo polyglutamylation of selected putative substrates. The previously tested substrates of polyglutamylation (Fig. 2), as well as nucleolin which was not tested in vitro because it was hard to produce in bacteria, were co-transfected into HeLa cells with an active or inactive (ATPase-dead) version of TTLL4_C639 in fusion with EYFP. After 24 hours of transfection, the cells were lysed in Laemmli buffer, run on SDS-PAGE, transferred onto nitrocellulose and probed with anti-GFP antibody to check the expression level of the overexpressed proteins (upper panel) and with GT335 antibody to visualize the extent of polyglutamylation of the substrates (lower panel). Orange arrows point to the different overexpressed substrates visualized with anti-GFP antibody. In the lower panel, green arrows indicate the substrates that have been polyglutamylated in vivo, while red arrows point to proteins that were not modified. In general, we considered a substrate as being modified, when: (i) a band of the size of the protein (as compared to the anti-GFP signal) is recognized by GT335, and (ii) its GT335-reactivity increases upon co-expression of active TTLL4_C639 and (iii) this band does not correspond to endogenous GT335-reactive proteins (compare with cells not expressing the given substrate). Note that cells co-overexpressing B23 and active TTLL4_C639 were dying, explaining the decreased overall GT335-reactivity in the corresponding lane. Molecular mass markers are indicated.

Fig. 5. In vitro and in vivo modification of NAPs and chimeras. The sequences of the NAP proteins (see Fig. S2A) and of different chimeras in which the C-terminal tails of NAP1, NAP2 and NAP4 were exchanged are schematized (A). Previously identified N-terminal and C-terminal polyglutamylation sites are shown in red according to their size (domain 1 and 3 in Fig. S2A). NAPs and their chimeras were expressed in bacteria, purified and subjected to in vitro polyglutamylation using recombinant TTLL4_C639 as a source of enzyme (B). The amount of [^{3}H]-glutamate incorporated in the different substrates was determined by scintillation counting. NAPs and chimeras were also cloned in fusion with

EYFP in mammalian expression vectors and transfected into HeLa cells. Following 24 hours of transfection, they were first immuno-precipitated and then probed with either anti-GFP antibody to see the expression level, or with GT335 antibody to detect their polyglutamylation by endogenous enzymatic activity (C).

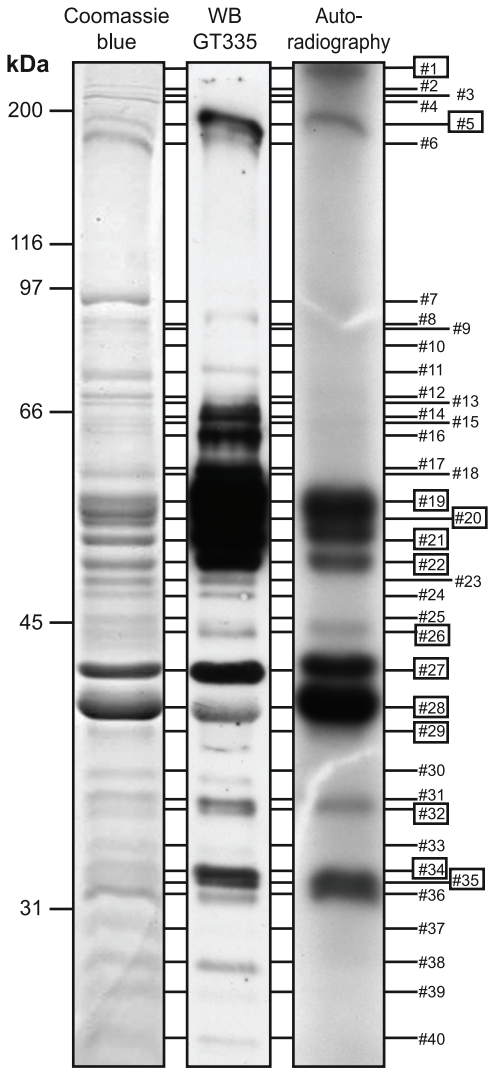
Fig. 6. In vitro polyglutamylation of proteins related to the new substrates. In vitro polyglutamylation assays were performed using recombinant TTLL4_C639 on proteins from the nucleophosmin / nucleoplasmin and EB families as well as on HMG-B1 (see sequences in Fig. S2A). The amount of [^{3}H]-glutamate incorporated in the different substrates was determined by scintillation counting.

TABLE

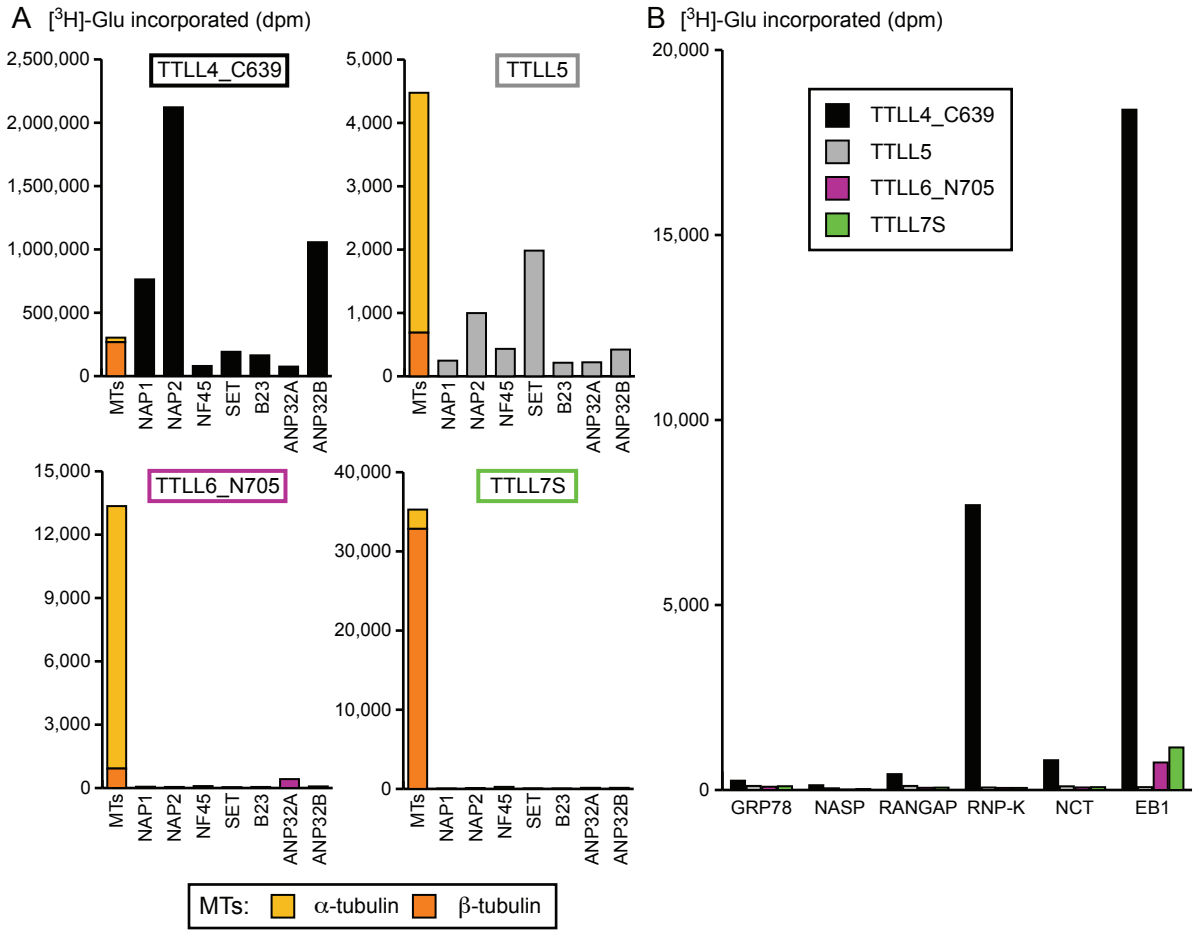
Table 1: Pre-selection of very likely targets of polyglutamylation.
Among the proteins identified within the 13 protein bands that were labeled with [^{3}H]-glutamate (Fig. 1 and Table S1), we selected those that possess long glutamate-rich stretches within their primary sequence and that thus are very likely targets of polyglutamylation. The band numbers correspond to those indicated in Fig. 1. In vitro polyglutamylation tests were performed on the proteins in bold and all of them did incorporate glutamate.

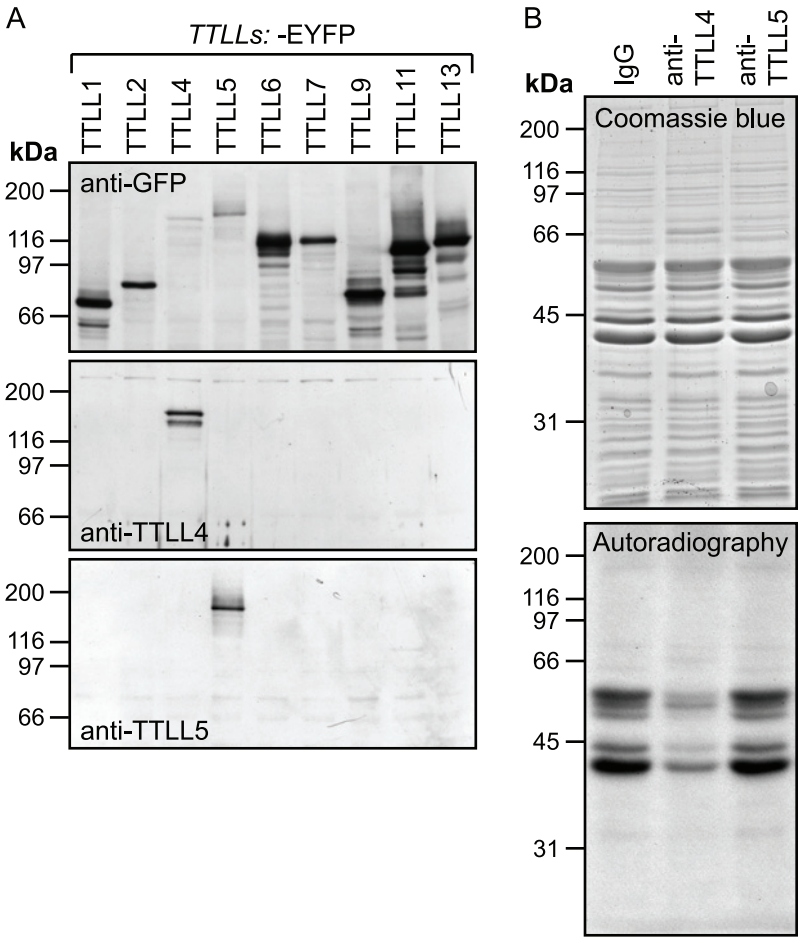
Band	Protein	Accession	Mass	pI
1	nucleoporin	gi|857368	357993	5.86
5	proline-, glutamic acid-, leucine-rich protein 1 (PELP1)	gi|24415383	119548	4.30
	Myb-binding protein 1A	gi|71153825	148762	9.34
19	alpha-tubulin	gi|32015	49761	4.95
20	**nucleosome assembly protein 1-like 4 (NAP2)**	gi|5174613	42797	4.60
21	beta-tubulin	gi|18088719	49640	4.75
22	**nucleosome assembly protein 1-like 1 (NAP1)**	gi|4758756	45346	4.36
26	**NF45 protein**	gi|532313	44669	8.26
27	**Protein SET**	gi|46397790	33469	4.23
28	PHAPII	gi|403009	32084	4.12
29	**B23 nucleophosmin**	gi|825671	30919	4.71
32	acidic nuclear phosphoprotein 32 family, member E (ANP32E)	gi|13569879	30674	3.77
34	**acidic nuclear phosphoprotein 32 family, member A (ANP32A)**	gi|5453880	28568	3.99
35	**acidic nuclear phosphoprotein 32 family, member B (ANP32B)**	gi|1498227	22263	4.19

van Dijk et al., Figure 1

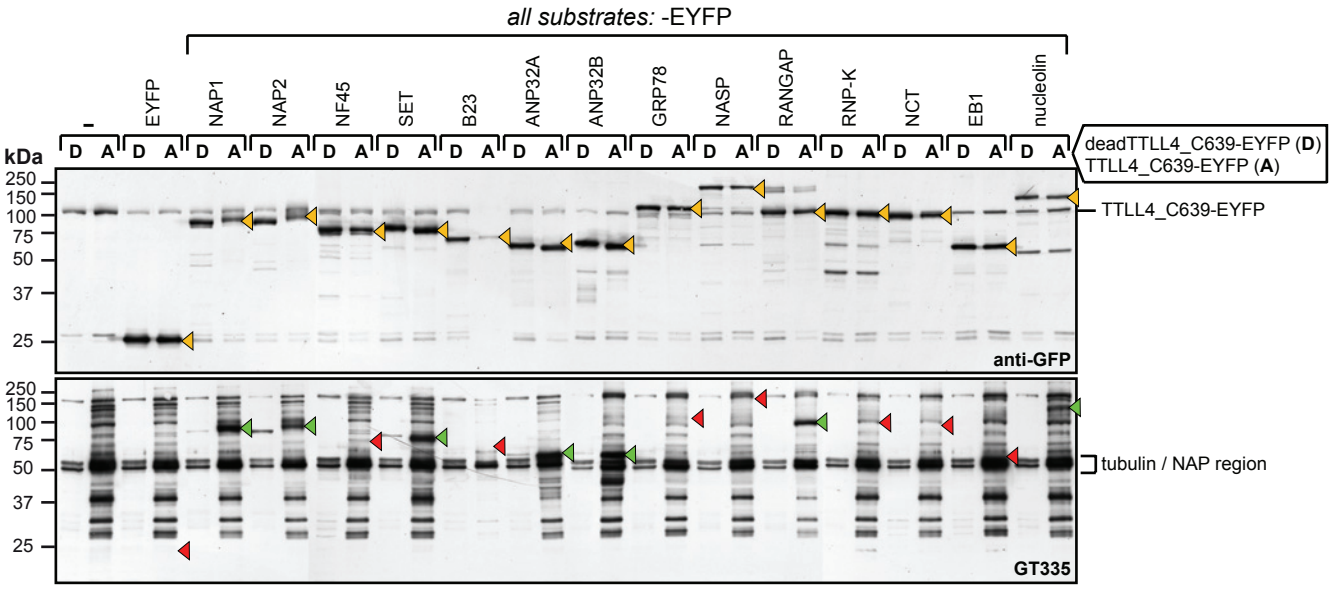


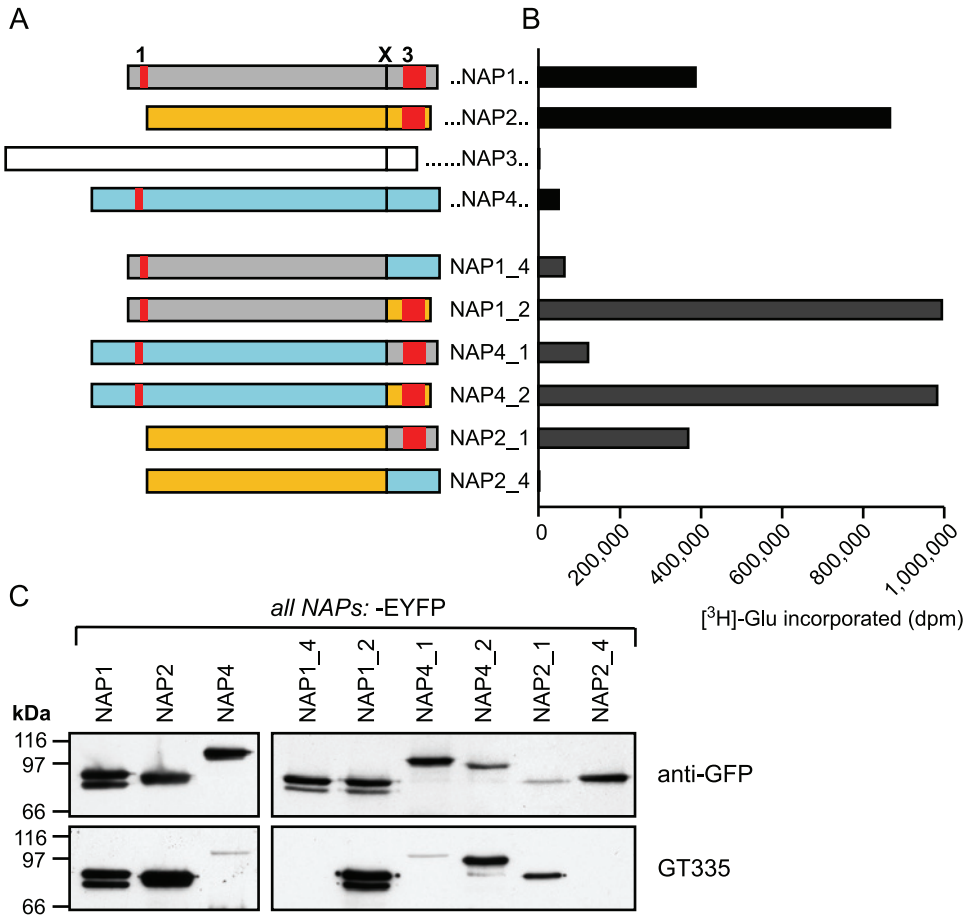
van Dijk et al., Figure 2



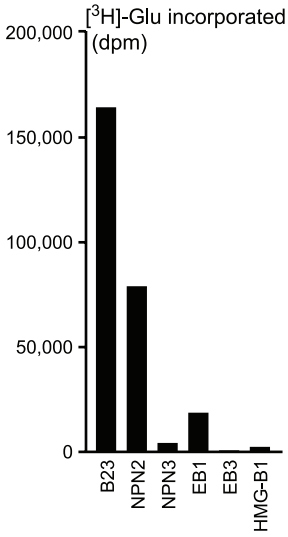

A
TTLLs: -EYFP
TTLL1
TTLL2
TTLL4
TTLL5
TTLL6
TTLL7
TTLL9
TTLL11
TTLL13
kDa
anti-GFP
200
116
97
66
anti-TTLL4
anti-TTLL5
B
IgG
anti-TTLL4
anti-TTLL5
Coomassie blue
45
31
Autoradiography

van Dijk et al., Figure 4




A
B
1
X
3
..NAP1..
...NAP2..
.....NAP3..
..NAP4..
NAP1_4
NAP1_2
NAP4_1
NAP4_2
NAP2_1
NAP2_4
0
200,000
400,000
600,000
800,000
1,000,000
[3H]-Glu incorporated (dpm)
C
all NAPs: -EYFP
NAP1
NAP2
NAP4
NAP1_4
NAP1_2
NAP4_1
NAP4_2
NAP2_1
NAP2_4
kDa
116
97
66
116
97
66
anti-GFP
GT335

van Dijk et al., Figure 6



SUPPLEMENTAL DATA

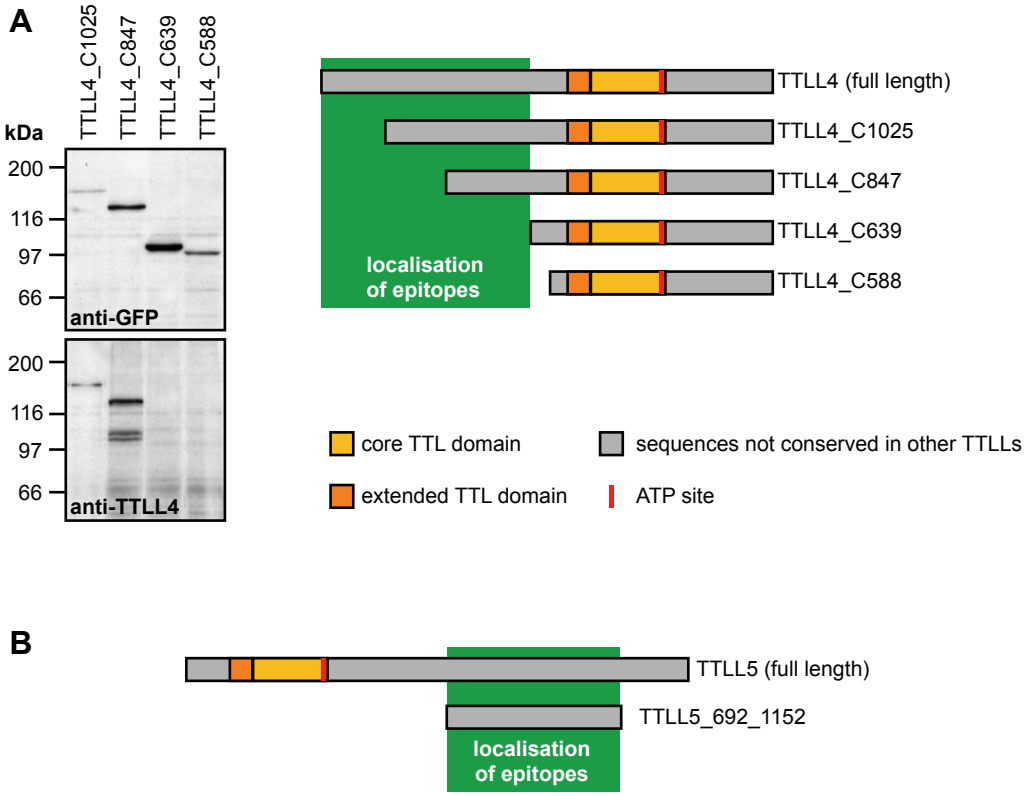


Figure S1: Epitopes of anti-TTLL4 and anti-TTLL5 polyclonal antibodies. (**A**) Anti-TTLL4 antibody was raised against the full-length protein. Out of four different truncated and EYFP-tagged versions of TTLL4 that were expressed in HeLa cells and analyzed on western blot, anti-TTLL4 recognized solely those containing N-terminal sequences located upstream of amino acid 555. Thus, all epitopes of this antibody are contained within the N-terminal domain of TTLL4, which is not conserved in any other TTLL. (**B**) Anti-TTLL5 antibody was raised against a fragment containing amino acids 692 to 1152 of TTLL5. This sequence is not conserved in any other TTLL protein. Schematic representation of TTLLs and their domains is according to van Dijk et al., 2007 (4).

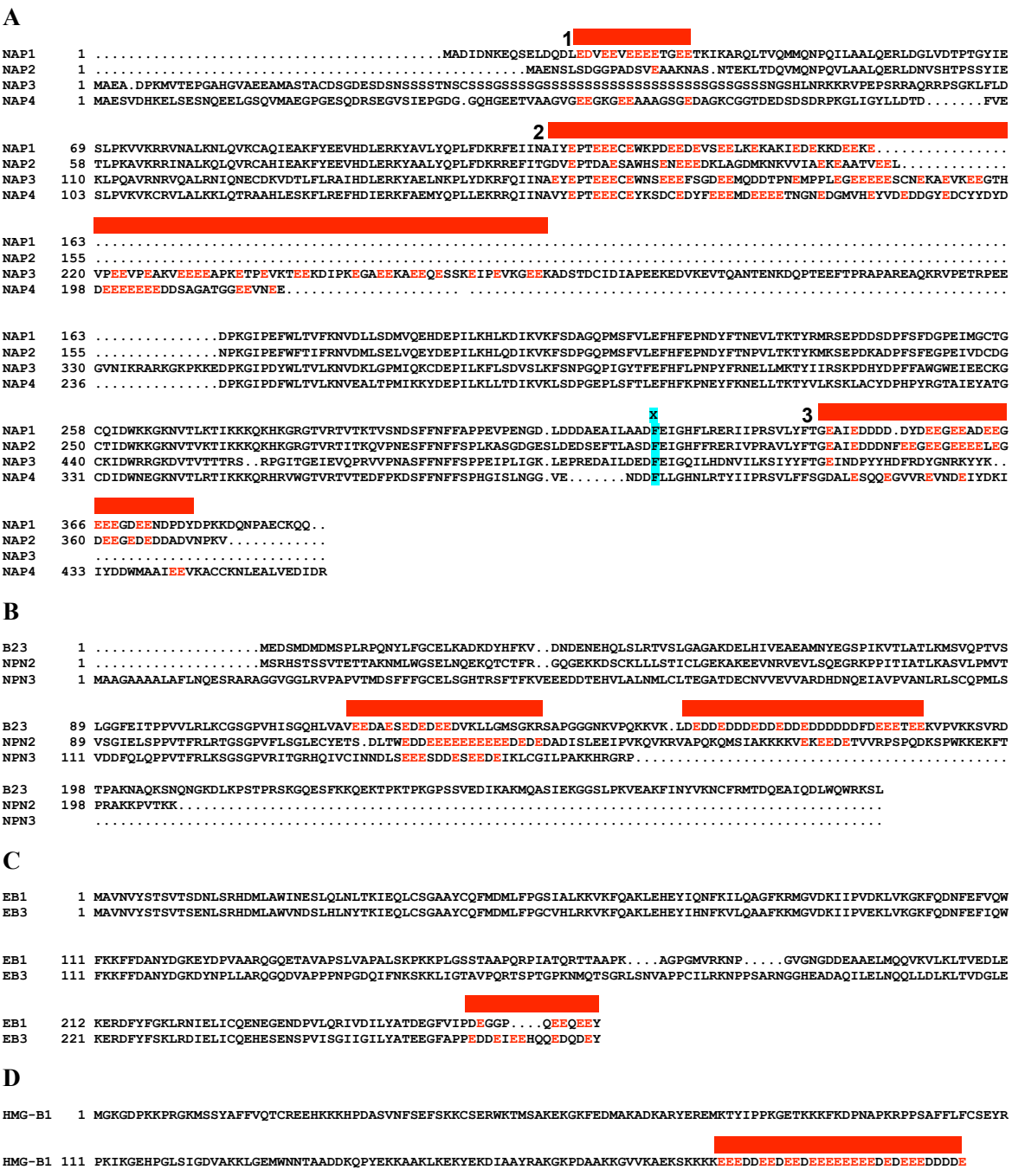


Figure S2: Sequence analysis of potential polyglutamylation sites. Murine primary sequences of members of the NAP (**A**) and nucleophosmin / nucleoplasmin (**B**) families as well as EB proteins (**C**) and HMG-B1 (**D**) are given. Sequence alignments were performed with MultiAlign (http://bioinfo.genopole-toulouse.prd.fr/multalin/) and corrected manually. Glutamate-rich domains have been highlighted and all glutamate residues in these regions are red. Note that not all regions are conserved between the different members of a family. For NAPs, the three glutamate-rich domains have been numbered and the site at which the C-terminal tails have been swapped in the chimeras have been underlayed with blue (X) (see Fig. 5A).

Supplementary Table 1: Identification of the GT335-affinity-purified fraction protein content. Forty protein bands were excised from the gel of Fig. 1, digested by trypsin and analyzed by nano-LC-MS/MS. The proteins identified in each band are listed. Proteins in bold contain at least one glutamate-rich sequence domain and are therefore considered as more likely substrates of polyglutamylation than others.

Band #	Protein	Accession #	Alternative names	Name used here	Mass	pI	Cove-rage	# pep-tides	RMS	Score
1	dynein, cytoplasmic 1, heavy chain 1	gi|119602166			525143	6.03	16%	53	17	1630
	plectin 1 isoform 3	gi|41322908			513393	5.61	7%	28	20	859
	nucleoporin	**gi|857368**			**357993**	**5.86**	**6%**	**16**	**24**	**524**
2	PRP8 pre-mRNA processing factor 8 homolog	gi|119610995			273382	8,94	24%	37	29	1266
3	**myosin, heavy polypeptide 9, non-muscle**	**gi|12667788**			**226392**	**5.50**	**40%**	**57**	**23**	**2749**
	activating signal cointegrator 1 complex subunit 3-like 1	gi|40217847			244353	5.73	11%	17	31	598
4	**translation initiation factor IF2**	**gi|4322304**	**TIF2**	**TIF2**	**138860**	**5.39**	**19%**	**15**	**14**	**385**
	IQ motif containing GTPase activating protein 1	gi|4506787			189134	6.08	7%	9	28	328
	eukaryotic translation initiation factor 3	**gi|4503509**			**166468**	**6.38**	**5%**	**4**	**24**	**164**
5	**proline-, glutamic acid-, leucine-rich protein 1**	**gi|24415383**	**PELP1, MNAR**	**PELP1**	**119548**	**4.30**	**27%**	**14**	**14**	**852**
	splicing factor 3b, subunit 1 isoform 1	gi|54112117			145738	6.65	17%	15	20	531
	Myb-binding protein 1A	**gi|71153825**			**148762**	**9.34**	**6%**	**4**	**13**	**85**
6	DEAH box polypeptide 9	gi|71153339			140869	6.41	34%	37	18	1573
	hnRNP U protein	**gi|32358**			**88890**	**5.60**	**27%**	**19**	**17**	**828**
	COPA protein	gi|23512328			139235	7.53	23%	23	24	744
	chromatin-specific transcription elongation factor	**gi|6005757**			**119838**	**5.50**	**21%**	**19**	**19**	**650**
	leucyl-tRNA synthetase	**gi|108773810**			**134379**	**6.95**	**14%**	**12**	**23**	**381**
	spliceosomal protein	gi|6006515			135507	5.15	13%	11	10	446
	Isoleucyl-tRNA synthetase, cytoplasmic	gi|730870			144866	5.86	12%	12	15	316
	Splicing factor 3B subunit 2	**gi|2498883**	**SAP145**		**97596**	**5.53**	**11%**	**8**	**12**	**183**
	DEAH box polypeptide 30 isoform 1	**gi|20336294**			**133854**	**8.99**	**9%**	**7**	**9**	**212**
	cellular myosin heavy chain	**gi|553596**			**154266**	**5.70**	**7%**	**8**	**29**	**182**
	DNA directed RNA polymerase II polypeptide B	gi|4505941			133811	6.44	7%	4	19	142
	Structural maintenance of chromosome 2-like 1	**gi|30173225**	**hCAP-E**		**135696**	**8.68**	**6%**	**4**	**15**	**142**
7	karyopherin beta 1	gi|19923142			97108	4.68	30%	16	8	853
	nucleolin	**gi|189306**			**76298**	**4.59**	**27%**	**14**	**14**	**1087**
	DEAD box polypeptide 21	**gi|50659095**			**87290**	**9.32**	**17%**	**10**	**9**	**270**
	testis expressed sequence 10	gi|8923269			105627	9.45	13%	8	7	431
8	**90kDa heat shock protein**	**gi|306891**	**HSP90**		**83242**	**4.97**	**33%**	**22**	**12**	**877**
	ILF3 protein	gi|40674401			76454	7.95	31%	18	24	724
	DEAH box polypeptide 15	**gi|68509926**			**90875**	**7,12**	**13%**	**11**	**23**	**342**
	DEAD box polypeptide 1	gi|4826686			82380	6.81	10%	3	29	131
	structure specific recognition protein 1	**gi|4507241**			**81024**	**6.45**	**8%**	**4**	**17**	**150**
	novel protein	**gi|5578958**			**81192**	**4.62**	**8%**	**3**	**27**	**172**
9	ATP-dependent DNA helicase II	gi|10863945			82652	5.55	31%	14	34	626
	Keratin, type II cytoskeletal 1 (Cytokeratin-1)	gi|1346343			65978	8.16	10%	6	31	377
	nuclear cap binding protein subunit 1, 80kDa	gi|4505343			91781	5,99	6%	4	10	145
10	**heterogeneous nuclear RNP R**	**gi|5031755**			**70899**	**8.23**	**28%**	**13**	**31**	**621**
	dead box, X isoform	gi|2580550			73226	6.73	17%	7	29	237
	solute carrier family 3, member 2 isoform	gi|61744475			71079	4.84	12%	5	29	242
	nucleolar protein 9	**gi|40217805**			**79226**	**9.33**	**6%**	**3**	**22**	**164**
	zinc finger protein 326 isoform 1	**gi|33946297**			**65613**	**5.08**	**6%**	**3**	**21**	**142**
11	**BiP protein**	**gi|6470150**	**GRP78**	**GRP78**	**70888**	**5.23**	**39%**	**19**	**30**	**986**
	polyadenylate binding protein II	gi|693937			58481	9.31	30%	12	23	543
	hepatocellular carcinoma associated protein	gi|4099969			64904	9.35	27%	13	25	483
	heterogeneous nuclear RNP M isoform a	gi|14141152			77464	8.84	14%	6	29	287
12	heat shock 70kDa protein 8 isoform 1	gi|5729877			70854	5.37	56%	27	22	1533
	heat shock-induced protein	gi|188488			70009	5.48	38%	12	10	694

13	**protein phosphatase 1G**	**gi|4505999**	**PP1G**		**59235**	**4.28**	**34%**	**14**	**26**	**709**
	growth regulated nuclear 68 protein	gi|226021			66881	8.95	22%	9	11	400
14	**ATP-dependent DNA helicase II, 70 kDa subunit**	**gi|4503841**			**69799**	**6.23**	**42%**	**20**	**8**	**857**
	histone-binding protein	**gi|184433**	**NASP, N1/N2**	**NASP**	**85139**	**4.27**	**15%**	**13**	**8**	**352**
	putative RNA binding protein KOC	gi|2105469			63681	8.99	10%	5	3	191
15	hnRNP Q1	gi|15809590			62561	7.18	40%	19	25	801
	transketolase	gi|37267			67751	7.90	40%	18	10	770
	Ran GTPase activating protein 1	**gi|4506411**	**RANGAP1**	**RANGAP**	**63502**	**4.63**	**26%**	**11**	**26**	**530**
	gamma subunit of CCT chaperonin	gi|671527			60292	6.23	22%	10	18	395
	RNA binding motif protein 39 isoform b	gi|4757926			58620	10.14	21%	7	18	266
	E2IG3	gi|6457340			63528	9.57	14%	4	21	272
	Similar to ribophorin I	gi|14124942			64542	6.10	5%	3	9	170
16	programmed cell death 8 isoform 1	gi|4757732			66859	9.04	28%	8	17	335
	tripartite motif-containing 26	**gi|4508005**			**62127**	**4.98**	**14%**	**3**	**16**	**105**
	Heterogeneous nuclear RNP L	gi|133274			60149	6.65	12%	4	15	218
	PBK1 protein	gi|3668141			58097	10.12	11%	4	15	131
	very-long-chain acyl-CoA dehydrogenase	gi|3273228			70391	8.88	6%	3	23	86
17	chaperonin containing TCP1, subunit 8 (theta)	gi|48762932			59582	5,54	33%	15	17	750
	chaperonin containing TCP1, subunit 6A isoform a	gi|4502643			57988	6.23	26%	7	18	405
	transformation upregulated nuclear protein	**gi|460789**	**hnRNP-K**	**RNP-K**	**51040**	**5.13**	**25%**	**8**	**10**	**405**
	importin-alpha homolog	gi|2654139			57737	4.80	24%	9	11	383
	chaperonin containing TCP1, subunit 5 (epsilon)	gi|24307939			59633	5.45	14%	5	23	303
18	T-complex protein 1 isoform a	gi|57863257			60306	5.80	32%	11	11	353
	5'-nucleotidase, cytosolic II	**gi|6912598**	**NCT**	**NCT**	**64928**	**5.75**	**27%**	**11**	**19**	**587**
	polypyrimidine tract-binding protein 1 isoform a	gi|4506243			59596	9,21	16%	5	140	139
19	vimentin	gi|37852			53653	5.06	60%	27	11	1255
	alpha-tubulin	**gi|32015**			**49761**	**4,95**	**42%**	**13**	**16**	**631**
20	**nucleosome assembly protein 1-like 4**	**gi|5174613**	**NAP2, NAP1_L4**	**NAP2**	**42797**	**4.60**	**49%**	**11**	**22**	**535**
21	**Tubulin, beta**	**gi|18088719**			**49640**	**4,75**	**47%**	**16**	**34**	**1066**
	RuvB-like 2	gi|5730023			51125	5,49	38%	15	39	786
	ATP synthase beta subunit	gi|179279			56896	5,38	16%	5	13	169
	RNP La	gi|337457			40569	8.47	12%	3	17	134
22	**nucleosome assembly protein 1-like 1**	**gi|4758756**	**NAP1, NAP1_L1**	**NAP1**	**45346**	**4.36**	**36%**	**9**	**42**	**315**
	ribosomal protein L4	gi|16579885			47667	11.07	23%	6	27	299
	eukaryotic translation initiation factor 2	gi|4503507			51077	8.66	21%	5	51	204
23	DNA-binding protein	gi|181914			35801	8.86	16%	3	31	151
	histone acetyltransferase 1 isoform a	gi|4504341			49481	5.52	14%	4	36	151
	translation initiation factor eIF-2 beta chain - human	gi|87941			38376	5.60	12%	3	47	152
	CGI-74 protein	**gi|4929617**			**47591**	**10.07**	**11%**	**3**	**29**	**106**
	heterogeneous nuclear RNP F	gi|4826760			45643	5.38	10%	3	41	123
24	DNA-binding protein B	gi|181486			39954	9,82	39%	8	54	554
	DnaJ protein homolog	gi|219588			44848	6.69	27%	5	50	227
	DnaJ subfamily A member 2	gi|5031741			45717	6.06	26%	6	33	273
	elongation factor Tu	gi|704416			49509	7.70	24%	5	27	402
	cytokeratin 18 (424 AA)	gi|30311			47305	5.27	20%	6	23	283
	keratin 17	gi|4557701			48076	4.97	17%	7	25	328
	translation initiation factor	gi|496902			46803	6.08	12%	3	38	114
	26S proteasome subunit p45	gi|976227			45624	8.23	11%	3	23	179
	farnesyl-protein transferase beta-subunit	gi|388758			43059	5.58	7%	3	48	90
25	ribosomal protein L3	gi|337580			45440	10.16	24%	7	46	261
	WD repeat protein 18	gi|21264122			47405	6.21	18%	3	32	174
	beta tropomyosin	**gi|6573280**			**29923**	**4,7**	**15%**	**4**	**30**	**193**
26	**NF45 protein**	**gi|532313**	**NF45, ILF2**	**NF45**	**44669**	**8.26**	**36%**	**10**	**35**	**553**
	alpha 1 actin precursor	gi|4501881			42024	5,23	32%	16	36	497
	proteasome subunit p42	gi|1526426			44133	7.10	11%	3	34	92
27	**Protein SET**	**gi|46397790**	**SET, TAF1, PHAPII**	**SET**	**33469**	**4.23**	**43%**	**13**	**44**	**666**
	PHAPII	**gi|403009**	**SET, TAF1, PHAPII**	**PHAPII**	**32084**	**4.12**	**53%**	**11**	**48**	**557**

28	**C protein**	**gi|306875**			**31947**	**5.10**	**33%**	**10**	**38**	**566**
	laminin-binding protein	gi|34234			31774	4.84	18%	3	43	175
	heterogeneous nuclear RNP A3	gi|34740329			39571	9,1	9%	3	26	100
29	**B23 nucleophosmin (280 AA)**	**gi|825671**	**B23, NPN1, NO38**	**B23**	**30919**	**4.71**	**32%**	**8**	**44**	**387**
	heterogeneous nuclear RNP A2/B1 isoform B1	gi|14043072			37407	8.97	27%	8	35	334
	RNA binding protein long isoform	gi|8051631			32444	9.20	24%	7	34	367
	heterogeneous nuclear RNP AB isoform a	gi|55956919			35945	6,49	13%	5	35	147
30	heterogeneous nuclear RNP A2/B1 isoform A2	gi|4504447			35984	8.67	50%	14	56	751
	ribosomal protein P0	**gi|4506667**			**34252**	**5.71**	**42%**	**9**	**58**	**455**
	annexin A2 isoform 2	gi|4757756			38580	7.57	34%	9	34	394
	heterogeneous nuclear RNP A0	gi|1911429			30882	9.34	23%	5	48	209
	glyceraldehyde-3-phosphate dehydrogenase	gi|31645			36031	8.26	17%	4	41	125
	heterogeneous nuclear RNP H3 isoform a	gi|14141157			36903	6,37	12%	3	48	112
31	DNA-binding protein TAXREB107	gi|9802306			32871	10.59	36%	10	43	459
	ribosomal protein L5	gi|550013			34426	9,76	9%	3	43	95
32	**acidic nuclear phosphoprotein 32 family, member E**	**gi|13569879**	**ANP32E, LANP-L**	**ANP32E**	**30674**	**3.77**	**44%**	**9**	**43**	**559**
	Heterogeneous nuclear RNP A1	gi|133254			38822	9.26	39%	14	52	699
	leucine rich repeat containing 59	gi|40254924			34909	9.61	36%	9	36	312
33	ribosomal protein S3a	gi|4506723			29926	9.75	45%	12	25	378
	THO complex 4	gi|55770864			26872	11.15	39%	5	36	192
	guanine nucleotide binding protein, beta polypeptide 2-like 1	gi|5174447			35055	7.60	25%	5	32	170
	tropomyosin 3 isoform 2	gi|24119203			29015	4,75	23%	7	35	303
	coiled-coil domain containing 124	gi|19923969			25820	9.54	21%	4	35	215
	microtubule-associated protein EB1	**gi|6912494**	**EB1**	**EB1**	**29980**	**5.02**	**20%**	**3**	**24**	**136**
	splicing factor, arginine/serine-rich 1 isoform 1	gi|5902076			27728	10.37	17%	3	20	107
	histone H1b	gi|356168			10359	11,03	15%	4	35	177
34	**acidic nuclear phosphoprotein 32 family, member A**	**gi|5453880**	**ANP32A, PHAPI, pp32**	**ANP32A**	**28568**	**3.99**	**38%**	**12**	**21**	**518**
	ribosomal protein S2	gi|15055539			31305	10,25	28%	7	24	269
	ribosomal protein L8	gi|4506663			28049	10.89	21%	6	21	187
	60S acidic ribosomal protein PO	**gi|5815233**			**27488**	**7.68**	**13%**	**3**	**9**	**116**
	cell division cycle 2 protein isoform 1	gi|4502709			34074	8.37	12%	3	11	114
35	S3 ribosomal protein	gi|7765076			26699	9.70	62%	12	24	567
	acidic nuclear phosphoprotein 32 family, member B	**gi|1498227**	**ANP32B, APRIL, PHAPI2A**	**ANP32B**	**22263**	**4.19**	**47%**	**13**	**26**	**617**
	ribosomal protein L7a	gi|4506661			29977	10.61	16%	3	20	162
36	ribosomal protein S4, X-linked X isoform	gi|4506725			29579	10,16	41%	13	21	490
	ribosomal protein L7	gi|35903			29164	10,6	41%	10	13	462
	small nuclear RNP polypeptide A'	**gi|50593002**			**28398**	**8,72**	**21%**	**4**	**12**	**284**
	ADP/ATP carrier protein (adenine nucleotide translocator 2)	gi|2772564			32831	9.71	18%	5	6	111
37	**ribosomal protein S8**	**gi|4506743**			**24190**	**10.32**	**54%**	**9**	**21**	**531**
	Csa-19	gi|531171			24844	9.94	31%	6	9	229
	ribosomal protein L13	gi|15431295			24247	11.65	29%	6	9	255
	heat shock protein 27	gi|662841			22313	7.83	24%	4	20	252
	homeobox prox 1	gi|7706322			28051	6.19	21%	4	6	184
	small nuclear RNP polypeptide B"	gi|4507123			25470	9.72	17%	3	15	105
38	Ribosomal protein L14	gi|12653649			23488	10.94	34%	6	14	307
	mitochondrial ribosomal protein S7	**gi|7705738**			**28144**	**10.00**	**28%**	**7**	**6**	**251**
39	60S ribosomal protein L15	gi|12006350			24145	11.62	32%	6	24	246
	ribosomal protein L13a	gi|6912634			23562	10.94	29%	9	28	348
	Rab5c-like protein, similar to Canis familiaris Rab5c protein	gi|508285			23553	8.87	27%	5	24	171
	von Hippel-Lindau binding protein 1	gi|4507873			22611	6.63	26%	4	26	242
	ribosomal protein L10	gi|5174431			24561	10.11	23%	4	17	232
	SEC22 vesicle trafficking protein homolog B	gi|94429050			24578	6,43	20%	3	24	136
	ribosomal protein L9	gi|15431303			21850	9,96	54%	8	30	313
	L21 ribosomal protein	gi|619788			17646	10.38	40%	6	32	193
	ribosomal protein L18	gi|4506607			21621	11.73	38%	6	32	324
	ribosomal protein	gi|337518			21835	10.09	36%	7	42	309
	hypothetical protein LOC84310	gi|14150149			22070	9.65	34%	4	31	197

40	ribosomal protein L18a	gi|11415026			20749	10.73	34%	4	30	149
	ribosomal protein S5	gi|550021			22763	9.59	28%	5	32	291
	Rab7	gi|4105819			23517	5,71	26%	4	30	173
	ribosomal protein S9	**gi|550023**			**22558**	**10.74**	**26%**	**8**	**28**	**323**
	Ras-related protein Rab-11A	gi|4758984			24378	6,12	25%	3	26	180
	ribosomal protein L17	gi|4506617			21383	10.18	23%	5	31	200