A role for Q/N-rich aggregation-prone regions in P-body localization

Citation for published version:

Digital Object Identifier (DOI):
10.1242/jcs.024976

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Journal of Cell Science

Publisher Rights Statement:
RoMEO green

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
A role for Q/N-rich aggregation-prone regions in P-body localization

Martin A. M. Reijns*, Ross D. Alexander, Michael P. Spiller‡ and Jean D. Beggs§

Wellcome Trust Centre for Cell Biology, University of Edinburgh, King’s Buildings, Mayfield Road, Edinburgh EH9 3JU, UK
*Present address: Medical Research Council, Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
‡Present address: Faculty of Life Sciences, Michael Smith Building, University of Manchester, Oxford Road, Manchester M13 9PT, UK
§Author for correspondence (email: jbeggs@ed.ac.uk)

Summary
P-bodies are cytoplasmic foci that are sites of mRNA degradation and translational repression. It is not known what causes the accumulation of RNA-degradation factors in P-bodies, although RNA is required. The yeast Lsm1-7p complex (comprising Lsm1p to Lsm7p) is recruited to P-bodies under certain stress conditions. It is required for efficient decapping and degradation of mRNAs, but not for the assembly of P-bodies. Here we show that the Lsm4p subunit and its asparagine-rich C-terminus are prone to aggregation, and that this tendency to aggregate promotes efficient accumulation of Lsm1-7p in P-bodies. The presence of glutamine- and/or asparagine-rich (Q/N-rich) regions in other P-body components suggests a more general role for aggregation-prone residues in P-body localization and assembly. This is supported by reduced P-body accumulation of Ccr4p, Pop2p and Dhh1p after deletion of these domains, and by the observed aggregation of the Q/N-rich region from Ccr4p.

Introduction
Cytoplasmic mRNA-processing bodies (P-bodies) contain a variety of protein factors, some of which are involved directly in decapping (Dcp1p, Dcp2p), some in translational repression and/or activation of decapping (Pat1p, Dhh1p, Edc1p, Edc2p, Edc3p, Scd6p) and others are involved in deadenylation (Ccr4p, Pop2p, Not1p, Not2p, Not3p, Not4p, Not5p, Pan2p, Pan3p) or 5’ to 3’ decay (Xrn1p). In addition, factors involved in nonsense-mediated decay (Pat1p, Dhh1p, Edc1p, Edc2p, Edc3p, Scd6p) and degradation of mRNAs, but not for the assembly of P-bodies. This seems to require RNA (Teixeira et al., 2005). More recently, various proteins in budding yeast have been implicated directly in P-body assembly, and the understanding of their physical and functional interactions is gathering pace. This includes Edc3p, Lsm4p (Decker et al., 2007), Pat1p (Pilkington and Parker, 2008), and Ded1p (Beckham et al., 2008).

Lsm4p of Saccharomyces cerevisiae encodes an essential protein of 187 amino acids (aa). It is one of the seven subunits of the Lsm1-7p and Lsm2-8p (comprising Lsm2p to Lsm8p) complexes, the latter of which is needed for efficient pre-mRNA splicing through its role in U6 small nuclear RNA (snRNA) stability (Achsel et al., 1999; Mayes et al., 1999; Pannone et al., 1998; Salgado-Garrido et al., 1999) and localization (Spiller et al., 2007a) as well as U4/U6 di-snRNA formation (Verdone et al., 2004). The N-terminal 92 aa of Lsm4p include the Sm domain, which is involved in protein-protein and protein-RNA interactions within the Lsm complexes (Cooper et al., 1995; Hermann et al., 1995; Séraphin, 1995). This region is highly conserved between budding yeast and humans (Fig. 1B). The C-terminal 95 aa are rich in asparagine (N; 36%) and serine (S; 17%), giving this region a highly hydrophilic character. It is less conserved than the N-terminus, however, homologues from various Saccharomyces species contain similar asparagine-rich stretches that vary in length and position. A notable exception is Lsm4p from S. kluveri that has a glutamine (Q)-rich region (Fig. 1A). This N and/or Q-rich character of the Lsm4p C-terminus is conserved throughout the budding yeasts (supplementary material Fig. S1). By contrast, most Lsm4p homologues from higher organisms have an abundance of arginine and glycine residues in their C-termini, often in the form
of RG repeats (Fig. 1B and supplementary material Fig. S2), that are important for interactions with the SMN complex. Symmetrical dimethylation of the arginine residues is thought to be important for regulation of snRNP assembly (Brahms et al., 2001; Paushkin et al., 2002). S. cerevisiae does not have a known SMN complex equivalent, providing a possible explanation for the absence of RG repeats in yeast Lsm4p. Experiments with Lsm4p of Kluyveromyces lactis suggest that the Lsm4p C-terminus is needed for efficient RNA degradation (Mazzoni et al., 2003a; Mazzoni et al., 2003b).

Here we describe the role of yeast Lsm4p and its C-terminus in Lsm protein aggregation. We show that the ability of Lsm4p to aggregate, although not essential, promotes efficient accumulation of Lsm1-7p in P-bodies. Many other P-body components contain Q/N-rich regions suggestive of a more general role for such aggregation-prone residues in efficient accumulation of these factors in P-bodies. In support of this hypothesis we show that the Q/N-rich region of Ccr4p is prone to aggregation under normal growth conditions and shows increased focal localization under stress conditions. Furthermore, we show that the Q/N-rich region of Ccr4p is essential for its accumulation in microscopically visible P-bodies, whereas those of Pop2p and Dhh1p, although not essential for P-body localization, promote their efficient accumulation in these cytoplasmic foci.

**Results and Discussion**

**Overproduced Lsm4p or its C-terminus accumulate in foci**

While investigating the localization of various GFP-Lsm4p fusions, we observed that GFP fused to full-length Lsm4p (GFP-Lsm4) or to its C-terminus (GFP-Lsm4C; aa 92-187) accumulates in cytoplasmic foci as well as in larger aggregates, in a variable percentage of cells (10-60%), whereas GFP fused to the N-terminal half of Lsm4p (GFP-Lsm4AC; aa 1-92), containing the Sm domain, does not (Fig. 2A). The number of GFP-Lsm4 foci increases after hypo-osmotic shock, indicating that aggregation can be triggered by stress, and that these newly formed foci are probably P-bodies. This is confirmed by colocalization of Dcp2-RFP with GFP-Lsm4p in foci formed after stress. By contrast, Dcp2-RFP is not particularly enriched in the larger Lsm4p aggregates during log-phase growth, suggesting that these are probably not P-bodies (Fig. 2B). In cells expressing GFP-Lsm4AC as the only copy of Lsm4p, Dcp2-RFP localizes to foci after osmotic shock, showing that P-bodies are formed (Fig. 2B). However, GFP-Lsm4AC localizes throughout the cell, indicating its failure to accumulate in P-bodies even under stress conditions. The virtual absence of GFP-Lsm4ΔC in microscopically visible P-bodies is not due to reduced levels of this truncated protein, as shown by western analysis (see below).

Other Lsm proteins aggregate upon Lsm4p overproduction

To investigate a potential link between these Lsm4p aggregates and P-bodies, the presence of other proteins was examined. We observed accumulation of all tested GFP-tagged Lsm proteins (Lsm1p, Lsm2p, Lsm6p, Lsm7p and Lsm8p) in similar cytoplasmic aggregates during log-phase growth of cells overproducing non-tagged Lsm4p ($\text{P}_{\text{GAL-LSM4}}$ strain grown on galactose; Fig. 2C and supplementary material Fig. S3). This was even the case with the normally nuclear Lsm8p, indicating that not only Lsm1-7p but also Lsm2-8p aggregates when Lsm4p is present at high levels. Aggregates were observed mostly in the cytoplasm and occasionally in the nucleolus, judging from colocalization with the nucleolar protein Nop1p (data not shown). By contrast, normal localization was observed for each of the Lsm proteins in the absence of excess Lsm4p (Fig. 2C), although each of these GFP-tagged proteins was moderately overexpressed from the $\text{MET25}$ promoter. Lsm protein aggregates are therefore likely to be not physiologically significant, but simply the result of aggregation of Lsm4p-containing complexes when Lsm4p is present at higher than normal levels. As aggregation was observed with direct interaction partners of Lsm4p in the ring-shaped Lsm complex (Lsm1p, Lsm2p and Lsm8p) as well as with physically more distant subunits (Lsm6p and Lsm7p) it seems possible that overexpressed Lsm4p drives the aggregation of entire Lsm1-7p and Lsm2-8p complexes, probably via its C-terminus. This aggregation is specific to the Lsm proteins, as GFP alone localized throughout the cells regardless of Lsm4p levels (Fig. 2C), and the...
Q/N-rich sequences and P-body localization

exclusively nuclear Lhp1p fused with GFP remained nuclear under these conditions and did not form aggregates, although many cells showed abnormal nuclear morphology, which is a phenotype associated with Lsm4p overproduction (supplementary material Fig. S3).

A role for Lsm4p in Lsm1-7p P-body localization

The accumulation of GFP-Lsm4 and GFP-Lsm4C in foci when overexpressed, and failure of GFP-Lsm4ΔC to aggregate even under stress conditions, suggests a role for the Lsm4p C-terminus in targeting Lsm1-7p to P-bodies. As a complete Lsm1-7p complex is apparently needed for localization to P-bodies (Ingelfinger et al., 2002; Tharun et al., 2005) the C-terminal deletion is likely to affect localization of the entire Lsm1-7p complex. The localization of GFP-Lsm1 to P-bodies was therefore examined in cells producing either Lsm4p or Lsm4ΔCp (both non-tagged) from the native LSM4 promoter (i.e. not overproduced). In comparison with the accumulation of GFP-Lsm1 in P-bodies (for colocalization of GFP-Lsm1 with Dcp2-RFP see supplementary material Fig. S4) following hypo-osmotic stress of LSM4 cells, there was a reduction in the intensity of GFP-Lsm1 foci that formed in lsm4ΔC cells (Fig. 3A), as well as an apparent delay in their formation. To quantify this delay, the number of cells that displayed visible P-bodies 5 minutes and 1 hour after hypo-osmotic shock was counted in the LSM4 and lsm4ΔC strains. Whereas 85% of LSM4 cells displayed foci 5 minutes and 1 hour after hypo-osmotic shock was counted in the LSM4 and lsm4ΔCp strains. Whereas 85% of LSM4 cells displayed foci 5 minutes after hypo-osmotic shock, with only a small increase (to 90%) after 1 hour, only 4% of lsm4ΔC cells displayed foci after 5 minutes, increasing to 73% after 1 hour (Fig. 3B), with the majority of these foci still weaker than those observed in LSM4 cells. The localization of GFP-Lsm2 and GFP-Lsm6 to P-bodies after hypo-osmotic stress was similarly reduced in the lsm4ΔC strain compared with the LSM4 strain (Fig. 3C,D). Taken together, these data strongly suggest that the C-terminal domain of Lsm4p although not actually essential, is nevertheless important for efficient accumulation of Lsm1-7p in P-bodies under stress conditions. As the Lsm4p C-terminal domain seems to be important for efficient recruitment of Lsm1-7p to P-bodies, the C-terminal deletion might also have an effect on the accumulation of other proteins in P-bodies. However, no significant effect was seen on the localization of either Dcp1p or Dcp2p to P-bodies (Fig. 3E, and data not shown).

Detrimental effects of GFP-tagging Lsm4ΔCp

The complete absence of GFP-Lsm4ΔCp from P-bodies after osmotic shock seems to contradict the mere reduction in P-body accumulation of GFP-Lsm1, Lsm2 and Lsm6 in the lsm4ΔC (non-tagged) strain. However, upon closer inspection, GFP-Lsm4ΔC was observed to localize weakly to P-bodies after hypo-osmotic shock in a small fraction of cells (<1%), and to accumulate in cytoplasmic foci in more than 90% of cells grown into late stationary phase (data not shown). Its reduced accumulation in P-bodies is likely to reflect negative effects of the GFP-tag in combination with the C-terminal deletion, possibly by reducing its incorporation into the Lsm1-7p complex. The negative effect

<table>
<thead>
<tr>
<th></th>
<th>GFP</th>
<th>DAPI</th>
<th>merge</th>
<th>Merge</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>GFP Lsm4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GFP Lsm4C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GFP Lsm4ΔC</td>
<td>N.D.</td>
<td>N.D.</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2 Overproduction of Lsm4p or its C-terminus leads to aggregation in cytoplasmic foci. (A) GFP-Lsm4 (pMPSlsm4), GFP-Lsm4C (pMPSlsm4D2) and GFP-Lsm4ΔC (pMPSlsm4D1) were overexpressed from the MET25 promoter in BY4741 cells grown in SD-Ura-Met. Localization was examined in cells during log-phase growth. (B) Colocalization of Lsm4p aggregates with Dcp2-RFP (pRP1155) was examined in BY4741 cells grown in SD-Ura-Leu-Met during log-phase growth or 20 minutes after hypo-osmotic shock. Dcp2-RFP (pMR171) localization was examined in PGAL-LSM4 cells expressing GFP-Lsm4ΔC grown in SD-Ura-His-Met (to prevent competition between GFP-Lsm4ΔC and endogenous Lsm4p for incorporation into Lsm1-7p), 20 minutes after hypo-osmotic shock. (C) Localization of GFP-Lsm1 (pGFP-N-Lsm1), GFP-Lsm6 (pMPSlsm6), Lsm8-GFP (pMR83) and GFP (pGFP-N-FUS) was examined in log-phase cells overexpressing Lsm4p (PGAL-LSM4 cells grown in SDGal-Ura) and in cells with normal levels of Lsm4p (PGAL-LSM4 cells with PUS1 grown in SD-Ura-Met). Nuclear DNA stained with DAPI is shown in blue.
of the GFP-tag is emphasized by a slow growth defect of the GFP-Lsm4ΔC strain at all temperatures compared with the lsm4ΔC strain (with non-tagged protein expressed from its native promoter), which shows slower growth only at 37°C (supplementary material Fig. S5). We cannot formally rule out the possibility that the difference between the non-tagged lsm4ΔC and the GFP-Lsm4ΔC strains is caused by their different levels of expression (native promoter vs MET25 promoter), although this is more likely to lead to the opposite of what we observe. While this manuscript was in preparation Mazzoni et al. reported that the asparagine-rich N-terminal region of the K. lactis Lsm4 protein, KlLsm4p, which is able to functionally replace its S. cerevisiae homologue, is essential for its own localization to P-bodies in budding yeast (Mazzoni et al., 2007). However, these authors only investigated the localization of a GFP-tagged version of this protein, and not the localization of other Lsm proteins in a strain expressing non-tagged klLsm4ΔCp. Thus the effect of the deletion may not have been distinguished from the additional, detrimental effect of the tag.

Absence of the Lsm4p C-terminus affects mRNA decay
To determine whether the Lsm4p C-terminal deletion affects mRNA decay, degradation of a PGK1pGmini reporter transcript (Mitchell and Tollervey, 2003) was investigated. This reporter is expressed from the GAL1 promoter, allowing its transcription to be switched off by growth on glucose. The rate of subsequent disappearance of the reporter transcript is used as a measure of its 5’ to 3’ degradation through the major mRNA decay pathway. A small effect was observed, as the mRNA half-life increased from 3.4±0.9 minutes in wild type to 4.7±1.1 minutes in the lsm4ΔC strain, on the basis of the quantitative reverse-transcriptase PCR (qRT-PCR) data presented in Fig. 4C. Half lives calculated using data obtained from the northern blot were slightly higher compared with those determined by qRT-PCR, but the relative difference between the two strains was similar. In addition, the steady-state level of this transcript appears to be about 40% higher in the lsm4ΔC cells compared with the LSM4 cells (Fig. 4B). By contrast, no effect was observed on the splicing of pre-U3 RNA, compared with 12 hours of Lsm5p depletion (Fig. 4D), suggesting that Lsm4ΔCp does not detrimentally affect formation of Lsm2-8p or stability of U6 snRNA. It therefore seems unlikely that the stability or formation of Lsm1-7p is reduced because of this C-terminal deletion, unless the assembly requirements of these two complexes are significantly different. A similar effect on mRNA degradation was reported for klLsm4ΔC in K. lactis (Mazzoni et al., 2003a), whereas a seemingly stronger effect was observed for klLsm4ΔC in S. cerevisiae (Mazzoni et al., 2003b). The latter may reflect reduced incorporation of the mutant K. lactis Lsm4p into the S. cerevisiae Lsm1-7p complex. Decker et al. did not find a significant change in the half-lives of PGK1pG or MFA2pG reporter transcripts in the absence of the C-terminal 97 aa of Lsm4p, nor did they report on increased steady-state levels of these transcripts (Decker et al., 2007). The reason for this difference remains unclear; however, the strains used in these studies were constructed in different ways and in different genetic backgrounds. We cannot formally rule out that the effect we see on the PGK1pG half-life is caused by reduced expression and/or stability of Lsm4ΔCp, as we have no antibody to compare its level with that of full-length Lsm4p. However, the absence of an effect on splicing argues against this.
Q/N-rich regions in other P-body components

Investigation of the aa sequences of all core components of P-bodies in yeast (Parker and Sheth, 2007) reveals Q and/or N-rich stretches of varying length in many of them, most of which are conserved between various *Saccharomyces* species (Fig. 5, Table 1 and supplementary material Fig. S6). Some (Lsm4p, Ccr4p, Pop2p and Not1p) were previously found in a genome-wide screen looking for yeast proteins with Q/N-rich domains (Michelitsch and Weissman, 2000). Michelitsch and Weissman used an algorithm to count these residues in consecutive aa 80-mers for each of the predicted open reading frames, finding an average Q/N-content of 7.7 per 80-mer in the yeast proteome (Michelitsch and Weissman, 2000). We counted Q, N and P residues in a similar fashion in each of the P-body core components. Our results (Table 1) show that all of the 20 proteins tested score above average for Q/N content (Graphic representations are shown in Fig. 5B,C and supplementary material Fig. S6). Interestingly, some of the *Saccharomyces* homologues show further extensions of Q repeats, e.g. Edc3p, Not3p, Not4p and Not5p (supplementary material Fig. S6). In addition, many of these polypeptides contain high numbers of proline residues in or just downstream of these Q/N-rich regions (Table 1). This is a feature that is also found in other aggregation-prone proteins, e.g. huntingtin, aggregation of which causes Huntington disease (Michelitsch and Weissman, 2000). Proline-rich regions often form extended and flexible regions, in many proteins apparently reaching out to facilitate interactions with other proteins, with phosphorylation having a potential regulatory role. Binding via these proline-rich domains is generally not very specific, but can be both very rapid and strong (Williamson, 1994; Kay et al., 2000). Furthermore, proteins with Q/N-rich domains have...
Fig. 5 Other P-body components contain Q/N-rich regions. (A) Alignment of the N-terminal region of S. cerevisiae Ccr4p with homologues from closely related Saccharomyces species (see Fig. 1 legend for abbreviations). The Q/N-rich region is boxed, with Q residues highlighted in grey, N residues highlighted in black and P residues in red. (B) Schematic representation of P-body components with areas rich in Q and/or N residues indicated in white (approximately to scale; Not1p is broken to fit; lengths are indicated in numbers of aa). (C) Q, N and P residues were counted in aa 80-mers of Ccr4p, Pop2p, Not1p and Dhh1p starting at position 1, shifting ten aa at a time.

Table 1. Q, N and P residues counted per amino acid 80-mers in 20 different P-body components

<table>
<thead>
<tr>
<th>Protein</th>
<th>Q/N*</th>
<th>Q</th>
<th>N</th>
<th>P in Q/N-rich</th>
<th>P (in best 80-mer)*</th>
<th>Percentage of P (in best 80-mer)</th>
<th>P-rich close to Q/N-rich</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lsm4</td>
<td>38</td>
<td>5</td>
<td>33</td>
<td>1</td>
<td>2</td>
<td>2.5</td>
<td>yes</td>
</tr>
<tr>
<td>Pop2</td>
<td>37</td>
<td>9</td>
<td>28</td>
<td>6</td>
<td>12</td>
<td>15.0</td>
<td>yes</td>
</tr>
<tr>
<td>Ccr4</td>
<td>34</td>
<td>11</td>
<td>23</td>
<td>4</td>
<td>10</td>
<td>12.5</td>
<td>yes</td>
</tr>
<tr>
<td>Not1</td>
<td>30</td>
<td>26</td>
<td>4</td>
<td>1</td>
<td>12</td>
<td>15.0</td>
<td>yes</td>
</tr>
<tr>
<td>Edc3</td>
<td>29</td>
<td>19</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>6.3</td>
<td>yes</td>
</tr>
<tr>
<td>Dhh1</td>
<td>25</td>
<td>25</td>
<td>0</td>
<td>17</td>
<td>19</td>
<td>23.8</td>
<td>yes</td>
</tr>
<tr>
<td>Dcp2</td>
<td>24</td>
<td>13</td>
<td>11</td>
<td>8</td>
<td>20</td>
<td>25.0</td>
<td>yes</td>
</tr>
<tr>
<td>Pat1</td>
<td>23</td>
<td>19</td>
<td>4</td>
<td>18</td>
<td>23</td>
<td>28.8</td>
<td>yes</td>
</tr>
<tr>
<td>Scd6</td>
<td>22</td>
<td>8</td>
<td>14</td>
<td>2</td>
<td>15</td>
<td>18.8</td>
<td>yes†</td>
</tr>
<tr>
<td>Dcp1</td>
<td>22</td>
<td>4</td>
<td>18</td>
<td>2</td>
<td>4</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>Edc1</td>
<td>22</td>
<td>5</td>
<td>17</td>
<td>10</td>
<td>10</td>
<td>12.5</td>
<td>yes</td>
</tr>
<tr>
<td>Not4</td>
<td>21</td>
<td>11</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>12.5</td>
<td>yes</td>
</tr>
<tr>
<td>Pan3</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>6</td>
<td>15</td>
<td>18.8</td>
<td>yes</td>
</tr>
<tr>
<td>Not3</td>
<td>19</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>13</td>
<td>16.3</td>
<td>yes†</td>
</tr>
<tr>
<td>Not5</td>
<td>15</td>
<td>8</td>
<td>7</td>
<td>3</td>
<td>12</td>
<td>15.0</td>
<td>yes†</td>
</tr>
<tr>
<td>Xrn1</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>24</td>
<td>30.0</td>
<td></td>
</tr>
<tr>
<td>Pan2</td>
<td>15</td>
<td>7</td>
<td>8</td>
<td>2</td>
<td>10</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>Lsm1</td>
<td>14</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Not2</td>
<td>11</td>
<td>8</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>Edc2</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>11.3</td>
<td></td>
</tr>
</tbody>
</table>

*Values indicate the number of residues in the highest scoring 80-mer for that protein.
†P-rich region near Q/N-rich region scoring <20.
terminal prion-like domain of Lsm4p (Decker et al., 2007). Rnq1 prion protein can, at least in part, functionally replace the C-terminal prion-like domain of Sup35p, which generally leads to larger and more abundant P-bodies by promoting de novo appearance of [PSI+], the prion-form of the Q/N-rich regions. Lsm4p is semi-conserved between Saccharomyces species, at least in the high content of N regions as well as other parts of these proteins for these interactions.

Previously been shown to promote aggregation of heterologous proteins with similar domains (Derkatch et al., 2004). Indeed, Lsm4p was found as one of nine Q/N-rich proteins that, when overproduced, promote de novo appearance of [PSI+], the prion-form of the Q/N-rich Sup35 protein (Derkatch et al., 2001). On the basis of this behaviour as well as its structural similarities to Sup35p, these authors proposed that Lsm4p itself is a prion protein. Furthermore, Decker et al. showed that the prion-like Q/N-rich domain of the Rnq1 prion protein can, at least in part, functionally replace the C-terminal prion-like domain of Lsm4p (Decker et al., 2007).

Q/N-rich regions affect P-body localization

It is plausible that Q, N and/or P-rich regions have a role in the accumulation of proteins in P-bodies. We tested Q/N-rich regions from Ccr4p, Pop2p and Dhh1p for their ability to aggregate and/or accumulate in P-bodies when fused to GFP. The Q/N-rich N-terminal region of Ccr4p fused to GFP [Ccr4p(1-229)] aggregates in cytoplasmic foci under normal growth conditions (Fig. 6A,B) in ~20% of cells, and foci increase in numbers under stress conditions, with more than 50% of cells showing multiple foci per cell. Although the dynamics of increased focal accumulation resembled that of P-body formation, suggesting that the Q/N-rich N-terminus of Ccr4p is sufficient for P-body localization, we found that the majority did not colocalise with Dcp2-RFP (Fig. 6C). GFP-fusions of the Q/N-rich regions of Pop2p [Pop2p(1-156)] and Dhh1p [Dhh1p(427-506)], however, do not aggregate under normal growth conditions but show weak focal concentration in a low percentage of cells (~1%) when stressed, although the majority of cells do not show a change in localization (Fig. 6A). However, GFP-fusions of Pop2p and Dhh1p deleted for these domains [Pop2ΔN(147-433)] and [Dhh1ΔC(1-427)] do show decreased P-body localization compared to full-length Pop2p and Dhh1p (Fig. 7A), and Ccr4p deleted for 147 aa at its N-terminus [Ccr4ΔN(148-837)] completely fails to accumulate in cytoplasmic foci under stress conditions. We quantified the P-body localization of these proteins by counting the number of visible foci per cell at a set time after osmotic shock (Table 2). These numbers are an indication of the level of P-body localization, as a reduction in P-body accumulation will lead to a reduction in the number of visible P-bodies, which generally have variable sizes and/or intensities. Interestingly, deletion of a further 102 aa from the Ccr4p N-terminus [Ccr4ΔN(250-837)] leads to exclusively nuclear localization (Fig. 7B). The latter suggests that Ccr4p normally shuttles between the nucleus and cytoplasm, and that its nuclear export depends on sequences within the N-terminal domain. The tendency for aggregation of these Q/N-rich regions is further emphasized by the fact that full-length Pop2p expressed from the MET25 promoter aggregates in bright nuclear foci when tagged at the C-terminus (Pop2-GFP, Fig. 7C), at a much lower rate when tagged at the (Q/N-rich) N-terminus (data not shown) and not at all in the absence of this N-terminus (Pop2ΔN-GFP, Fig. 7C). As these experiments were performed in the presence of natively expressed non-tagged proteins, which may contribute to the observed absence of GFP-Ccr4ΔN concentration in P-bodies, we investigated the localization of this protein in ccr4Δ as well as xn1ΔΔ strains. Whereas GFP-Ccr4ΔN still failed to concentrate in microscopically visible foci in the absence of native Ccr4p (Fig. 7D), some weak foci were observed in the absence of Xrn1p (supplementary material Fig. S7), which generally leads to larger and more abundant P-bodies by preventing 5′-to-3′ degradation of transcripts. For Ccr4ΔNp and Pop2ΔNp the reduced P-body localization is not due to reduced levels of these truncated proteins as western analysis showed no difference between levels of full-length and mutant proteins (Fig. 7E). As the level of Dhh1ΔCp was only 60% of that of full-length Dhh1p we cannot rule out that its reduced P-body localization is, in part, due to the lower protein level.

In summary, although not absolutely essential, Q/N-rich sequences in Pop2p, Ccr4p and Dhh1p contribute to efficient accumulation of these proteins in P-bodies under stress conditions. This is most obvious for Ccr4p, which, in the absence of its N-terminal 147 aa is not microscopically detectable in P-bodies in otherwise normal cells. Increased focal accumulation under stress conditions of the N-terminal 229 aa fused to GFP, suggests that this region is capable of regulated aggregation in response to stress. The fact that the majority of these foci do not colocalise with Dcp2p, suggests that additional parts of Ccr4p are necessary for proper P-body localization, most probably through additional protein-protein interactions. It would therefore be interesting to further investigate the requirements of the Q/N-rich regions as well as other parts of these proteins for these interactions.

Is a mechanism for protein accumulation in P-bodies conserved?

As the C-terminal region of S. cerevisiae Lsm4p is semi-conserved between Saccharomyces species, at least in the high content of N regions...
and/or Q residues (supplementary material Fig. S1), the ability to promote Lsm1-7p accumulation in P-bodies is likely to be conserved in these yeasts as well as in other budding yeasts. In fact, this was shown to be true for the budding yeast *K. lactis* Lsm4p produced in *S. cerevisiae* (Mazzoni et al., 2007). The human homologue, however, does not show a significant enrichment in Q or N residues, apart from a short stretch of five glutamines. Indeed, full-length human LSM4 fused to GFP did not aggregate when overexpressed in wild-type yeast cells, nor did it accumulate in foci under stress conditions. Surprisingly, it mostly accumulated in the nucleus instead (data not shown). As it was not able to support viability in the absence of native Lsm4p expression (data not shown), it might be unable to form a functional complex with yeast Lsm proteins. It is possible that residues in other human LSM1-7 complex (comprising LSM1 to LSM7) members normally contribute to its accumulation in P-bodies. Notably, the short N or C-terminal extensions of LSM1, LSM2, LSM3 and LSM7 proteins contain relatively high levels of glutamine residues. However, if, as we propose, Q/N-rich sequences contribute to a rapid response to stimuli in yeast, this may not be needed in human cells, as the LSM1-7 complex accumulates in P-bodies even under normal growth conditions.

Q/N-rich regions do not seem to be conserved in the human homologues of budding yeast Ccr4p, Pop2p and Dhh1p (CNOT6, Journal of Cell Science 121 (15))

![Fig. 7](image-url) Q/N-rich regions from Ccr4p, Pop2p and Dhh1p contribute to efficient accumulation of these proteins in P-bodies. (B) Localization of GFP-tagged full-length Ccr4p (pMR212), Pop2p (pMR214), Dhh1p (pMR210) or truncated versions of these proteins (Ccr4ΔN(148-837) from pMR218, Pop2ΔN(147-433) from pMR215, Dhh1ΔC(1-427) from pMR211) before and after osmotic shock. All GFP-fusions were expressed in BY4741 cells and localization was examined in cells during normal growth (normal) or 20-40 minutes after osmotic shock (stress). (B) Localization of GFP-tagged Ccr4ΔN(148-837) and Ccr4ΔN2 (aa 250-837; pMR213) in fixed cells with DAPI-stained nuclear DNA. (C) Localization of C-terminally GFP-tagged full-length Pop2p (pMR216) or Pop2ΔNp (pMR217); DAPI-stained nuclear DNA in blue. (D) Localization of GFP-Ccr4, GFP-Ccr4ΔN and Dcp2-RFP in ccr4Δ cells (Y10387) 30 minutes after hypo-osmotic shock. (E) Anti-GFP western blot analysis of full-length and truncated Lsm4, Ccr4, Pop2 and Dhh1 proteins. Curiously, GFP-Ccr4 (122 kDa) migrates faster than GFP-Ccr4ΔN (106 kDa), but slower than GFP-Ccr4ΔN2 (95 kDa; Fig. 7E), and all three GFP-Ccr4 proteins migrate faster than their predicted molecular weights. The presence or absence of the highly polar N-terminal region of Ccr4p causes a change in the effective charge of the entire protein (predicted charges at pH 7 are –5.8, –4.0 and –5.5 respectively) and might affect protein conformation resulting in unusual migration during SDS-PAGE. Nop1p was used as a loading control.
CNOT8 and DDX6, respectively) proteins either; they are significantly shorter, lacking the N-terminal and C-terminal Q/N-rich regions respectively (supplementary material Fig. S8). Perhaps the function of these protein domains has been replaced by alternative domains, possibly in other polypeptides with which they interact. For example, GW182 contains an internal Q/P-rich region that is essential, but not sufficient, for its own P-body localization and that of Ago1 (Behm-Ansmant et al., 2006). Another P-body component specific to higher eukaryotes, Ge-1/Hedls, contains a C-terminal repetitive sequence rich in hydrophobic residues that is essential for P-body localization and parts of which aggregate in cytoplasmic foci that are not P-bodies (Yu et al., 2005). In addition EDC4 (also known as Ge-1, Hedls), DCP2 and TNRC6B from humans as well as other higher eukaryotes contain high levels of Q and/or N residues (Decker et al., 2007). Thus, alternative aggregation-prone regions might have replaced some of the yeast Q/N-rich domains in higher eukaryotes, at least some of which are likely to have a role in P-body assembly.

Aggregation of P-body components through their Q/N-rich regions could promote efficient P-body formation. Whether this is really the case and, if so, whether this occurs through prion-like aggregation or through specific interactions via putative modular ‘polar zipper’ protein-protein interaction domains (Perutz et al., 1994; Michelitsch and Weissman, 2000) remains to be determined. The importance of the Q/N-rich protein Edc3p (also known as Lsm16p) in combination with Lsm4p in P-body assembly in yeast, which came to light while this manuscript was being revised (Decker et al., 2007), is in support of this hypothesis. An intriguing question is how Lsm4p aggregation, and that of other P-body components, is prevented under normal growth conditions. Post-translational modifications, e.g. phosphorylation of Lsm4p or other (Lsm) proteins, probably have a role. Such modifications could allow the cell to respond quickly and efficiently to changes in conditions, and might regulate the levels and intracellular localizations of Lsm1-7p and Lsm2-8p, in addition to promoting P-body localization. Such a mechanism could also regulate the competition between these two complexes that was observed by Spiller et al. (Spiller et al., 2007b). Similarly, post-translational modifications, e.g. of the N-terminal region of Ccr4p, could allow P-body localization of other proteins involved in RNA degradation. Q-rich regions in mouse TIA-1 and PUM2 have previously been shown to contribute to protein accumulation in stress granules (Gilks et al., 2004; Vessey et al., 2006). We now show that at least some of the Q/N-rich domains in P-body components have a role in the assembly of these RNA processing bodies. The presence of Q/N-rich regions in many other proteins that are involved in various aspects of RNA metabolism (Michelitsch and Weissman, 2000; Decker et al., 2007) hints at the possibility of a more general role for these prion-like domains in functional protein aggregation, in addition to stress-granule and P-body assembly.

Table 2. Quantification of P-body localization

<table>
<thead>
<tr>
<th>Protein</th>
<th>Foci per cell</th>
<th>Cells examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ccr4</td>
<td>2.7±0.5</td>
<td>138</td>
</tr>
<tr>
<td>Ccr4AN (aa148-837)</td>
<td>0</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>Pop2</td>
<td>4.8±1.6</td>
<td>80</td>
</tr>
<tr>
<td>Pop2AN (aa147-433)</td>
<td>1.9±0.6</td>
<td>93</td>
</tr>
<tr>
<td>Dihl1</td>
<td>10.7±3.3</td>
<td>35</td>
</tr>
<tr>
<td>Dihl1AC (aa1-427)</td>
<td>3.5±1.1</td>
<td>83</td>
</tr>
</tbody>
</table>

Materials and Methods

Plasmids and strains
For a complete list of plasmids and strains used see supplementary material Tables S1 and S2.

Microscopy
Cells were grown at 30°C to mid-log phase in synthetic dropout (SD) medium. To stress cells, cultures were centrifuged and cells were resuspended in water. Live cells were placed on microscopy slides and examined by bright-field and/or fluorescence microscopy using a Leica WF4000 fluorescence microscope. Fixing of cells followed by DAPI staining was performed as previously described (Spiller et al., 2007b). Images were captured using LeicaFW4000 software (Scanalytics, Fairfax, VA) with a CH-250 16-bit, cooled CCD camera (Photometrics, Tucson, AZ).

RNA analyses
 Cultures were grown at 30°C in synthetic dropout medium containing 2% (w/v) galactose. Transcription of the PGK1pGmini reporter gene was stopped by addition of glucose to 4% (w/v) and 20 ml culture with an OD600 of 0.5 were snap-chilled at the indicated times after the addition of glucose. RNA extractions and northern blot analyses of 6% acrylamide/urea gels were as described (Maries et al., 1999). The following oligonucleotide probes were used for northern hybridizations: to detect the PGK1pGmini reporter transcript 5’-AACCCGCGTCGAAGTAAACC-3’, to detect scR1 RNA 5’-ATCCCGCCGCTTCTACCAC3’ and to detect U3 RNA 5’-GTTATGGGACTCTCACTCA-3’.

Quantitative reverse-transciptase PCR
Ten μg of total RNA were treated with DNase1 (0.9 μg U Q1, Promega) according to the manufacturer’s instructions. cDNA was prepared from 5 μg of DNase-treated RNA in a 10 μl reaction: 1× first strand synthesis buffer, 2.5 mM DTT, 10 μg RNase inhibitor (Roche), 0.75 mM dNTPs, 7.5 μg ThermoScript RNaseH (Invitrogen) and 500 nM of PGK1pGmini-specific primer (5’-AGCGTAAAGGATGGGGAAAGGAAG-3’), according to the manufacturer’s instructions. A negative control reaction was performed in the absence of reverse transcriptase (RT). Any remaining RNA was hydrolysed by incubating reactions for 1 hour at 37°C after addition of 15 μl of 0.1 mg/ml RNaseA (Roche). Quantitative PCRs (qPCRs) were performed with SYBR Green JumpStart Taq ReadyMix (Sigma) in a Stratagene MX3000P real-time PCR machine in 10 μl reactions: 6 μl containing 5 μl 2× SYBR Green ReadyMix, 300 nM of each primer (F: 5’-ATTTGAAAAATGAAATCAGGAGGATTTTG-3’; R: 5’-AGGCTAAAGGATGGGGAAAGGAAG-3’) and 0.5X ROX, plus 4 μl of cDNA template (diluted 1 in 20 after RT-PCR). Cycling parameters were as follows: 2 minutes at 94°C, then 50 cycles of 10 seconds at 94°C, 10 seconds at 63°C and 20 seconds at 72°C. Each qPCR reaction was performed in triplicate for each repeat RT reaction.

Western analysis
For crude protein extracts (Volland et al., 1994), yeast cells were lysed in 0.5 ml of 0.2 M NaOH on ice for 10 minutes, followed by TCA precipitation (final 5% w/v) for 10 minutes on ice. After centrifugation, the pellet was resuspended in 35 μl of dissociation buffer (0.1 M Tris-HCl pH 6.8, 4 mM EDTA, 4% SDS, 20% (v/v) glycerol, 2% (v/v) β-mercaptoethanol, 0.02% (w/v) BPB) and 15 μl of 1 M Tris base. Samples were heated at 95°C for 10 min before separation by SDS-PAGE. Proteins were transferred to PVDF membrane and detected with mouse anti-GFP (BD Bioscience) or anti-Nop1p antibodies, and sheep anti-mouse IgG-HRP (Amersham Bioscience).

Polypeptide alignments
Amino acid sequences of P-body components were obtained from the Saccharomyces Genome Database (http://www.yeastgenome.org/) or the NCBI Entrez Protein database (http://www.ncbi.nlm.nih.gov/sites/entrez). Alignments were made using the ClustalW Multiple Sequence Alignment tool (Thompson et al., 1994) inside Jalview 2.2 (Clamp et al., 2004).

We thank Roy Parker and David Tolfer for reagents, Jon Houseley for the critical reading of the manuscript, and David Barrass, Daniela Hahn and Olivier Cordin for assistance. This work was funded by a Wellcome Trust Prize Studentship 71448 and Royal Society support to MAMR, a studentship from The Darwin Trust of Edinburgh to MPS and Wellcome Trust Grant 067311. J.D.B. is the Royal Society Darwin Trust Research Professor. M.A.M.R. designed and performed all experiments under the supervision of J.D.B., except the qPCR experiment which was performed by R.D.A.; M.P.S. first suggested a
References


Séraphin, B. (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* 14, 2089-2098.


