THE EMBO JOURNAL

Manuscript EMBO-2009-73108

A systematic survey of in vivo obligate chaperonin-dependent substrates

Kei Fujiwara, Yasushi Ishihama, Kenji Nakahigashi, Tomoyoshi Soga, Hideki Taguchi

Corresponding author: Hideki Taguchi, University of Tokyo

Review timeline:

Submission date:	10 November 2009
Editorial Decision:	23 December 2009
Revision received:	08 February 2010
Editorial Decision:	26 February 2010
Revision received:	03 March 2010
Additional correspondence:	04 March 2010
Accepted:	04 March 2010

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 23 December 2009

Thank you for submitting your manuscript for consideration by The EMBO Journal. It has now been seen by three referees whose comments to the authors are shown below. As you will see while referee 2 raises concerns regarding the conceptual advance provided by the study and is not in favour of publication of the study here the other two referees are more positive and would support publication of the study here after appropriate revision. Taking together all issues that were raised we have come to the conclusion that we would be able to consider a revised version of this manuscript. However, you need to address the concerns raised by the referees in an adequate manner and to their satisfaction in particular those regarding the comprehensive nature of study and the criteria used to identify GroE-dependent proteins as put forward by referee 3. In addition, during the initial assessment of the study concerns were raised whether the solubility is a sufficient criterion for correct protein folding as the proteins are His tagged and this is known to affect the solubility properties of proteins. This issue should be addressed as well.

I should remind you that it is EMBO Journal policy to allow a single round of revision only and that, therefore, acceptance or rejection of the manuscript will depend on the completeness of your responses included in the next, final version of the manuscript as well as on the final assessment by the referees.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process initiative, please visit our website:

http://www.nature.com/emboj/about/process.html

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Editor
The EMBO Journal

———

REFEREE REPORTS:

Referee #1 (Remarks to the Author):

This manuscript presents analysis of E.coli chaperonin substrates previously identified by Kerner et al. [Cell 122, pp. 209-220 (2005)] as potential obligate GroEL substrates. Here, the authors test actual effects of depleting GroEL using a well-controlled approach and conclude that only about 60% of suggested class III substrates strictly require GroE for solubility in vivo, while other 40% were found to be less dependent on theGroE system. They use their analysis to create a new grouping of chaperonin-dependent substrates, class IV, and attempt to learn what the properties of an In vivi GroEL substrate are. This study is of interest to broad scientific audience. The paper contributes significantly to our understanding of the role of chaperones by pointing to structural and biochemical properties of a typical GroE substrate and identifying the essential identified class IV substrates.

The only thing lacking in this very comprehesive study is in vitro refolding of several class III- and class III+ (class IV) GroE substrates in the absence and presence of chaperone systems like DnaK and GroE. The suggested experiment is of particular importance for the substrates which were not originally assigned to be class III substrates.

Note that there is a huge amount of data in the Supplementary Material. Much is not requiresd for this study. The authors might consider removing some for a later paper.

Specific points that should be addressed:
Results (pp. 6-8): In the manuscript, obligate GroE in vivo substrates were indentified as those 'whose soluble abundances were reduced during depletion to less than 50% of that found during arabinose growth in MGM100 (as a genetic control) and druing glucose growth in MG1655 (as a sugar control)'. First, the authors do not explain why the cutoff value was chosen to be 50% and how different the substrate sets would be if the cutoff value changed. Second, it is only mentioned, but not shown, how different from normal (as in MG1655) the levels of GroEL and GroES were at the time of lysis. A Western-blot with anti-GroEL and anti-GroES antibodies with MGM100 cells grown in arabinose and glucose as well as MGM1655 cells would provide required evidence. Thirdly, depletion of GroE might have resulted in depletion of proteins which are not GroE substrates. One can consider GroE depletion to less than 10% of its basal level as a stress, and not only GroE substrates might be downregulated as a part of cellular stress response. And, finally, the authors do not mention that their proteomics-based approach detected only about a half (43 of 84) of potential class III substrates suggested by Kerner et al. (2005) (Fig. 7A in the referred to paper).

Results (pp. 12-13): According to proteomics data, solubilities of 24 class II and 3 class I GroE substrates were reduced more than 50% in GroE-depleted MGM100 E.coli cells (Fig. 1C). Together they make 27 non-class III GroE substrates meeting the authors' criteria for obligate GroE in vivo substrate. On the other hand, on p. 13 the authors report: 'we found 8 new in vivo obligate GroE substrates not previously identified as Class III proteins'. The reason why other 19 proteins were not included is not clearly stated in the text.

Discussion (pp. 17-19): Here the authors mention that GroE may have a maintenance function for relatively unstable proteins, suggesting Class III- substrates as candidates, but this directly

contradicts the experimental data: Class III- GroE substrates are not rapidly degraded in vivo under GroE-depleted conditions, thus being 'resistant to proteases even when overexpressed' (p. 18).

Discussion (pp. 19-20): In the first sentence the authors mention that their 'survey revealed that 60% (49 out of 83) of Class III substrates were Class III+'. However, according to Fig. 1C, they detected total 43 out of 84 class III substrates suggested by Kerner et al. (2005), and only 24 of them were reduced more than 50% and, thus, represent Class III+ substrates.

Unfortunately, some experimental results are not properly discussed in the context of the existing literature:
(1) Introduction section. On p. 4 (bottom) the authors say that 'in vivo GroE dependency of class III susbtrates has not been tested, except that DapA, GatY, MetK, ADD and YajO were verified as requiring GroE for folding in yeast growing on a synthetic medium (Kerner et al., 2005)'. In addition to the experiment mentioned, however, solubility of GroE class III substrates has been tested in E.coli under conditions of substrate overexpression or GroE depletion (pp. 211-212 of the paper referred to, Fig. 2)
(2) Results section. On pp. 8(bottom)-9(top) the authors claim that their results on everexpression in GroE-depleted cells is in 'complete consistence with the previous data on the GroE dependency'. It must be noted, however, that in the study by Kerner et al. (2005) METK and DAPA, class III GroE substrates, behaved differently in GroE-depleted E.coli cells (strain MC4100): While DAPA was degraded, METK aggregated (Fig. 2 in the cited paper). In the manuscript by Fujiwara et al. the authors show that both DAPA and METK aggregated (Fig. 2A) as indicated by the presence of the protein in the total lysate, but not soluble fraction. This point should be properly discussed since the difference may result from E.coli strain, induction time, medium etc.

Referee #2 (Remarks to the Author):

Hartl and co-workers identified in a study published in 2005 a set of GroE substrates in vivo that they divided into 3 classes depending on the extent to which they require the GroE system for folding. In this paper, it is shown using various criteria (i.e. extent of aggregation, activity and enrichment as an indication of solubility upon GroE deletion) that about 40% of class III proteins, suggested by Hartl and co-workers to be obligate substrates, are actually not obligate. In addition, some previously unidentified obligate substrates are described. The paper also contains some bioinformatics analysis of class IV proteins defined in the present study to be truly obligate. Some aspects of the paper are not convincing and, although it is clearly written, extensive editing is still needed. A revision should address the points below.

Comments:
1. p. 7, bottom paragraph - I would define candidate GroE obligate substrates as those depleted by 50% during arabinose growth in MGM100 (as do the authors) and, but unlike the authors, as those NOT depleted during glucose growth in MG1655.
2. p. 8, line 5 from top - 24 out of 43 is not 40%.
3. Figure 3 - Observing enzyme activity in GroE-depleted cells does not necessarily mean that folding of these enzymes is GroE-independent since a small amount of folded enzyme may suffice depending on the substrate concentration and the enzyme's kinetic parameters. The discussion in the paper must refer to this caveat.
4. p. 13 - the analysis of the sequence features of classes III- and IV ignores previous work by Noivirt-Brik et al. (2007) that should be discussed and cited. This point is also relevant regarding the discussion of the U. urealyticum proteome (p. 16).
5. Figure 5D - it is hardly surprising that differences in the solubilities of class IV and III- proteins are observed given that solubility was a major criterion in assigning them to these classes. Such circular reasoning should be avoided.
6. The paper would benefit from a comparison with the data of Chapman et al.

Minor comments:
1. p. 4, bottom lines - the authors appear to be confused when they write 'folding in yeast growing on a synthetic medium'.
2. Figure 4C - the color-coding needs to be defined in the legend.
3. Is there a difference between classes III+ and IV? If not, why use both notations?

4. p. 20 - Apetri & Horwich should not be cited for accelerated folding which they argue does not occur.

Referee #3 (Remarks to the Author):

Fujiwara et al. characterize obligate GroE chaperonin substrates by quantifying the soluble expression of almost 1000 E. coli proteins with and without GroE depletion. The work initially undertakes a systematic screen for GroE-dependent proteins but then focuses only on the subset of proteins previously identified as GroEL-interacters, plus a few newly-identified GroE-dependent proteins. The major contribution of the paper is the systematic investigation of the soluble expression and biophysical description of the set of proteins that was previously identified by Hartl's group as highly enriched in the GroE-interacting fraction. This analysis is novel and important in that it distinguishes GroEL-interacters from GroE-dependent proteins and then provides plausible explanations for the distinction, namely that positive charge promotes binding without dependence and that the GroE-dependents are rich in ala and gly and therefore fold slowly and aggregate.

It is not clear how well this set of roughly 50 proteins represents the complete set of GroE-dependent proteins. The authors could project this fraction of 50/1000 onto the complete set of soluble proteins, but this is fraught with dangerous assumptions. Nevertheless, the authors have proceeded to speculate that the handful of essential proteins within the 50 could account for GroE essentiality. This presumes that basically all of the GroE-dependent proteins are now known. What if this is only half or less of the GroE-dependent proteins? The 1000 tested are likely to be the most abundant ones. Do we have any reason to expect the GroE-dependent proteins to be abundant? The authors should at least state the number of soluble proteins in E. coli so that we can compare with the number of tested proteins and then judge how well the GroE-dependent fraction may have been probed.

The group of GroE-dependent proteins identified in this work is not thoroughly described. The authors do not even provide the number of proteins that satisfy the criteria for "candidate GroE obligate substrates." They spend a lot of time on the 43 GroEL-interacters previously identified by Hartl, but how were found by this new method?

The paper's title, "comprehensive survey", is not justified without addressing the above two issues and making a persuasive argument that they have achieved comphensiveness.

The authors use poorly explained criteria to identify the GroE-dependents, those reduced in strain MGM100 by switching from ara to glc and those reduced in MGM100 in glc relative to MG1655 in glc. Part of Figure 1 indicates the fraction of proteins having reduced expression using the former, but it does not show the fraction reduced by the latter criterion. [What they show is the fraction reduced by ara to glc in MG1655, which is not a criterion and therefore is a bit misleading because it makes you think you've seen a description of both criteria.] They should describe (with numbers and a graph as in Fig. 1B) the similarity of the two protein sets captured by the criteria used and explain why these two criteria are preferable to replicates of ara vs. glc in MGM100.

On page 19, the authors should cite Aoki et al. (PMID 9405415) for showing that GroEL binding can be stabilized predominantly by electrostatics.

A number of typos on pp 20-21.

The authors should state that the TIM barrel is not simply the commonest fold (which would have been a trivial explanation for its dominance in GroE-dependence), and they should identify the folds c.2 and c.37 in Figure 5.

They should cite Lindquist for mutations and genetic diversity on p. 22.

1st Revision - authors' response 08 February 2010

Response to Editor
Editor's specific concern:
--- during the initial assessment of the study concerns were raised whether the solubility is a sufficient criterion for correct protein folding as the proteins are His tagged and this is known to affect the solubility properties of proteins. This issue should be addressed as well.

Our response:
We guess that our previous work on a global aggregation analysis using a reconstituted translation system (Niwa et al. PNAS 2009), in which the solubility of in vitro translated proteins was assessed by His-tagged sequences, might have been confused with the present study (Fujiwara et al.). In the present study, none of the proteins contains a His-tag. Overall, the results in the present study were obtained using proteins without any tag sequence. Only the results of a few substrates in Fig. 2B (DadA, YbjS and TrmD) and all of the results in Supplemental Fig S4 were obtained with HA-tagged proteins for specific detection by western blotting. Even in the HA-tagged proteins, we confirmed that the attachment of the HA-tag does not affect the chaperonin requirement, at least for FolE, SuhB (Class IIIñ), MetK, and HemB (Class IV) in cells (Fig. 2A, B and Supplemental Fig. S4B).

Response to Referee #1

Referee's comment:
The only thing lacking in this very comprehensive study is in vitro refolding of several class IIIñ and class III+ (class IV) GroE substrates in the absence and presence of chaperone systems like DnaK and GroE. The suggested experiment is of particular importance for the substrates which were not originally assigned to be class III substrates.

Our response:
As Referee #1 suggested, we conducted in vitro folding assays for several substrates and have described the results on p14 and in Supplemental Figure S8. In the folding assay, we translated a Class IIIñ (FolE) and several Class III+ (Class IV) proteins (DapA, which was originally assigned to Class III, and SerC and KdsA, which were not assigned to Class III) by a reconstituted cell-free translational system (PURE system, Shimizu et al. Nat. Biotechnol. 2001; Niwa et al. PNAS, 2009), which does not contain any chaperones. The requirements of chaperone systems, DnaK (DnaK, DnaJ, and GrpE) and GroE (GroEL and GroES), for the folding were monitored by the solubility and the appearance of a folded structure, defined as a sharp band in native PAGE. As shown in Supplemental Figure S8, FolE (Class IIIñ) was soluble and formed a folded structure, even in the absence of chaperones. In contrast, all of the Class IV proteins tested (DapA, SerC, and KdsA) were aggregation-prone without chaperones. The addition of the DnaK system increased the solubilities of the Class IV proteins to a greater or lesser extent, but folded structures were not detected in native PAGE, implying that the soluble but unfolded structures in the presence of DnaK might be easily degraded in vivo. The Class IV proteins were soluble and formed folded structures only in the presence of GroE. The in vitro folding assay further confirmed our conclusion that the Class IV substrates, including the substrates that were not assigned as Class III (SerC and KdsA), stringently require GroE for correct folding.

Referee's comment:
Note that there is a huge amount of data in the Supplementary Material. Much is not required for this study. The authors might consider removing some for a later paper.

Our response:
After consideration, we have removed several Supplemental Figures (former S1, S3 and S13) in the revised manuscript.

Referee's comment:
Specific points that should be addressed:
Results (pp. 6-8): In the manuscript, obligate GroE in vivo substrates were indentified as those 'whose soluble abundances were reduced during depletion to less than 50% of that found during

arabinose growth in MGM100 (as a genetic control) and during glucose growth in MG1655 (as a sugar control)'.
First, the authors do not explain why the cutoff value was chosen to be 50% and how different the substrate sets would be if the cutoff value changed.

Our response:
Among the previously identified in vivo obligate GroE substrates, MetK was the most soluble (46% solubility) under the GroE-depleted conditions. Therefore, we chose 50% as the cutoff value to minimize potential false negatives. By this criterion, 252 proteins were rough candidates for in vivo obligate GroE substrates, among the 986 proteins detected in the proteomics. We have included the above description in the revised manuscript.

Referee's comment:
Second, it is only mentioned, but not shown, how different from normal (as in MG1655) the levels of GroEL and GroES were at the time of lysis. A Western-blot with anti-GroEL and anti-GroES antibodies with MGM100 cells grown in arabinose and glucose as well as MGM1655 cells would provide required evidence.

Our response:
As Referee #1 recommended, we conducted western blotting with anti-GroEL and anti-GroES to confirm their depletion under glucose conditions. The western blots have been included as Supplemental Figure S1 in the revised manuscript.

Referee's comment:
Thirdly, depletion of GroE might have resulted in depletion of proteins which are not GroE substrates. One can consider GroE depletion to less than 10% of its basal level as a stress, and not only GroE substrates might be downregulated as a part of cellular stress response.

Our response:
This is an important point. Certainly, subsets of proteins are down-regulated or up-regulated in GroE-depleted cells, as a part of the cellular stress responses mediated by transcriptional shifts, and as exemplified by MetE up-regulation. Therefore, proteomics cannot simply identify in vivo obligate GroE substrates, but nevertheless provides valuable rough candidates. Therefore, we developed methods, independent of proteomics, to verify the GroE requirement, by the stress-independent expression of individual candidate proteins (Figs. 2 and 4, and Supplemental Figs. S3, S4, and S7).

Referee's comment:
And, finally, the authors do not mention that their proteomics-based approach detected only about a half (43 of 84) of potential class III substrates suggested by Kerner et al. (2005) (Fig. 7A in the referred to paper).

Our response:
Since the latest innovative proteomics (Masuda et al., Mol. Cell. Proteomics, 2009) was used in our work, the potential of proteomics in the present study is superior to that reported by Kerner et al. The coverage of about half of our proteomics for Class III proteins would be attributed to differences in the expression profiles of the proteins, depending on the E. coli strains, cultivation or medium conditions and so on.

Referee's comment:
Results (pp. 12-13): According to proteomics data, solubilities of 24 class II and 3 class I GroE substrates were reduced more than 50% in GroE-depleted MGM100 E.coli cells (Fig. 1C). Together they make 27 non-class III GroE substrates meeting the authors' criteria for obligate GroE in vivo substrate. On the other hand, on p. 13 the authors report: 'we found 8 new in vivo obligate GroE substrates not previously identified as Class III proteins'. The reason why other 19 proteins were not included is not clearly stated in the text.

Our response:
We used proteomics only for a rough screening; the cutoff value (50% reduction under GroE-depleted conditions) yielded 252 candidates, which included many false positives. Our criterion for

an in vivo obligate GroE substrate is whether the candidate proteins expressed under GroE-depleted conditions are soluble (e.g. Fig. 2), and is not simply based on the reduction in the GroE-depleted cells. Therefore, a substantial number of the candidates identified by the proteomics were not in vivo obligate GroE substrates. Indeed, among the proteomics-based candidates tested in our work, only four Class II proteins (KdsA, PyrC, SerC, and NuoC) were bona fide obligate GroE substrates in vivo (Figs. 2 and 4). In addition to the above-mentioned four Class II substrates, we newly identified four in vivo obligate substrates that had not appeared among the GroEL interactors (i.e. Classes I, II, and III) by either the proteomics or a homology search. In total, eight (4 previous Class II and 4 non-GroEL interactors) were identified.
Since our description on this issue was insufficient, we have modified the corresponding sentences to clarify this point.

Referee's comment:
Discussion (pp. 17-19): Here the authors mention that GroE may have a maintenance function for relatively unstable proteins, suggesting Class III- substrates as candidates, but this directly contradicts the experimental data: Class III- GroE substrates are not rapidly degraded in vivo under GroE-depleted conditions, thus being 'resistant to proteases even when overexpressed' (p. 18).

Our response:
We agree with Referee #1's comment. The discussion on the maintenance function has been altered to the following.
One plausible explanation is that the Class IIIñ substrates might interact with GroEL for relatively longer periods after translation, irrespective of the GroE requirement for folding. The prolonged interaction with GroEL, probably due to a strong affinity, would result in an enrichment of the substrates in the GroE complex, leading to the assignment of these proteins as Class III substrates, which were originally defined as being enriched with GroEL (Kerner et al., 2005). In this context, the previous observation that some of proteins >60 kDa exhibited very slow release from GroEL (Houry et al., 1999) seems to be correlated with our assignment of all Class III proteins >70 kDa as Class IIIñ.

Referee's comment:
Discussion (pp. 19-20): In the first sentence the authors mention that their 'survey revealed that 60% (49 out of 83) of Class III substrates were Class III+'. However, according to Fig. 1C, they detected total 43 out of 84 class III substrates suggested by Kerner et al. (2005), and only 24 of them were reduced more than 50% and, thus, represent Class III+ substrates.

Our response:
We apologize for the confusing description. As described above, proteomics was used for a rough screening of the candidate substrates. Also, our proteomics did not cover all of the Class III proteins (Fig. 1C), as Referee #1 pointed out. Therefore, the GroE requirement was examined by individually expressing the candidates to investigate their solubility under GroE-depleted conditions (Fig. 2, Supplemental Figures S3 and S4). The expression assay, which fully covered all of the Class III substrates, revealed that 49 out of 83 Class III substrates were GroE dependent in vivo, which is Class III+ in our definition.

Referee's comment:
Unfortunately, some experimental results are not properly discussed in the context of the existing literature:
(1) Introduction section. On p. 4 (bottom) the authors say that 'in vivo GroE dependency of class III susbtrates has not been tested, except that DapA, GatY, MetK, ADD and YajO were verified as requiring GroE for folding in yeast growing on a synthetic medium (Kerner et al., 2005)'. In addition to the experiment mentioned, however, solubility of GroE class III substrates has been tested in E.coli under conditions of substrate overexpression or GroE depletion (pp. 211-212 of the paper referred to, Fig. 2)

Our response:
We have removed the phrase "in yeast growing on a synthetic medium".

Referee's comment:

(2) Results section. On pp. 8(bottom)-9(top) the authors claim that their results on overexpression in GroE-depleted cells is in 'complete consistence with the previous data on the GroE dependency'. It must be noted, however, that in the study by Kerner et al. (2005) METK and DAPA, class III GroE substrates, behaved differently in GroE-depleted E.coli cells (strain MC4100): While DAPA was degraded, METK aggregated (Fig. 2 in the cited paper). In the manuscript by Fujiwara et al. the authors show that both DAPA and METK aggregated (Fig. 2A) as indicated by the presence of the protein in the total lysate, but not soluble fraction. This point should be properly discussed since the difference may result from E.coli strain, induction time, medium etc.

Our response:
As Referee #1 pointed out, the expression levels affected the fates of obligate GroE substrates. For example, MetK was degraded at low expression levels (e.g. leaky expression without IPTG, Supplementary Fig. S4B), and aggregated at high expression levels (e.g. overexpression by 1 mM IPTG, Fig. 2A). Therefore, we added the following sentences and modified the corresponding section (p. 9, top) accordingly.
Note that DapA was degraded in the previous report (Kerner, 2005), probably reflecting the difference in the expression levels. Except for the difference of whether DapA was degraded or aggregated, the present results are consistent with the previous data on the GroE dependency, suggesting that the overexpression strategy reflects the in vivo GroE dependency.

Response to Referee #2

Referee's comment:
1. p. 7, bottom paragraph - I would define candidate GroE obligate substrates as those depleted by 50% during arabinose growth in MGM100 (as do the authors) and, but unlike the authors, as those NOT depleted during glucose growth in MG1655.

Our response:
Actually, we had considered these criteria in our initial assessment. For the criteria suggested by Referee #2, GatY, one of the in vivo tested obligate GroE substrates previously identified by Kerner et al., was omitted from the candidates of the obligate GroE substrates. The difference would be derived from the observation that the expression levels of a subset of proteins were repressed during glucose growth, even in MG1655. For example, a protein that has emPAIglucose/emPAIarabinose = 1/10 and 4/10 in MGM100 and MG1655, respectively, meets our criteria, but is excluded in the criteria suggested by Referee #2.

Referee's comment:
2. p. 8, line 5 from top - 24 out of 43 is not 40%.

Our response:
We have corrected the value to 44%.

Referee's comment:
3. Figure 3 - Observing enzyme activity in GroE-depleted cells does not necessarily mean that folding of these enzymes is GroE-independent since a small amount of folded enzyme may suffice depending on the substrate concentration and the enzyme's kinetic parameters. The discussion in the paper must refer to this caveat.

Our response:
We agree with the comment. In particular, the function of FolE in GroE-depleted cells was estimated by the concentration of the metabolite (Fig. 3A), and not by the direct measurement of the enzymatic activity. Therefore, we have now included the following caveat in the revised manuscript. "However, several caveats should be stated regarding the activity of the Class IIIñ proteins. Observing enzyme activity in GroE-depleted cells does not necessarily mean that the folding of these enzymes is GroE-independent, since a small amount of folded enzyme may suffice, depending on the substrate concentration and the enzyme's kinetic parameters." (Discussion, p.20 top)

Referee's comment:
4. p. 13 - the analysis of the sequence features of classes III- and IV ignores previous work by

Noivirt-Brik et al. (2007) that should be discussed and cited. This point is also relevant regarding the discussion of the U. urealyticum proteome (p. 16).

Our response:
We have cited the reference and analyzed the sequences of the GroE substrates with FoldIndex. The FoldIndex distribution of the Class IV substrates was almost the same as that of the E. coli cytosolic proteome (Fig. S9B), suggesting that FoldIndex is not correlated with the in vivo GroE dependency. Nevertheless, we note that the FoldIndex distribution of Class IIIñ was lower than that in the E. coli cytosolic proteome (Fig. S9B), suggesting that FoldIndex might predict the preferential binding of proteins to GroE in cells. Regarding the Ureaplasma counterparts, we barely observed a significant difference between the 5 homologs in Ureaplasma and the corresponding Class IV proteins. A description of these analyses has been included in the revised manuscript (p15, p17 and Supplemental Figure S9B).
The tRNA adaptation index (tAI), which was also analyzed by Noivirt-Birk et al., is not discussed because this index is based on DNA sequences.

Referee's comment:
5. Figure 5D - it is hardly surprising that differences in the solubilities of class IV and III- proteins are observed given that solubility was a major criterion in assigning them to these classes. Such circular reasoning should be avoided.

Our response:
The global solubility data used in Fig. 5D were obtained by a reconstituted in vitro translation system, which is completely chaperone-free (Niwa et al., PNAS, 2009). Therefore, the in vitro data, which are independent of the current in vivo analysis, provide the inherent aggregation propensities of proteins in the absence of chaperones. Although the result shown in Fig. 5D might be expected, we consider it to be worth showing.

Referee's comment:
6. The paper would benefit from a comparison with the data of Chapman et al.

Our response:
We have included a comparison with the data of Chapman et al. in the Discussion section (p. 22, bottom).

Referee's comment:
Minor comments:
1. p. 4, bottom lines - the authors appear to be confused when they write 'folding in yeast growing on a synthetic medium'.

Our response:
"folding in yeast growing on a synthetic medium" has been removed.

Referee's comment:
2. Figure 4C - the color-coding needs to be defined in the legend.

Our response:
The color-coding has been defined in the Figure legend.

Referee's comment:
3. Is there a difference between classes III+ and IV? If not, why use both notations?

Our response:
Since the Class III+ substrates were defined as obligate GroE substrates derived from the Class III proteins identified by Kerner et al., the newly identified 8 obligate GroE substrates other than the Class III proteins (e.g. previous Class II proteins) could not be incorporated in Class III+. Therefore, Class IV was introduced as in vivo obligate substrates. Since ~85% (49/57) of the Class IV substrates were Class III+, the use of both notations in the manuscript might be confusing. The immediate use of Class IV instead of the introduction of Class III+ might be possible, but we still use "Class III+" to emphasize the fact that 60% of the Class III members identified by Kerner et al.

were actually in vivo obligate GroE substrates (Class III+), partly verifying the original proposal that the preferential GroE interactors are in vivo obligate substrates. In addition, only the notation of "Class IIIñ", one of the novel findings in our work, in Class III proteins would possibly cause confusion, as classes other than Class IIIñ were not defined.

Referee's comment:
4. p. 20 - Apetri & Horwich should not be cited for accelerated folding which they argue does not occur.

Our response:
The citation has been omitted.

Response to Referee #3

Referee's comment:
It is not clear how well this set of roughly 50 proteins represents the complete set of GroE-dependent proteins. The authors could project this fraction of 50/1000 onto the complete set of soluble proteins, but this is fraught with dangerous assumptions. Nevertheless, the authors have proceeded to speculate that the handful of essential proteins within the 50 could account for GroE essentiality. This presumes that basically all of the GroE-dependent proteins are now known. What if this is only half or less of the GroE-dependent proteins? The 1000 tested are likely to be the most abundant ones. Do we have any reason to expect the GroE-dependent proteins to be abundant? The authors should at least state the number of soluble proteins in E. coli so that we can compare with the number of tested proteins and then judge how well the GroE-dependent fraction may have been probed.

Our response:
The number of cytosolic proteins in E. coli has been predicted to be ~2,200 (Niwa et al., 2009), whereas our proteomics quantified 986 proteins. Although we tested ~70 additional proteins that could not be quantified by proteomics for the in vivo GroE requirement, about half of the E. coli proteins were not investigated for their GroE dependency. In addition, we newly identified several obligate GroE substrates that were not quantified by proteomics. Therefore, we agree with Referee #3 that the ~57 Class IV substrates we verified do not represent a complete set of GroE-dependent substrates. Regarding the comprehensiveness, our intention was to thoroughly investigate all of the Class III proteins identified by Kerner et al. for the in vivo GroE requirement for folding.
Since our description of the strategy might have been insufficient and confusing, we have modified the related parts for clarification. In addition, according to Referee #3's suggestion, we have described the number of cytosolic proteins in E. coli, and stated a caveat for the possibility of other GroE obligate substrates among the proteins that were not tested in our assay, in the Discussion.

Referee's comment:
The group of GroE-dependent proteins identified in this work is not thoroughly described. The authors do not even provide the number of proteins that satisfy the criteria for "candidate GroE obligate substrates." They spend a lot of time on the 43 GroEL-interacters previously identified by Hartl, but how were found by this new method?

Our response:
We acknowledge the confusion regarding our description of in vivo GroE substrates. By the criteria we used in our proteomics, 252 proteins were candidates for in vivo obligate GroE substrates among the detected 986 proteins (see below for details on the candidates). The 252 candidates contained many false positives, including proteins that were reduced for other reasons, including the cellular stress response mediated in GroE-depleted cells. Our criterion for an in vivo obligate GroE substrate is whether the candidate proteins expressed under GroE-depleted conditions are soluble (e.g. Fig. 2), and is not simply based on the reduction in the GroE-depleted cells.
Regarding the 43 GroE-interactors, it is true that we detected 43 Class III proteins in our proteomics. However, in the next section (Fig. 2, Supplemental Figs. S3, S4) we comprehensively examined all Class III proteins (83 proteins), including not only the 43 proteins but also 40 other proteins that had not been quantified by our proteomics in the individual expression assay. The verification of the GroE requirement revealed that 49 out of 83 Class III proteins were in vivo obligate GroE substrates

(Class III+).

Referee's comment:
The paper's title, "comprehensive survey", is not justified without addressing the above two issues and making a persuasive argument that they have achieved comphensiveness.

Our response:
Based on above argument, we have replaced the word "comprehensive" with "systematic" in the title of the paper. The new title is "A systematic survey of in vivo obligate chaperonin-dependent substrates".

Referee's comment:
The authors use poorly explained criteria to identify the GroE-dependents, those reduced in strain MGM100 by switching from ara to glc and those reduced in MGM100 in glc relative to MG1655 in glc.

Our response:
The cutoff value of 50% was set to minimize false negatives, as the highest solubility of the known in vivo obligate GroE substrates was 46%, for MetK. MetK was used for the cutoff, as described in Kerner et al. Since the solubility of MetK was 46% in our proteomics, the cutoff value was set to 50% of emPAIglucose/emPAIarabinose in MGM100 cells. In addition to this criterion, to minimize false positives we chose those reduced in strain MGM100 by switching from ara to glc, and those reduced in MGM100 in glc relative to MG1655 in glc. As a result, 347 proteins remained upon the cutoff of 50% in ara vs. glc in MGM100, and 329 proteins remained upon the cutoff of 50% in glc in MGM100 vs. glc in MG1655. Then, 252 proteins met both criteria. The above description has been included in the revised manuscript.

Referee's comment:
Part of Figure 1 indicates the fraction of proteins having reduced expression using the former, but it does not show the fraction reduced by the latter criterion. [What they show is the fraction reduced by ara to glc in MG1655, which is not a criterion and therefore is a bit misleading because it makes you think you've seen a description of both criteria.
They should describe (with numbers and a graph as in Fig. 1B) the similarity of the two protein sets captured by the criteria used and explain why these two criteria are preferable to replicates of ara vs. glc in MGM100.

Our response:
The numbers of proteins and the reasons why we used the criteria described above have been included in the revised manuscript. As Referee #3 pointed out, Figure 1B might be misleading. Therefore, we have included a color map showing MGM100 glc /MG1655 glc in the middle of Figure 1B.

Referee's comment:
On page 19, the authors should cite Aoki et al. (PMID 9405415) for showing that GroEL binding can be stabilized predominantly by electrostatics.

Our response:
Thank you for the suggestion. We have cited the work by Aoki et al. to show the electrostatic contribution in the GroEL-substrate interaction.

Referee's comment:
A number of typos on pp 20-21.

Our response:
We have carefully corrected the typos. In addition, the revised manuscript has been edited by a professional scientific editing service, as also suggested by Referee #2.

Referee's comment:
The authors should state that the TIM barrel is not simply the commonest fold (which would have been a trivial explanation for its dominance in GroE-dependence)

Our response:
We have removed the following sentence in the revised version. "Although the TIM barrel is the commonest fold among GroE substrates we propose that this is because it is the commonest aggregation-prone folds, rather than because of any property of the fold."

Referee's comment:
..., and they should identify the folds c.2 and c.37 in Figure 5.

Our response:
We apologize for forgetting to identify the two folds in the legend. c.2 is NAD(P)-binding Rossmann-fold domains, and c.37 is P-loop containing nucleoside triphosphate hydrolases. The notations have been described in the legend of Figure 5.

Referee's comment:
They should cite Lindquist for mutations and genetic diversity on p. 22.

Our response:
The two relevant papers (Rutherford and Lindquist, 1998; Queitsch et al., 2002) have been cited in the revised manuscript.

2nd Editorial Decision 26 February 2010

Many thanks for submitting the revised version of your manuscript EMBOJ-2009-73108R. It has now been seen again by referees 1 and 3. Both referees now fully support publication, but while referee 3 has no further comments, referee 1 does raise a number of issues on your revised manuscript that need to be addressed in a further round of minor revision. Most of these will require only text changes, but he/she does also point out that data on the solubility of some of the tested substrates is missing, and should be added to the appropriate figure.

I would therefore like to invite you to revise your manuscript according to the referee's comments. Once we have this revised version, I hope we should be able to accept your mansucript for publication without the need for further input from the reviewer in question.

Yours sincerely,

Editor
The EMBO Journal

———

REFEREE REPORTS:

Referee #1 (Remarks to the Author):

The authors have improved the manuscript greatly after the revision. All points I raised in my last review were addressed, and the manuscript has been modified accordingly. The revision was extensive and there now are new items that could be revised to improve the manuscript. Nonetheless, the revised manuscript requires only minor revision. The points are listed below:

Results section (p. 7): In the sentence "We note that a significant number of proteins were increased in the GroE-depleted cells..." the authors imply not the number of protein species, but their cellular abundances. I suggest re-phrasing this sentence, e.g. "We note that expression levels of a significant number of proteins were increased in the GroE-depleted cells..."

Results section (p. 7): While the first selection criterion (more than 50% solubility drop in glucose-treated MGM100 cells) has been justified by the authors by referring to the highest solubility of an obligate GroE substrate under the experimental conditions (46% for the MetK protein), the second selection criterion (more than 50% reduction in solubility in MGM100 cells compared to MG1655 cells in the presence of glucose) is still poorly justified. The authors do not clearly reason the choice of the 50% cut-off value. Will the results be different with a 40% cut-off value? 30%? 70%? This is of particular importance in the situation when a GroE substrate is an enzyme involved in E.coli sugar metabolism. They should discuss this issue properly.

Results section (pp. 9-10): In the text the authors claim that they tested GroE dependency of all 84 Class III substrates proposed by Kerner et al. (2005) except for Ypt1, a plasmid origin protein. They say that "...Class III was divided into 49 Class III+ and 34 Class III- substrates" and refer to the Figures 2C, Supplemental Figures S3 and S4. In the figures 2A, 2B, S3 and S4, however, the data are shown for all 34 Class III- proteins, but only for 46 Class III+ proteins. For instance, no solubility data can be found for the protein Nfo which is mentioned on page 11 and listed in the Table 1 as a Class III+ protein. As a result, on page 11 the authors refer to the data they do not actually show. The authors should show the solubility data for the three missing proteins.

Results section (p. 11): The authors say that "Dysfunction of an enzyme should cause an accumulation of its precursor and/or a reduction in the product of the catalyzed reaction". While this statement holds in vitro, in the cell enzymes often share precursors and products. Thus, a situation may occur when upregulation of some enzymes and downregulation of others will result in smaller than one would expect changes in the amounts of precursor and/or product. This point should be underlined.

Results section (p. 15): Neither in the Results nor in the Methods section the authors do not mention how they calculated hydrophobicity of the substrates. They should provide more details.

Discussion section (p. 20): Since most of the GroE substrates were not assayed for activity in this study, then, strictly speaking, it is incorrect to say that "...in vivo folding of ~40% (34 out of 84) of the Class III substrates was independent on GroE..." I suggest to refer to the solubility, which is supported by the experimental data, and not folding.

Discussion section (p. 20-22): On page 20 the authors say that "...in vivo GroE dependency is not determined by the enrichment of the GroE interaction" and later, on pages 21-22, discuss the role of Class III- substrates which strongly interact with GroE, but do not need it for folding. It is unclear though how much of the GroE capacity Class III+ vs. Class III- substrates occupy. It must depend at least on chaperone abundance, protein abundance and their binding affinity. This should be discussed in more detail.

Discussion section (p. 22): Please, change 83 to 84 in the sentence "Our comprehensive analysis revealed that 60% (49 out of 83) of the Class III substrates...". In the original paper by Kerner et al. (2005) there were 84 proposed Class III substrates.

Referee #3 (Remarks to the Author):

The authors have addressed the issues raised in the first review.

2nd Revision - authors' response 03 March 2010

Response to Referee #1

Referee's comment:

Results section (p. 7): In the sentence "We note that a significant number of proteins were increased in the GroE-depleted cells..." the authors imply not the number of protein species, but their cellular abundances. I suggest re-phrasing this sentence, e.g. "We note that expression levels of a significant number of proteins were increased in the GroE-depleted cells..."

Our response:
We have rephrased the sentence according to the Referee #1's suggestion (p.7).

Referee's comment:
Results section (p. 7): While the first selection criterion (more than 50% solubility drop in glucose-treated MGM100 cells) has been justified by the authors by referring to the highest solubility of an obligate GroE substrate under the experimental conditions (46% for the MetK protein), the second selection criterion (more than 50% reduction in solubility in MGM100 cells compared to MG1655 cells in the presence of glucose) is still poorly justified. The authors do not clearly reason the choice of the 50% cut-off value. Will the results be different with a 40% cut-off value? 30%? 70%? This is of particular importance in the situation when a GroE substrate is an enzyme involved in E.coli sugar metabolism. They should discuss this issue properly.

Our response:
The solubility of MetK was used for both criteria; the solubility of MetK in the second selection criterion is 46%, which is coincidentally same as that in the first selection criterion. Therefore, the choice of the cutoff value below 46% excludes MetK from the candidates. In contrast, higher cutoff values such as 70% greatly increase the number of false positives including Class I, II and III-proteins. To clarify this point, we added the description about the second selection criterion in the revised manuscript (p.7).

Referee's comment:
Results section (pp. 9-10): In the text the authors claim that they tested GroE dependency of all 84 Class III substrates proposed by Kerner et al. (2005) except for Ypt1, a plasmid origin protein. They say that "...Class III was divided into 49 Class III+ and 34 Class III- substrates" and refer to the Figures 2C, Supplemental Figures S3 and S4. In the figures 2A, 2B, S3 and S4, however, the data are shown for all 34 Class III- proteins, but only for 46 Class III+ proteins. For instance, no solubility data can be found for the protein Nfo which is mentioned on page 11 and listed in the Table 1 as a Class III+ protein. As a result, on page 11 the authors refer to the data they do not actually show. The authors should show the solubility data for the three missing proteins.

Our response:
Thank you for pointing out this issue. Although we showed 49 Class III+ proteins in the figures, we made a spelling mistake. "NanA" in previous Fig. S3 was Nfo. Npl, which is shown in just below Nfo (former "NanA"), is synonymous with NanA. (http://biocyc.org/ECOLI/NEW-IMAGE?type=ENZYME&object=ACNEULY-MONOMER). We have replaced Supplemental Figure S3 in this revision.
Regarding the number of Class III+ proteins, total number of Class III+ proteins is 49 (3 in Fig. 2A, 4 in Fig. 2B, 20 in Fig. S3, 12 in Fig. S4B, and 10 in Fig. S4C). To clarify this point, we described the details about the number in the revised manuscript (p10).

Referee's comment:
Results section (p. 11): The authors say that "Dysfunction of an enzyme should cause an accumulation of its precursor and/or a reduction in the product of the catalyzed reaction". While this statement holds in vitro, in the cell enzymes often share precursors and products. Thus, a situation may occur when upregulation of some enzymes and downregulation of others will result in smaller than one would expect changes in the amounts of precursor and/or product. This point should be underlined.

Our response:
We agree with Referee #1's comment. We have added the following sentence in the related Discussion (p 21). "We also note that a situation may occur when upregulation of some enzymes and downregulation of others will result in smaller than one would expect changes in the amounts of precursor and/or product."

Referee's comment:
Results section (p. 15): Neither in the Results nor in the Methods section the authors do not mention how they calculated hydrophobicity of the substrates. They should provide more details.

Our response:
We calculated hydrophobicity by the Kyte-Doolittle method (Kyte, J., and Doolittle, R.F. 1982 J Mol Biol 157, 105-132), as described in Supplemental methods. We cited the reference in the main text of the revised manuscript (p. 15).

Referee's comment:
Discussion section (p. 20): Since most of the GroE substrates were not assayed for activity in this study, then, strictly speaking, it is incorrect to say that "...in vivo folding of ~40% (34 out of 84) of the Class III substrates was independent on GroE..." I suggest to refer to the solubility, which is supported by the experimental data, and not folding.

Our response:
The word "folding" in the sentence has been replaced with "solubility" (p. 20).

Referee's comment:
Discussion section (p. 20-22): On page 20 the authors say that "...in vivo GroE dependency is not determined by the enrichment of the GroE interaction" and later, on pages 21-22, discuss the role of Class III- substrates which strongly interact with GroE, but do not need it for folding. It is unclear though how much of the GroE capacity Class III+ vs. Class III- substrates occupy. It must depend at least on chaperone abundance, protein abundance and their binding affinity. This should be discussed in more detail.

Our response:
We agree with Referee #1's comment on the GroE capacity. In addition to description that our proteomics did not provide the enrichment of the substrates in the GroE complex, we have discussed on this issue as Referee #1 suggested (p. 21).

Referee's comment:
Discussion section (p. 22): Please, change 83 to 84 in the sentence "Our comprehensive analysis revealed that 60% (49 out of 83) of the Class III substrates...". In the original paper by Kerner et al. (2005) there were 84 proposed Class III substrates.

Our response:
We have changed the number of the Class III from 83 to 84 in the sentence (p. 23).

Additional correspondence 04 March 2010

Many thanks for submitting your revised manuscript - from looking through your response to the referee's comments, I am happy that you have addressed his/her concerns. However, before we can accept your manuscript for publication, I notice one remaining problem. As stated in our Guide to Authors, we do not permit "Supplementary Discussion" sections: our view is that all discussion points pertinent to the conclusions of the study should be included in the main text. I therefore need to ask you to remove this section from your supplementary file (as well as the references to it in the main text), and - if appropriate - incorporate the points made here into the main discussion section.

In addition, we would like to encourage authors to include a statement as to author contributions in their acknowledgements section, and I would also like to ask you to include this.

I apologise for not pointing these issues out in my previous correspondence, and I suggest that the easiest way forward would be for you to e-mail me a revised manuscript text file and Supplementary Information file that we can upload in place of the previous versions. Once we have these, we should then be able to accept the manuscript without further delay.

Additional correspondence 04 March 2010

I modified the manuscript according to your advice. I've attached the main text and supplemental data in Word format. I look forward to hearing from you.

