A versatile selection system for folding competent proteins using genetic complementation in a eukaryotic host.
Lyngsø, Christina; Kjaerulff, Søren; Muller, Sven; Bratt, Tomas; Mortensen, Uffe H.; Dal Degan, Florence

Published in:
Protein Science

DOI:
10.1002/pro.337

Publication date:
2010

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

Citation for published version (APA):
A versatile selection system for folding competent proteins using genetic complementation in a eukaryotic host

Christina Lyngsø,¹* Søren Kjaerulf,² Sven Müller,² Tomas Bratt,² Uffe H. Mortensen,³ and Florence Dal Degan⁴

¹The Laboratory for Molecular Cardiology, The Danish National Research Foundation Centre for Cardiac Arrhythmia, Department of Biomedical Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen N, Denmark
²Pharmexa A/S, Agern Allé 1, DK-2970 Hoersholm, Denmark
³Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark
⁴Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen N, Denmark

Received 18 May 2009; Revised 14 December 2009; Accepted 17 December 2009
DOI: 10.1002/pro.337
Published online 15 January 2010 proteinscience.org

Abstract: Recombinant expression of native or modified eukaryotic proteins is pivotal for structural and functional studies and for industrial and pharmaceutical production of proteins. However, it is often impeded by the lack of proper folding. Here, we present a stringent and broadly applicable eukaryotic in vivo selection system for folded proteins. It is based on genetic complementation of the Schizosaccharomyces pombe growth marker gene invertase fused C-terminally to a protein library. The fusion proteins are directed to the secretion system, utilizing the ability of the eukaryotic protein quality-control systems to retain misfolded proteins in the ER and redirect them for cytosolic degradation, thereby only allowing folded proteins to reach the cell surface. Accordingly, the folding potential of the tested protein determines the ability of autotrophic colony growth. This system was successfully demonstrated using a complex insertion mutant library of TNF-α, from which different folding competent mutant proteins were uncovered.

Keywords: protein folding; selection; protein engineering; folding quality control; heterologous expression; eukaryote; in vivo selection; library construction; solubility; TNF-α; recombinant protein; Schizosaccharomyces pombe; yeast; fission yeast; protein vaccine

Introduction

The recombinant expression of properly folded proteins is essential for the biotechnological and pharmaceutical applications of proteins. However, because of incorrect folding, numerous native and modified proteins are not efficiently expressed in a soluble form and typically end up aggregating into inclusion bodies.

Additional Supporting Information may be found in the online version of this article.

Abbreviations: CAT, chloramphenicol acetyl-transferase; cDNA, coding DNA; CPY, carboxypeptidase Y; E. coli, Escherichia coli; ELISA, enzyme-linked immunosorbent assay; GFP, green fluorescent protein; PADRE, pan-DR helper T-cell epitope; PCR, polymerase chain reaction; RANK-L, receptor activator of nuclear factor kappa B ligand; S. pombe, Schizosaccharomyces pombe; Tat, twin-arginine translocation; TNF-α, tumor necrosis factors alpha; TRAIL, TNF-related apoptosisinduced ligand.

Søren Kjaerulf’s current address is ChemoMetec A/S, Gydevang 43, DK-3450 Allerød, Denmark.
Sven Müller’s current address is Dako A/S, Produktionsvej 42, DK-2600 Glostrup, Denmark.
Tomas Bratt’s current address is Novo Nordisk A/S, Novo Allé, Bagsvaerd, Denmark.
*Correspondence to: Christina Lyngsø, Department of Biomedical Sciences, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen N, Denmark. E-mail: lyngsoe@sund.ku.dk
To bypass this difficulty, different in vivo assays for detection of soluble recombinant proteins have been developed in E. coli. In these assays, the protein of interest is fused C-terminally to a reporter protein. Foldable protein can then be selected based on the comprehension that a misfolded insoluble protein, in most cases, causes elimination of the activity of a C-terminally fused reporter protein. Several reporter genes, for example, chloramphenicol acetyltransferase, green florescent protein (GFP), and β-galactosidase, have been used in this type of assay. Recently, a novel genetic selection assay for protein folding was described in E. coli, based on the observation that the bacterial twin-arginine translocation (Tat) pathway depends on correct folding of the protein prior to transport. Proteins of interest were fused C-terminally to the reporter β-lactamase and directed through the Tat pathway, through a signal peptide. Hereafter, the survival of E. coli cells on the selective medium was correlated with the solubility of the tested proteins. This assay represents a significant improvement to the previously described methods as they simply discriminate between soluble and insoluble proteins.

However, a major limitation of these systems is that E. coli lacks the elaborate protein quality control machinery of eukaryotic cells and their ability to perform important posttranslational modifications on proteins. Most proteins of interest in biotechnology are indeed of eukaryotic origin and are therefore best expressed in a eukaryotic host. Consequently, two screening systems have been developed in Saccharomyces cerevisiae. These assays are based on the comprehension that proteins are directed to the secretory pathway will be exported to the cell wall, only if they fold. In contrast, misfolded protein variants will be exported to the cytoplasm for proteosomal degradation. In one assay, the tested protein is fused to a small tag for antibody detection followed by fluorescence activated cell sorting. In another assay, a protein library is likewise combined with an epitope tag, and screening is performed on plates. One drawback of these assays is that they require extensive screening and additional assaying to get rid of false-positives. To eliminate these obstacles, we have developed a versatile genetic screening and selection system, based on the ER protein quality control in the fission yeast Schizosaccharomyces pombe.

Therapeutic recombinant human proteins will play an increasingly important role in combating human disease. Such proteins need to fold correctly to retain functionality. For example, vaccination against TNF-α, which is involved in chronic inflammation, could represent an attractive anti-TNF-α treatment for example, by the use of AutoVac technology. This technology depends on inserting a ubiquitous T-helper cell epitope into a native protein of interest. The presence of the T-helper cell epitope in the protein will increase its immunogenicity for example, to circumvent tolerance to harmful self-proteins. It is crucial that the epitope insertion does not disrupt the folding and three-dimensional structure of the protein to preserve native B-cell epitopes in the protein vaccine.

As proof of concept, we have applied our folding selection assay to identify foldable species within a library of human TNF-α insertion mutants, where an oligonucleotide encoding the 13 amino acid long T-helper cell epitope PADRE was inserted randomly into the TNF-α sequence using a transposon-based method. Using our selection system, we have successfully identified 20 folded variants in this library.

Results

Design of a eukaryotic protein-folding screening and selection system

The selection system presented here is designed to screen protein libraries to identify foldable species. The selection system relies on the ability of the secretory pathway of S. pombe to eliminate unfolded proteins through the ER quality control system. S. pombe has a well-organized and morphologically distinct golgi apparatus that performs many posttranslational modifications similar to those in higher eukaryotic cells. Importantly, it is an easily cultured single-cell eukaryote with a well developed genetic toolbox including the ability to introduce vector based gene libraries by transformation.

The S. pombe invertase gene was chosen as selectable reporter, for five main reasons. First, the invertase catalyses the hydrolysis of sucrose into glucose and fructose that, unlike sucrose, can be taken up by the cell and used for growth. This nutritional marker permits efficient selection on medium with sucrose, a sole carbon source, as only cells expressing functional invertase will survive. Second, invertase is a secretory protein and is therefore not expected to influence transport negatively through the secretory pathway. Third, the invertase activity produced from a surviving clone can be quantitated, and potentially, it provides a measure for folding efficiency. Fourth, invertase has already proven to be useful in screens, and fifth, it had proven tolerant to many different amino terminal extensions, hence, allowing fusion to numerous targets.

In our system, the cDNA encoding the proteins subjected to selection is fused in frame to the 5’-end of the invertase gene. This configuration was chosen for two reasons. First, the folding of the protein variant, rather than that of the invertase, is likely to determine the overall secretion efficiency of the
chimeric protein, as it reduces the possibility that the reporter protein may act as a primer for folding. Second, it ensures that no reporter protein activity is expressed when species in the library of protein variants contain a stop codon. To direct the different protein library variants to the ER, they are N-terminally fused to the signal peptide from carboxypeptidase Y (Cpy) from S. pombe, an export signal previously reported to direct TNF-α and other recombinant proteins to the secretory pathway of S. pombe (Fig. 1, Supporting Information Fig. A, panel B).

After transformation of an invertase deficient S. pombe strain (PY-17) with a plasmid-based library, folded proteins are subsequently selected in a simple growth-based assay on media containing sucrose as the main carbon source. Minute amounts of glucose (0.01, 0.02, or 0.03%) were added to the selective media to allow all newly transformed cells to survive on the sucrose media until the invertase gene begins to express from the plasmid. As glucose is readily used up by the plated cells, transformants, which are not expressing invertase, cannot propagate to form visible colonies. In this way, transformants expressing invertase can easily be selected from those that do not. The appropriate glucose concentrations were determined empirically through a series of tests on control strains [Fig. 2(B)].

The assay discriminates between foldable and nonfoldable proteins

To establish, whether the method indeed discriminates between folded and misfolded proteins, six test proteins were fused to invertase and transformed into PY-17. This set of proteins contains the three foldable species, GFP, human TNF-α, and human RANK-L, and a variant of human IL-5 that cannot fold and be secreted (IL5.14, generated in a previous study) and two randomly picked variants of TNF-α (variants 59 and 60) from our library (see later). Finally, PY-17 was also transformed with a construct, where a stop codon was inserted between the signal peptide and the invertase sequences.

We first verified that only invertase fusions with foldable proteins supported cell growth on selective medium (Fig. 2). PY-17 transformed with a construct containing invertase, TNF-invertase, RANKL-invertase, or GFP-invertase could grow on a nonselective medium and on a selective medium, whereas PY-17 transformed with an IL5.14-invertase construct or a construct, where a stop codon was inserted between the signal peptide and the invertase sequences, could only grow on a nonselective medium. Next, we investigated the presence of the invertase fusion proteins and their subcellular localization using invertase activity measurements (Table 1), immunocytochemistry, and Western blots (Fig. 3). Only strains transformed with invertase or invertase C-terminally fused with a foldable protein exhibited invertase activity. Invertase fusions with IL5.14 and TNF-α variants 59 and 60 did not show any activity. Invertase activity was solely associated with intact cells as no invertase activity could be detected in the culture medium. This indicates that the fusion proteins were correctly secreted but remained associated with the cell of S. pombe through the invertase moiety, because the invertase resides in the cell wall. This is important to keep cross-feeding of cells that do not express invertase activity by cells that do to a minimum. Western blot experiments of total cell extracts (including cell and cell wall content) and of culture medium demonstrated that the cellular localization of the hybrid proteins was determined by the invertase moiety. Hence, GFP and TNF-α that were not extended by invertase were secreted into the medium when directed through the secretory pathway with the use of the signal peptide of Cpy. In agreement with these results, we demonstrated that the GFP and TNF-α invertase fusions were localized to the cell surface by fluorescence microscopy [Fig. 3(B)]. In contrast, IL5.14, a protein species that does not fold efficiently, is retained within the cell when expressed alone or fused with invertase.

As expected, GFP-invertase, TNF-α-invertase, and RANKL-invertase transformants supported growth on selective medium and displayed significant levels of invertase activity, whereas the IL5.14-invertase transformant did not. This shows that the selection system is indeed able to differentiate between foldable and nonfoldable proteins. The cells transformed with invertase fusions with TNF-α variants 59 and 60 displayed no invertase activity and growth. This indicates that the assay can distinguish between folded and misfolded TNF-α variants in our library and that TNF-α variants 59 and 60 contain insertions at nonpermissive sites. In the following, variants 59 and 60 are used as negative controls for folding. Altogether, the growth on selective medium correlated with expression of the invertase fusion proteins at the cell surface and the invertase activity.

Construction of a TNF-α epitope insertion library

To test the efficiency of our selection assay, we used a TNF-α epitope insertion library as a model. Previously, a DNA transposition-based linker insertion strategy has been used to generate genes encoding proteins with random pentapeptide insertion. We have modified this strategy to insert a 45-mer oligonucleotide encoding the T-helper cell epitope PADRE (AKFVAAWTLKAAA) into the pentapeptide sequence (Supporting Information Fig. B), to produce a library of TNF-α insertion mutants. Our strategy results in 17 amino acid long inserts within the TNF-α sequence.
Figure 1. Summary of the library generation and selection strategy used to identify foldable proteins in *S. pombe*.

(A) Construction of the TNF-α insertion Library: (Step 1) The Entranceposon (Finnzymes) and transposase react with the acceptor plasmid (modified *E. coli* vector pUC19, containing the target cDNA (Supporting Information Fig. A, panel A), resulting in random integration of the transposon herein. TNF-α is represented in blue and the transposon in cyan-blue. The resulting plasmid collection is transformed into *E. coli* and amplified on LB solid medium containing both ampicillin and kanamycin, ensuring selection only of plasmids with integrated transposon. (Step 2) The transposon can be integrated both in the vector backbone and in the TNF-α gene. To eliminate the plasmids with transposons integrated into the vector backbone, the TNF-α + entranceposons sequences are reintroduced into the original pUC19 vector by use of the restriction enzymes Nco1 and Sac1. (Step 3) Transformants are again selected on LB solid medium containing ampicillin and kanamycin. This is important to eliminate any wild type TNF-α from the pool of variants that later could show up as false positives in the screen. (Step 4) Now, the entranceposon is ready to be replaced with the PADRE epitope. To do so, the entranceposon is removed by use of the restriction enzyme Not1. The open plasmid is ligated together with a DNA fragment encoding the PADRE epitope, having compatible ends. The resulting plasmid is amplified on LB solid medium containing ampicillin and kanamycin. The TNF-α library is ready for introduction into the screening vector. (B) Library generation and selection assay: The target cDNA (Red) is subjected to transposon (Cyan-blue) and transposase, and consequently the transposon is introduced randomly into the cDNA. In this study, the target cDNA encodes TNF-α. The transposon is genetically engineered. Hence, it can be removed by the restriction enzyme Not1. After digesting the library with Not1, the transposon is interchanged by an insert DNA fragment of choice (purple) containing compatible sticky ends. In this study, this insert contains the PADRE sequence [for details of the library construction, see Fig. 1(A) and materials and methods]. Next, the library of cDNA's is fused to the selectable genetic marker invertase (Blue) and the Cpy signal sequence (Green) to produce a target-reporter plasmid library. This library is introduced in an invertase null strain of *S. pombe* (PY-17). After translation, the protein enters the secretory pathway. Here, misfolded proteins are retained and eventually targeted for degradation by the proteasome, whereas folded proteins go to the surface. Here, the invertase marker supports growth on selective sucrose media by hydrolysing sucrose to the fermentable monosaccharides glucose and fructose. Accordingly, only colonies encoding folding competent protein variants will survive on selective sucrose media.
The constructed library contained approximately 17,500 individual species. If integration into the 474 bp TNF-α gene is truly random, this exceeds the number of different theoretically attainable integration sites by approximately 20 times. The randomness of sequence integration was determined for 44 arbitrarily picked colonies. Each clone contained a single insertion and when the integration points of all variants were compared they were distributed along the TNF-α sequence and did not appear to cluster. Also, no favorable sequence orientation was observed (Fig. 4(A) upper panel and Table II). Hence, this random distribution, combined with the relative size of the library, suggests that the library contains most of the possible protein variants.

**Identification of foldable TNF-α variants using the protein folding screening and selection assay**

The TNF-α library was transferred to the *S. pombe* expression vector, transformed into PY-17 and grown on plates on selective medium. For selection, transformants were plated on sucrose medium supplemented with the three different limiting concentrations of 0.01, 0.02, and 0.03% glucose, as described above. A total of ~ 36,000 transformants (12,000 on each selection medium) were subjected to selection (Table III).

The growth rate of the transformants was followed by visual inspection. As soon as the colonies reached a diameter of 1–2 mm (7, 8, or 9 days after transformation), they were picked from the selection plates and individually streaked onto fresh plates containing pure sucrose medium. This was done to reduce growth of satellite colonies that otherwise appear because of the cross feeding by glucose and fructose diffusing from the positive colonies (data not shown). Eighteen colonies were picked from plates containing 0.01% glucose and seven plates with 0.02% glucose.

On sucrose medium containing 0.03% glucose, 1068 colonies had formed after 7 days of growth. This was significantly more than the growth in the other two glucose concentrations, indicating that this glucose concentration is too high to provide the proper selective pressure needed to select for proteins that fold efficiently. Therefore, all colonies obtained on sucrose medium containing 0.03% glucose were replica-plated onto pure sucrose media. Following replica-plating, 11 colonies showed robust growth. These were picked after 9 to 10 days and restreaked as above. Thus, in total, 36 colonies were selected from the three different selective media. Finally, these colonies were restreaked twice on solid sucrose media to validate their phenotype. After this selection step, 26 transformants remained, which equals to 0.07% of all transformants screened. These clones were named clone 1–26.

After selection, the possibility existed that the growth of a colony was because of the isolation of a suppressor mutation in the *S. pombe* genome or resulted from the transformation by more than one plasmid into a single cell. Therefore, plasmids were rescued from all 26 colonies, retransformed into PY-17, and the transformants evaluated for growth ability on sucrose (data not shown). All the selected plasmids were able to complement invertase activity, except for four (clone 1, 20, 22, and 26). These four clones were discarded without further characterization. Hence, the final number of the selected clones with foldable TNF-α variants by the assay was 22.

**Sequence analysis of the selected TNF-α insertion mutants**

The plasmids from the 22 positive colonies were sequenced to identify the insertion point and the insert sequence within the individual TNF-α variants [Fig. 4(A), lower panel, and Supporting Information Fig. C]. Of these 22 clones, 20 were unique and two TNF-α variants were found twice (clones 7/24 and 9/17 were pair wise identical). Fourteen unique permissive insertion sites were identified, four of which could accommodate at least two different insert sequences.

To evaluate the effect of the screening procedure on the distribution of the epitope integration sites in TNF-α, the sequences of the 20 unique TNF-α variants were compared to those of the 44 randomly picked clones obtained from the library before selective screening [Fig. 4(A)]. Before selection, insertions were distributed evenly along the TNF-α sequence, regardless of the secondary structure of the protein. After screening, insertion sites were mapped to the regions of low structural complexity, as they clustered within the N-terminal region, C-terminus, or in loops. Only two inserts were found in a sequence encoding a β-strand (strand A and F, respectively) [Figs. 4(A,B)]. These results strongly indicate that the active selection for foldable protein has been taking place during the screen.

Importantly, when the collection of foldable TNF-α variants was inspected for the presence of the PADRE epitope, this sequence was found in six cases and at five different insertion sites (clone 4, 7/24, 12, 18, and 25). These proteins are therefore TNF-α vaccine candidates.

**The selected TNF-α variants ability to fold is not dependent on the invertase moiety**

The fact that the target protein is physically fused to invertase, our screen opens the possibility that invertase influences the folding efficiency and kinetics of the target protein moiety. We therefore expressed the five TNF-α vaccine candidates identified from our library in *S. pombe* without the invertase moiety to verify their folding properties. Wild-type TNF-α
Figure 2.

A

Non-selective media

[Images of petri dishes showing bacterial growth under non-selective conditions]

Selective media

[Images of petri dishes showing bacterial growth under selective conditions]

B

<table>
<thead>
<tr>
<th>2% Glucose</th>
<th>2% Sucrose</th>
<th>2% Sucrose</th>
<th>2% Sucrose</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01% Glucose</td>
<td>0.02% Glucose</td>
<td>0.03% Glucose</td>
<td>2% Sucrose</td>
</tr>
</tbody>
</table>

[Images of petri dishes showing bacterial growth under various glucose and sucrose concentrations]
was used as a positive control and the two clones 59 and 60, which did not support growth on selective medium (see above), were used as negative controls. Secretion of TNF-α was monitored using ELISA [Fig. 5(A)]. Only wild-type TNF-α and variant 25 were expressed at detectable levels in S. pombe. The absence of detectable levels of variants 4, 7, 12, and 18 in S. pombe indicates that the selection assay is more sensitive than the TNF-α ELISA test. Also, it is known that S. pombe secretes high levels of proteases in the growth media making exported folded proteins prone to partial degradation.21,25 Such degradation might affect only the detection in the ELISA assay and not the ability to promote growth of S. pombe colonies on sucrose. Therefore, the variants were tested for soluble expression in E. coli and Drosophila S2 cells, two commonly used hosts for protein recombinant expression, which we have previously used to express wild-type TNF-α (Nielsen, et al. and unpublished data). Wild-type and all five selected TNF-α variants, but not variants 59 and 60, were expressed as soluble proteins at detectable levels in E. coli and S2 cells [Fig. 5(A)]. The presence of TNF-α variant in S2 cell cultures was also analyzed by immunoblotting [Fig. 5(B)] and confirmed with the ELISA results. These expression data confirm that our selection system can distinguish foldable from nonfoldable proteins in a complex library of insertion mutants and that the invertase moiety does not drive or influence folding of the target passenger protein. In other hand, the expression levels of TNF-α variants were much lower in S. pombe than in the other two expression systems, indicating that our S. pombe strain, although being an excellent host for selection, is not well-suited for high yield production of recombinant proteins.

Discussion

TNF-α is a well-suited model to test our system, because the results from our screening method can be compared to many existing single amino acid residue substitutions, deletions, and in-frame insertions in TNF-α, which have previously been investigated for their effect on folding, structure, and activity.26–28 TNF-α is a homotrimer, where each monomer folds into a β-sandwich containing two stacked β-pleated sheets, each formed by five anti-parallel β-strands that adopt a classical jelly-roll topology with loops of variable length protruding out of the core.29,30 [Fig. (4B)].

Our selection assay identified 14 insertion permissive sites in TNF-α. More than half of the insertions were located within the first 26 N-terminal amino acid residues of TNF-α, which contain β-strand A. However, most of these sites were located in the sequence before strand A and in the loop between strands A and A’ (loop AA’). In the crystal structure of TNF-α, the first 10 amino acids residues are disordered, and mutations and deletions in this region generally do not affect the folding of the protein.29 In TRAIL, a member of the TNF protein family, loop AA’ contains a 15 residue-long extension, hence permissive insertion sites were to be expected in loop AA’ of TNF-α. Most of the remaining insertions were found in loops, where structural flexibility is expected. Interestingly, no insertion was found in the large flexible loop EF, where we have previously successfully inserted PADRE by rational design.26 This may reflect that our library screening was not exhaustive or that other loops are more permissive to insertions than loop EF. Indeed, we have previously shown that sliding of PADRE by one residue within loop EF can cause misfolding.26

Importantly, one insertion site (variant 11) would not have been predicted by rational design as it locates in the buried β-strand F in TNF-α. All other insertion sites are located near the surface of the protein [(Figs. 4(B,C)]. This demonstrates the power of this random genetic screen to uncover novel permissive insertion sites. We speculate that the buried PADRE insert integrates as part of the

---

**Figure 2.** Growth selection assay on plates to assess the expression of control proteins (A and B) and optimisation of the glucose concentration in the screening plates (B). (A) The invertase deficient S. pombe strain PY-17 was transformed with an expression vector containing unfused invertase (cpy-inv), GFP-invertase fusion (cpy-GFP-inv), IL5.14-invertase fusion (cpy-IL5.14-inv), TNF-α-invertase fusion (cpy-TNF-α-inv), or unfused invertaseQ176R (cpy-invQ176R), an invertase mutant with reduced activity (unpublished data). The transformants were streaked out on plates containing nonselective (EMM) or selective (EMM-sucrose) medium. (B) The invertase deficient S. pombe strain PY-17 was transformed with the same constructs as in (A) and with an expression vector containing invertase with a stop codon inserted between the signal peptide and the invertase coding region (cpy-stop-inv) or RANKL-invertase fusion (cpy-RANKL-inv). Transformants were directly plated out onto solid EMM-media containing 2% of glucose (nonselective medium) or 2% of sucrose supplemented with 0, 0.01, 0.02, or 0.03% of glucose (media of different selective pressure). An uneven distribution of colonies at the 0, 0.01, and 0.02% glucose concentrations were observed even though all transformed cells contain plasmids giving a positive phenotype. This could be due to a higher invertase activity in some colonies because of a higher plasmid copy number in these colonies resulting to a relatively faster growth rate. These cells then drive the neighbour colonies to grow faster because of cross-feeding. Alternatively, subtle differences in the environment on the plates, for example, temperatures or moistness result in the differences in growth rate.
β-strand causing a lateral displacement of the TNF-α residues. This extension is then accommodated in the adjacent EF or FG loop.

The data presented here demonstrate that we have developed a novel in vivo selection assay for protein folding, based on genetic complementation of the selectable growth marker gene invertase in S. pombe. After testing our assay with six proteins of known folding competence, we applied it to the screening of a TNF-α insertion mutant library and identified 20 unique folding competent TNF-α variants out of the 36,000 transformants screened. Recombinant expression of five of these TNF-α variants, without the invertase moiety, in two heterologous cell factories shows that our assay is indeed predictive of the folding of the target protein during subsequent recombinant expression in heterologous production hosts and provides a validation of the method. Our assay has a number of advantages over the existing ones. First, as it is a growth-based selection assay, only transformants expressing a foldable protein will form colonies. This greatly facilitates the identification of interesting candidates and the necessity of replica-plating and top-agar or color

<table>
<thead>
<tr>
<th>Construct</th>
<th>Total activity</th>
<th>Extracellular activity</th>
<th>Activity in culture medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cpy&lt;sup&gt;a&lt;/sup&gt;</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Cpy-inv</td>
<td>1287</td>
<td>1261</td>
<td>130</td>
</tr>
<tr>
<td>Cpy-TNF-inv</td>
<td>359</td>
<td>337</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Cpy-GFP-inv</td>
<td>122</td>
<td>120</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Cpy-RANKL-inv</td>
<td>494</td>
<td>485</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Cpy-ILA14-inv</td>
<td>19</td>
<td>16</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Cpy-TNF50-inv</td>
<td>50</td>
<td>n.d&lt;sup&gt;b&lt;/sup&gt;</td>
<td>n.d</td>
</tr>
<tr>
<td>Cpy-TNF60-inv</td>
<td>&lt;5</td>
<td>n.d</td>
<td>n.d</td>
</tr>
</tbody>
</table>

*S. pombe* cells were grown in EMM-fructose medium ON at 30°C. The number of cells in each sample was normalised to $5 \times 10^6$ cells/ml before measuring invertase activity using a quantitative liquid assay<sup>16</sup>. The invertase assay was performed on cell cultures (Total activity) as well as on intact cells (Extracellular activity) and culture growth medium, separately. Invertase activity is expressed in nanomoles of glucose released per minute per $5 \times 10^6$ cells at 30°C. The numbers are average values from two measurements.

<sup>a</sup> Cpy, signal peptide of carboxypeptidase Y.

<sup>b</sup> n.d., not determined.

Recombinant expression of five of these TNF-α variants, without the invertase moiety, in two heterologous cell factories shows that our assay is indeed predictive of the folding of the target protein during subsequent recombinant expression in heterologous production hosts and provides a validation of the method. Our assay has a number of advantages over the existing ones. First, as it is a growth-based selection assay, only transformants expressing a foldable protein will form colonies. This greatly facilitates the identification of interesting candidates and the necessity of replica-plating and top-agar or color.

---

**Figure 3.** Expression and sub-cellular localisation of control proteins fused to invertase. (A) Western blot analysis of *S. pombe* cell cultures transformed with control proteins GFP, TNF-α, or IL5.14 either without signal peptide or with the signal peptide of Cpy for extracellular expression (cpy-protein) or with the signal peptide of Cpy and fused to invertase (cpy-protein-inv). Fifty μg of crude protein extracts from the cell + cell wall fraction (C + W) or 10 μL of culture supernatant (S) or 10 μL of 10× concentrated culture supernatant (10S) were loaded in each lane. (B) Fluorescence microscopy of *S. pombe* cells expressing GFP-invertase fusion (cpy-GFP-inv) or TNF-α-invertase fusion (cpy-TNF-α-inv).
Figure 4. Distribution of insertion sites in the TNF-α structure. (A) Distribution of insertion sites along the TNF-α amino acid sequence before (upper panel) and after (lower panel) selection. Bars represent insertion sites, and their heights the number of insertions found at each site. The representation does not discriminate between the different reading frames and orientations. A two-dimensional representation of TNF-α indicating the positions of β-strands is shown below the diagrams. Blue rectangles represent β-strands situated in the inner-sheet of the molecule, whereas orange ones represent strands of the outer-sheet. (B) Graphic 3D representation of insertion points in TNF-α after selection. The TNF-α monomer is displayed as a solid ribbon with β-strands of the inner- and outer-sheets represented in blue and orange, respectively. The positions of insertion sites are located C-terminally to the amino acid residues represented in black in the accentuated subunit. For specific details of the individual insertions see supplementary Figure C. (C) Graphic 3D representation of the TNF-α trimer from two different perspectives, showing that insertion points are mostly located at the surface of the protein.
is reduced or avoided. Second, as it is performed in a eukaryotic host, our system exploits the elaborate eukaryotic protein quality control, which could be more stringent than a simple solubility screen in *E. coli*. Recently, Boder and coworkers showed that the protein quality control of a yeast (*Saccharomyces cerevisiae*) could not be distinguished between folding variants of a highly thermostable, nonnatural, de novo-designed protein.31 However, this seems to highlight a particularity of proteins of extreme thermostability or of nonnatural proteins, rather than a general limitation of the eukaryotic protein quality control.

Our assay is simple, fast, has a low frequency of false positives, and is well-suited for high-throughput screening of large protein libraries. The uncovered foldable sequences can then be expressed in a suitable expression host for example, insect cells as demonstrated here. Our approach is versatile, as no structural or functional information on the target protein is required apart from its sequence and many modifications including point mutations, insertions, or deletions may be analyzed, making our approach universal. In conclusion, the assay presented here is a true folding assay combining the simplicity of the *E. coli* on-plates systems with the advantages and versatility of a eukaryotic system.

**Materials and Methods**

All standard cloning procedures were carried out according to Sambrook and Russell (2001).32 Restriction enzymes and T4 ligase were purchased from New England Biolab and primers from DNA technology. PCR reactions were performed using Expand PCR system (Roche) according to the manufacturer’s instructions. The prepared plasmids were transformed into *E. coli* ElectroMAX™ DH10B™

### Table II. Distribution of Epitope Insertion Sites and Orientation

<table>
<thead>
<tr>
<th>Insert seq. frequency</th>
<th>RF of insertion point (33%)</th>
<th>Resulting insert DNA seq. (16%)</th>
<th>Resulting amino acid insert seq.</th>
<th>Translated seq prior selection (25%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>prior selection</td>
<td>pUC19 library</td>
<td><em>S. pombe</em> library</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>RF1-d (PADRE)</td>
<td>6 (33%)</td>
<td>10 (38%)</td>
<td>16 (36%)</td>
<td>6 (13%)</td>
</tr>
<tr>
<td>RF1-r</td>
<td>10 (22%)</td>
<td>16 (37%)</td>
<td>26 (33%)</td>
<td>10 (37%)</td>
</tr>
<tr>
<td>RF2-d</td>
<td>5 (28%)</td>
<td>8 (31%)</td>
<td>13 (30%)</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>RF2-r</td>
<td>9 (20%)</td>
<td>17 (40%)</td>
<td>26 (31%)</td>
<td>9 (20%)</td>
</tr>
<tr>
<td>RF3-d</td>
<td>7 (38%)</td>
<td>8 (31%)</td>
<td>15 (34%)</td>
<td>8 (20%)</td>
</tr>
<tr>
<td>RF3-r</td>
<td>7 (16%)</td>
<td>7 (16%)</td>
<td>14 (31%)</td>
<td>14 (31%)</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>26</td>
<td>44</td>
<td>44 (25%)</td>
</tr>
</tbody>
</table>

The insert can be found in three reading frames (RF1, RF2, RF3) and in two orientations (d: direct or r: reverse) translating into 6 different insert sequences, two of which contain stop codons (See supporting data C). Frequency of each possible insert sequence in TNF-α in 44 clones picked randomly in the library for sequence analysis before selection (Panel A) and in 22 clones isolated from the library after selection (Panel B).

### Table III. Number of Colonies Obtained After Each Screening Step Using Three Different Selection Pressures

<table>
<thead>
<tr>
<th>Glucose concentration</th>
<th>0.01%</th>
<th>0.02%</th>
<th>0.03%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of screened colonies</td>
<td>12,000</td>
<td>12,000</td>
<td>12,000</td>
<td>36,000</td>
</tr>
<tr>
<td>Positive colonies before restreak</td>
<td>18</td>
<td>7</td>
<td>1,068 (11)</td>
<td>36 (0.1%)</td>
</tr>
<tr>
<td>Positive colonies after restreak</td>
<td>12</td>
<td>5</td>
<td>9</td>
<td>26 (0.07%)</td>
</tr>
</tbody>
</table>

Colonies obtained on 0.03% glucose medium were subjected to an additional round of selection on sucrose medium; the resulting number of colonies is given in parenthesis.
(Invitrogen) by electroporation using a Gene Pulser II (Bio-Rad) and grown in liquid or agar solid LB-media (Invitrogen) containing 50 μg/mL carbenicillin (Sigma). All plasmids are sequenced in a region covering modifications to check the fidelity of the constructs, using the BigDye kit (Applied Biosystem) on an ABI PRISM 310 Sequence analyzer (Applied Biosystem).

Construction of the S. pombe invertase knockout strain
The S. pombe invertase knockout strain PY-17 (mat1-P Δmat2,3::LEU2− leu1 Δinv1::ura4+) was generated from Eg660 obtained from O. Nielsen (University of Copenhagen). The inv1 locus was disrupted in the Eg660 wild type S. pombe strain (kindly provided by O. Nielsen, University of Copenhagen) by replacing the complete open reading frame of inv1 with the S. pombe ura4 gene to produce the strain, mat1-P Δmat2,3::LEU2− leu1 Δinv1 (PY-17). Approximately, 1000 bp of sequences upstream of the initiating ATG of inv1 was amplified by PCR using S. pombe genomic DNA as template and the following oligonucleotides: 5′-ATCTTAGACA AAATACATCTTATGGAATCGTTAAG-3' and 5′-AGAATTCGAAATGCCACTAGCTAATATATT TC-3'. Around 1000 bp of sequences located downstream of the inv1 stop codon was amplified by PCR using S. pombe genomic DNA as template and the following oligonucleotides: 5′-AATCGATGCCATTAA AGACTATCTGGAATTGC-3' and 5′-AGGTACCTTGGC AAAAATACA TATTTGTGG-3'. The two PCR products were linked to the S. pombe ura4 gene through EcoRI and ClaI sites and the chimeric molecule was inserted into pBluescriptIIISK(−) using KpnI and XbaI sites. A linearized DNA fragment carrying the disrupted inv1 gene was used to transform a wild-type haploid strain. To confirm the disruption of the inv1 gene, ura + transformants were analyzed by PCR to verify correct integration of construct.

Construction of the screening vector
A 1680 bp fragment encoding amino acids 23–581 of inv1 was amplified by PCR using S. pombe genomic DNA as template and the following oligonucleotides: 5′-ACCATGGTCCCGCCTCA CTATATGTAA-3′ and 5′-AGCGGCCGCTTAAGCAATTCAGA TAGTCCTTAA ATG-3′. The resulting PCR product was digested with NeoI and NotI and inserted into Neoc/NotI restricted pCpy-GFP³, thereby fusing the inv1 gene with the Cpy secretion signal sequence (pNMT-Cpy-inv1). Next, the screening vector (pCpy-inv1-lib) was generated for in-frame insertion of protein encoding libraries between the Cpy signal peptide and the inv1 gene. First, the Sac1 site found in the downstream of the nmt terminator in the pNMT-Cpy-inv1 vector was destroyed by Sac1 restriction followed by end-filling with T4 DNA polymerase and religation. Second, a stop codon and a Sac1 site were inserted immediately after the Neo1 site located in the fusion point of the Cpy signal sequence and inv1 gene as described in the following. The 5′ of the inv1 gene was amplified by PCR using the oligonucleotides: 5′-TATCCATGGCAATGAGCTCCTCCCGCTACATTATGG TAAAAGC-3′ and 5′-CAATGGTATAGCAATTCAGC CTG-3′. The PCR product was restricted with NeoI and BspEI and ligated into Neo1/BspEI digested pNMT-Cpy-inv1, creating the final screening vector. For cloning of the library into the expression cassette the recognition sequences for the restriction enzymes NeoI and Sac1 are situated between Cpy and invertase. In between these restriction sites a stop codon is placed in-frame of the invertase gene to prevent background from uncut or religated plasmids in the screen. The library of variants was cloned from the pUC19-based library plasmid by excising the sequences by NeoI and Sac1 restriction and ligating into Neoc/Sac1 digested screening plasmid.

Construction of TNF-α variant library
For insertion mutagenesis, the MuA transposon kit from Finnzymes was used according to the
manufacturer’s directions. The template DNA was made by PCR amplification of a human wt-TNF-α synthetic gene made by GeneArt GMBH, using the primers 5'-GCACCATA TGCCATGGTGCCCTCAAGCTGGCCACGGCAG-3' and 5'-TTACGC-CAAGCTTGAGCTCCCAATGCGATAATGCCGAAGTAGACC-3'. Fragments were ligated into NdeI and Hind III digested pUC19 plasmid (Invitrogen).

The TNF-α coding expression cassette containing transposon inserts was removed from the vector using NdeI/SacI double digestion and reintroduced into a fresh pUC19 plasmid to remove plasmid containing transposon insertions outside of TNF-α. The transposon was then removed by the restriction enzyme NotI followed by de-phosphorylation, and interchanged with insert fragment encoding PADRE flanked by NotI restriction sites. For details see Figure 1(A) and Supporting Information Figure (B). As the insert integrates in all the three reading frames and in both orientations, it results in six different insert sequences. One of these insertions equals the epitope, three others result in different ORF peptides, and two introduce a stop codon in the TNF-α cDNA sequence. Hence, the library comprises cDNA species encoding TNF-α variants with inserts of four different amino acid sequences and nonproductive species (variants with stop codons).

To make the PADRE insert the two primers 5’-GC GCCCGCTAAGTTGCGACCTGAACCTAAGG CGG-3’ and 5’-CCGGTTCATCATTACTATTACTAAGCTCAGGG-C3’ were annealed and cloned into NotI-digested pVAX vector (Invitrogen). The NotI restriction site adds two additional alanines to the insert sequence. This plasmid was amplified in E. coli and the PADRE insert was recovered by NotI digestion.

**Immunocytochemistry**

About 2 × 10^6 cells were labeled with antibodies as described elsewhere. For labeling of TNF-α, monoclonal antibody Hyb 256-01 was used (1 μg in 100 μL of 50 mM Tris-Cl, 150 mM NaCl, 0.1% BSA, pH 8.0), followed by incubation with rabbit anti-mouse Alexa 488 (Molecular Probes, 1 μg in 100 μL of same buffer as above). The microscopic analysis was performed using an Axioplan 2 (Carl Zeiss) microscope equipped with a MicroMax (Roper Scientific) cooled CCD camera and MetaMorph software (Universal Imaging).

**S. pombe screening**

*S. pombe* transformations were performed using a PEG/lithium acetate procedure. Transformants were washed twice in media without glucose before plating. *S. pombe* cells were grown at 30°C. EMM media with the different carbon sources were prepared as described elsewhere. Selection media were always freshly made with filter-sterilized ultra pure sucrose (Invitrogen, Cat. #15503-022). The number of transformants was estimated to be ~ 4,000 on each 10 cm dish based on the number of colonies appearing on the nonselective 2% glucose media.

**S. pombe plasmid rescue, colony PCR, and invertase assay**

A single colony was inoculated in 5 mL EMM-sucrose media and grown at 30°C until cell density had reached 1 × 10^7 cells/mL. For invertase assay, cells were diluted exactly 5 × 10^6 cells/mL, and then the assay was preformed as described elsewhere. For colony PCR cells from 1 mL culture were mixed with 500 μL extraction buffer and 250 μL glass beads. Cell disruption was then performed mechanically by Fastprep for 15 s. After 1 min centrifugation at 13,000 rpm, 1 μL supernatant was used for a PCR with the primers 5’-CGTTTTCATATAAGTG ACCGGGAGTCAGGGG-C3’ and 5’-CCGGGTATCAT CATTACTATTACTAAGCTCAGGG-C3’. For plasmid rescue cells were harvested and washed once in 5 mL milliq H2O. 500 μL extraction buffer (100 mM Tris pH 7.6, 10 mM EDTA, 1% SDS, 0.2M NaCl) and glass beads until 2 mm below the liquid surface were added. Plasmids were extracted using phenol/ chloroform extraction. Briefly, after addition of 500 μL phenol/ chloroform/isoamylalcohol (25:24:1), cells were mechanically disrupted for 15 s in Fastprep. The phenol phase was removed after 5 min centrifugation at 20,000 rpm. A second phenol extraction was performed and the DNA extracted by EtOH precipitation. The plasmid preparation was transformed into DH10B (Invitrogen) and plasmids were sequenced.

**Recombinant expression of TNF-α variants in S. pombe**

The expression cassettes of the TNF-α variants were amplified by PCR using the primers 5’-CG TTATACATATAAGTG ACCGGGAGTCAGGGG-C3’ and 5’-CCGGGTATCAT CATTACTATTACTAAGCTCAGGG-CTC3’ and cut with the restriction enzymes NcoI and HindIII and cloned into the screening vector also line-argised with NcoI and HindIII, hereby removing the invertase moiety. The plasmid was transformed into PY-17. A single *S. pombe* colony was grown ON in a 3 mL preculture EMM-glucose medium containing 2% broad bean peptone and 15 μM thiamine to suppress the promoter. The ON cultures were used to inoculate fresh expression medium at OD_600 = 0.2. The expression medium was 10 mL EMM-glucose medium containing 5% broad bean peptone (Fluka) and CS-LEU amino acid supplement (Q-BIO gene), but no thiamine. Culture supernatants were harvested by centrifugation after 30 h of growth. Growth media were used for protein quantification and expression level was normalised relative to the
Recombinant expression of TNF-α variants in E. coli

The expression cassettes of the TNF-α variants were amplified by PCR using 5′-CGTTTTACATAAGTGACGGGAGCTCAGAGG GC-3′ and 5′-CCGGTCATATTACTACTATTACTACAGGCT-3′ and cloned into the pET28b vector by standard techniques using the restriction enzymes Neol and HindIII. E. coli BL21-star, containing the various expression constructs, were grown to mid-log phase (OD₆₀₀ = 0.4) in 50 mL LB media with 60 μg/mL kanamycin (Sigma). Then the expression was induced with addition of 0.7 mM IPTG (Boehringer Mannheim) and cells were further cultivated for 5 h. 1 mL of cells was harvested, pelleted, and resuspended in 1 mL of sonication buffer [20 mM Bis-Tris pH 6.0,40 mM NaCl, 10 mM EDTA, containing Complete<sup>TM</sup> protease inhibitor (Roche)]. Cells were sonicated on ice 4 times for 15 s, with 20 s pauses, and centrifuged at 20,000 rpm at 4°C for 20 minutes. Cell-lysis supernatants were used for protein quantification.

Recombinant expression of TNF-α variants in Drosophila S2 cells

The TNF-α and variants cDNA were amplified from the screening vector by the primers 5′-TCAGCTGAA TTCCTGATCTCATAATAGGAATTTACTCATATTACTG GCCCTGCTGGCCTTGGCTCTGGCCTCTGGGATGG TTCCCGATCTCAATATGAAGTTATGCATATTACTG and 5′-CGGTTCAGAGGATCTCATATTACTATTAGAGGGGCATGG TCGAGATCATCTCTCAAGGCCGCAACCG-3′ and 5′-CCGGTCATATTACGAATTTAGGAATTTACTCATATTACTG GCCCTGCTGGCCTTGGCTCTGGCCTCTGGGATGG TTCCCGATCTCAATATGAAGTTATGCATATTACTG and 5′-CGGTTCAGAGGATCTCATATTACTATTAGAGGGGCATGG TCGAGATCATCTCTCAAGGCCGCAACCG-3′ and 5′-CCGGTCATATTACGAATTTAGGAATTTACTCATATTACTG GCCCTGCTGGCCTTGGCTCTGGCCTCTGGGATGG TTCCCGATCTCAATATGAAGTTATGCATATTACTG. The 5′ includes the signal sequence of Bip, for directing the protein through the secretory pathway. The fragment was then introduced into p2Zop2F vector linearised with the restriction enzymes NeoI and HindIII. The fragment was amplified by PCR using 5′-CGGTTCAGAGGATCTCATATTACTATTAGAGGGGCATGG TCGAGATCATCTCTCAAGGCCGCAACCG-3′ and 5′-CCGGTCATATTACGAATTTAGGAATTTACTCATATTACTG GCCCTGCTGGCCTTGGCTCTGGCCTCTGGGATGG TTCCCGATCTCAATATGAAGTTATGCATATTACTG and cloned into the expression host and plasmid transfection was done using the Saint-18 transfection reagent (Cytotech) according to the manufacturer’s instructions. Cells were grown in 4 mL Excell-420 media (JRH Biosciences) with 10% FBS in 25 cm² T-flasks at 37°C for 4 days.

Quantification of TNF-α variants using ELISA

The assay was performed as described previously<sup>26</sup> with the following modifications. Maxisorp microtiter plates were coated with 100 μL of antimouse antibody (Dako Z0109), diluted 1:1000 in carbonate 0.1M, and pH 9.0. After blocking (1% BSA in PBS, pH 7.2) 100 μL of MAB210 (R&D Biosystems) diluted to 0.5 μg/mL in ELISA buffer (1% BSA, 0.05% Tween 20 in, PBS, pH 7.2) was transferred to the plates.

Western blots

Following electrophoresis on NuPAGE 4–12% SDS gels (Invitrogen), protein samples were transferred onto nitrocellulose membranes by semidyblotting. For immunoreaction incubation with rabbit TNF-α antisemur (1 h; dilution 1:10,000) or rabbit IL-5 antisemur or rabbit GFP antisemur (Abcam 6556) and with horse radish peroxidase-conjugated secondary antibody (Dako P448) was carried out in Tris 50 mM, NaCl 150 mM, EDTA 5 mM, Igepal 0.1%, and gelatin 0.5%. Antibodies were diluted 1:10,000 and incubations were performed at room temperature for 1 h. Detection was with ECL<sup>TM</sup> reagents (GE Healthcare).

Acknowledgment

The S. pombe strain Eg660 was obtained from O. Nielsen (University of Copenhagen). This work was funded by the Danish ministry of science, technology and innovation and Pharmexa A/S.

References