Applications of catalyzed cytoplasmic disulfide bond formation

Mirva J Saaranen and Lloyd W Ruddock*

Faculty of Biochemistry and Molecular Medicine and Biocenter Oulu, University of Oulu, Aapistie 7, 90220 Oulu, Finland

* Corresponding author. Email: lloyd.ruddock@oulu.fi

Abstract

Disulfide bond formation is an essential post translational modification required for many proteins to attain their native, functional structure. The formation of disulfide bonds, otherwise known as oxidative protein folding, occurs in the endoplasmic reticulum and mitochondrial inter-membrane space in eukaryotes and the periplasm of prokaryotes. While there are differences in the molecular mechanisms of oxidative folding in different compartments, it can essentially be broken down into two steps, disulfide formation and disulfide isomerization. For both steps, catalysts exist in all compartments where native disulfide bond formation occurs. Due to the importance of disulfide bonds for a plethora of proteins, considerable effort has been made to generate cell factories which can make them more efficiently and cheaper. Recently synthetic biology has been used to transfer catalysts of native disulfide bond formation into the cytoplasm of prokaryotes such as *Escherichia coli*. While these engineered systems cannot yet rival natural systems in the range and complexity of disulfide bonded proteins that can be made, a growing range of proteins have been made successfully and yields of homogenously folded eukaryotic proteins exceeding g/L yields have been obtained. This review will briefly give an overview of such systems, the uses reported to date and areas of future potential development, including combining with engineered systems for cytoplasmic glycosylation.

Abbreviations:

ER, endoplasmic reticulum; GRAS, generally regarded as safe; IMS, inter-membrane space; PDI, protein disulfide isomerase; POI, protein of interest; PTM, post-translational modification

Introduction

Disulfide bond formation is one of the most common post-translational modifications (PTM) found in proteins. Such oxidative folding as it is also known, occurs mainly in specialized compartments such as the endoplasmic reticulum (ER), intermembrane space (IMS) of mitochondria, and bacterial periplasm. The primary function of disulfide bonds is to stabilize the native structure of proteins, in particular in secreted or outer membrane proteins. Native disulfide bond formation is the rate limiting step of protein folding for many proteins and hence proteins that contain disulfide bonds may be difficult to obtain in high yields. This can severely inhibit progress in understanding a myriad of physiological processes and associated disease states, and can impose major limitations on the effective and economic expansion of the use of such proteins – for example in protein based therapeutics and analytics.

The mechanisms of disulfide bond formation are complex and vary between the different compartments in which native disulfide bond formation occurs (reviewed in [1]). At its simplest native disulfide bond formation can be broken down into two steps, dithiol oxidation and disulfide isomerization (Fig 1A). Each of these steps is catalyzed *in vivo*. In the ER of eukaryotes, the first step is catalyzed by a sulfhydryl oxidase, while the second step is catalyzed by a protein disulfide isomerase (PDI). Humans have three families of sulfhydryl oxidases, the Ero1, ERV/ALR and QSOX families, while the human ER contains over 20 PDI family members.

Disulfide bond formation in the cytoplasm is found naturally in some hyperthermophiles, under stress conditions and during some viral infections (reviewed in [2]). To our knowledge there have been no published applications of such natural systems. Engineered systems for cytoplasmic disulfide bond formation have primarily been focused on *E.coli*. *E.coli* is a well characterized, generally regarded as a safe (GRAS) organism for the production of various pharmaceutical and industrial enzymes and proteins, with fast and cheap scalability, making it ideal for biotechnological applications.

The first engineered systems for disulfide bond formation in the cytoplasm were based on the disruption of the natural reducing pathways present. Two such pathways exist (Fig 1B) and disruption of both, for example in *Δgor/ΔtrxB* strains, allows the production of some disulfide bonded proteins [3]. Subsequently a disulfide isomerase, DsbC, was added to this system [4] and this variant is commercially available as the SHuffle system from New England Biolabs. In this system thioredoxin “reverses” its usual function and transfers disulfide bonds into folding proteins

[5]. However, there is no active system for the formation of disulfide bond *de novo* i.e. there is no equivalent to a sulfhydryl oxidase. Instead disulfide bonds are introduced into thioredoxin based on its ability to reduce disulfide bonds – in effect thioredoxin acts to transfer disulfide bonds between proteins (Fig 1C). It is possible that the main route for disulfide bond formation in *Δgor/ΔtrxB* strains is via the action of ribonucleotide reductase, which generates one disulfide bond per catalytic cycle (reviewed in [6]) i.e. disulfide bond formation in folding proteins such strains may be directly linked to DNA biogenesis.

Subsequent systems introduced active catalysts for both of the steps in native disulfide bond formation i.e. a sulfhydryl oxidase and a disulfide isomerase [7,8]. There is no need to disrupt the reducing pathways (Fig 1D; [7]) and the system can be plasmid based and hence easily transferred between different *E.coli* strains. It works in all genetic backgrounds and all media tested to date, though appropriate consideration for the target protein of interest (POI) must be given and combining active catalysts with *Δgor/ΔtrxB* strains can be problematic as the cytoplasm becomes over-oxidizing and disulfide isomerases need to be in the reduced state to be active..

This review will focus only on systems which have active catalysts for both steps.

Initial publications [7,8] used an odd combination of protein folding catalysts, a sulfhydryl oxidase (Erv1p) from the intermembrane space of yeast mitochondria and the periplasmic *E.coli* catalyst of disulfide isomerization DsbC. This arose in part by chance, but also by design. Many sulfhydryl oxidases e.g. Ero1 family members contain structural disulfide bonds and so these cannot be easily used as the primary oxidase as they set up a chicken and egg situation where disulfide bonds are needed to generate a system which can make disulfide bonds. Erv1p does not have this issue. In addition, Ero1 family members do not act on folding proteins directly but instead oxidize PDI [9], while in house *in vitro* experiments indicated that Erv1p was able to act efficiently on folding proteins (unpublished data). While initial reports used DsbC, this was subsequently swapped for Human PDI as - with the exception of certain specialist proteins such as marine snail conotoxins [10], we are unaware of any disulfide bond containing protein tested *in vitro* which PDI is not able to help fold i.e. it appears to be the most efficient disulfide isomerase reported to date.

Subsequent publications sometimes replaced the oxidase (Erv1p) with either another sulfhydryl oxidase (e.g. QSox [11, 12, 13] or inverted the *E.coli* periplasmic transmembrane disulfide bond forming enzyme DsbB [14] such that its active site pointed to the cytoplasm or used a transmembrane disulfide bond forming enzyme VKOR from a hyperthermophilic organism that

naturally makes disulfide bonds in the cytoplasm [14]. However, Erv1p remains the most widely used oxidase in engineered systems for cytoplasmic disulfide bond formation.

Proteins made using cytoplasmic disulfide bond formation

The initial publication [7] on catalyzed cytoplasmic disulfide bond formation used two proteins of interest (POI), alkaline phosphatase (PhoA) and acid phosphatase (AppA), usually expressed in the periplasm of E.*coli*. These were chosen as they have simple disulfide patterns, have easily analyzed biological activity and had been previously used for examining systems for the production of disulfide bonded proteins e.g. [3]. PhoA is made in a soluble, but biologically inactive state in the absence of systems for disulfide bond formation. Addition of just a sulfhydryl oxidase generated high yields of active protein. This is the only protein we are aware of that contains more than one disulfide bond which requires no disulfide isomerase for efficient native disulfide production. AppA needed catalysts for both steps of native disulfide bond formation to be produced in an active state.

The second publication [8] greatly expanded on the number of protein types examined and looked at the expression of eukaryotic proteins. These included vtPA (9 disulfide bonds), BPTI and mutants (2 or 3 disulfide bonds), Ero1 (6 or 7 disulfide bonds), BMP (disulfide linked heterodimer, with each monomer containing 3 disulfide bonds), resistin (disulfide linked heterodimer, with each monomer containing 5 disulfide bonds). All were made in increased yields using a catalyzed disulfide forming system. The yields of several proteins were helped by the use of a fusion protein, MBP.

The use of fusion proteins can have three separate effects. Firstly, some, including MBP, can be used for purification of the protein. Secondly, the fusion protein may increase the solubility of the final product. Thirdly, and most importantly connected with disulfide bond containing proteins, the fusion protein may increase the solubility of folding intermediates. This can be a major issue with cytoplasmic disulfide bond formation due to differences in how disulfide bonds are made in this system compared with how they are made in the ER, IMS and periplasm. In natural systems the protein is translocated across a membrane in an unfolded state and disulfide bonds are made co-translocationally e.g. by the action of Ero1/PDI [15]. Most likely sequential disulfide bonds are formed first and these are subsequently isomerized to the native state. Proteins have probably evolved such that folding intermediates formed by this route are less prone to aggregation, though there is no evidence we are aware of for this. In catalyzed cytoplasmic disulfide bond formation there is no translocation across a membrane, hence no co-translocational disulfide bond formation

and hence the folding pathway may be different. It is impossible to predict which proteins will have issues with the solubility of folding intermediates and hence may benefit from the use of a fusion protein approach. However, we have used the system to produce >500 proteins and two rules of thumb have arisen. The first is that for POI with N-glycans in their natural state, a N-glycan density of > 1 per 100 amino acids generally implies that the POI cannot be made without the use of a solubilizing fusion partner. The second is that increasing density of disulfide bonds, e.g. >4 disulfide bonds per 100 amino acids, often causes solubility of folding intermediate issues. The use of fusion proteins has two downsides. Firstly, often the fusion protein must be removed at the end – usually via the action of proteases – adding to downstream processing and the (in)accessibility of the cleavage site may cause inefficient release of the mature POI. Secondly, the use of fusion proteins may generate false positive results i.e. they may allow the solubilization of incompletely folded or incorrectly folded proteins (Fig 2). MBP is an excellent choice for solubilization of folding intermediates, but it is prone to false positive results. Other proteins have also been used as fusion partners in this system, including Halo, SUMO and ubiquitin [16, 17], but there has to our knowledge been no systematic comparison of their efficiency and their propensity to generate false positive results.

There is now a rapidly growing body of literature using catalyzed cytoplasmic disulfide bond formation for the production of a wide range of proteins (Table 1). These proteins range in size from a 6 kDa monomer to a 107 kDa heterodimer and contain between 1 and 11 disulfide bonds. The highest published yield we are aware of for production is 240mg/L from shake flasks and 1.1g/L purified product from fed-batch fermentation in chemically defined media [18, 19]. Unpublished results both increase yields obtained and greatly widen the breadth of proteins made from 5 to 135 kDa in size for monomeric proteins and with up to 44 disulfide bonds – though typically proteins with more than 10 disulfide bonds need extensive optimization of the system. Proteins produced in the system have been successfully used for protein crystallization e.g. [20] demonstrating homogeneity of the material produced.

Other post-translational modifications in the cytoplasm of *E.coli*

Disulfides are often found in proteins with other PTM, some of which are essential for the function and/or stability of the protein. The most common PTM linked with disulfides is N-glycosylation of the side chain of asparagine residues. N-glycosylation of POI has been reported in both the periplasm [21, 22] and cytoplasm [23] of *E.coli* using engineered cell factories. Neither system uses eukaryotic N-glycosylation machinery as the oligosaccharide transfer machinery is a large complex

of multiple transmembrane proteins (reviewed in [24]). Instead bacterial N-glycosylation systems are used. In the periplasm the pgl system from *Cambylobacter jejuni* is used, while in the cytoplasm N-glycosyltransferase (NGT) from *Actinobacillus pleuropneumoniae* is used. To our knowledge nobody has reported combining catalyzed disulfide bond formation and N-glycosylation in the cytoplasm of *E.coli*. This possibly arises as NGT is inefficient at glycosylating heterologous POI.

The formation of other PTMs in the cytoplasm of *E.coli* are potentially possible, including where the enzymes forming the PTM themselves contain disulfide bonds (Fig 3). A recently published example of this is mucin-like O-glycosylation [25]. Initiation of mucin-like O-glycosylation is catalyzed by GalNAc transferases, with 20 family members reported in humans [reviewed in 26]. All are Golgi located transmembrane proteins and all contain structure stabilizing disulfide bonds. Several Golgi located glycosyltransferases have been expressed in the cytoplasm of *E.coli* using active mechanisms of disulfide bond formation to facilitate functional expression. These include B4GalT1 [27], ST3Gal1 [28], ST6Gal1 [28] and, of relevance to mucin-like O-glycosylation, GalNacT2 [25, 29, 30].

Given the relative simplicity of the evaluating the use of the system, it is likely that other PTMs combined with disulfide bond formation in the cytoplasm will be reported over the next few years.

Completing the loop

Catalyzed systems for disulfide bond formation in the cytoplasm arose from the application of knowledge about how natural systems catalyze such processes, but what about completing the loop? What we can learn about the mechanisms of protein folding using cytoplasmic disulfide bond formation?

To date there have been relatively few reports on this, but since many proteins involved in protein folding contain disulfide bonds themselves, there is considerable potential. In addition, the ability to make late stage folding intermediate mimics e.g. BPTI lacking one disulfide bond [8] in high yields opens up interesting possibilities.

Two studies have been published to date that give new insights into the mechanisms of protein folding and quality control using catalyzed cytoplasmic disulfide bond formation. The first publication from Riemer and co-workers used it to produce domains of ERp90, as part of a study identifying ERp90 as an interaction partner of ERFAD, a flavoprotein involved in ER-associated

degradation [31]. The second detailed a molecular analysis of the ER sulfhydryl oxidase Ero1 and identified two novel regulatory mechanisms: i) high cooperativity of oxygen binding – allowing disulfide bond formation to occur efficiently under hypoxic conditions but to be rapidly switched off under hyper-hypoxic conditions; ii) The formation of a dynamic regulatory mixed disulfide complex with PDI that involves cysteines in Ero1 that had previously been regarded as non-functional [32].

Given the potential in this area it is likely that more publications will come out over the next few years. This could set up a feedback system where cytoplasmic disulfide bond formation is used to make proteins which are used to identify novel mechanisms involved in protein folding, which in turn feeds into the generation of new, more efficient, protein cell factories.

Conclusions

Catalyzed, engineered systems for disulfide bond formation in the cytoplasm are still in their infancy, but they have potential for several areas of research, including functional and structural studies of secreted and outer membrane proteins, enzyme design, diagnostic production and even industrial scale production of disulfide bond containing proteins.

Declarations of interest

Patents for production systems used to make disulfide bond containing proteins in the cytoplasm are held by the University of Oulu with Lloyd Ruddock listed as the inventor.

Perspectives

- Disulfide bonds are found in a large number of proteins and hence having cell factories which can efficiently produce such proteins could have a significant impact in understanding disease and in disease intervention.
- Recently synthetic biology has been used to generate systems for disulfide bond formation in the cytoplasm of bacteria such as *E.coli*.
- More efficient systems for cytoplasmic disulfide bond formation will arise through understanding the natural mechanisms of oxidative folding and these will be combined with other enzymes to generate proteins which contain both disulfide bonds and other post translational modifications.

References

1. Herrmann, J.M. and Reimer, J. (2014) Three approaches to one problem: protein folding in the periplasm, the endoplasmic reticulum and the intermembrane space. *Antioxid Redox Signal*, **21**, 438-56 doi: 10.1089/ars.2014.5841

2. Saaranen, M.J. and Ruddock, L.W. (2013) Disulfide bond formation in the cytoplasm. *Antioxid Redox Signal*, **19**, 46-53 doi: 10.1089/ars.2012.4868

3. Derman, A.I., Prinz, W.A., Belin, D. and Beckwith, J. (1993) Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* **262**, 1744-1747 doi: 10.1126/science.8259521

4. Bessette, P.H., Åslund, F., Beckwith, J. and Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc Natl Acad Sci USA* **96**, 13703-13708 doi: 10.1073/pnas.96.24.13703

5. Stewart, E.J., Åslund, F. and Beckwith, J. (1988) Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J* **17**, 5543-5550 doi: 10.1093/emboj/17.19.5543

6. Kolberg, M., Strand, K.R., Graff, P. and Andersson, K.K. (2004) Structure, function, and mechanism of ribonucleotide reductases. *Biochim Biophys Acta* **1699**, 1-34 doi: 10.1016/j.bbapap.2004.02.007

7. Hatahet, F., Nguyen, V.D., Salo, K.E. and Ruddock, L. W. (2010). Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microb Cell Fact*, **9**, 67 doi:10.1186/1475-2859-9-67

8. Nguyen, V.D., Hatahet, F., Salo K.E.H., Enlund, E., Zhang C. and Ruddock, L.W. (2011) Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. *Microb Cell Fact*, **10**, 1 doi:10.1186/1475-2859-10-1

9. Frand, A.R. and Kaiser, C.A. (1999) Ero1p oxidizes protein disulfide isomerase in a pathway for disulfide bond formation in the endoplasmic reticulum. Mol Cell 1, 161-170 doi: 10.1016/s1097-2765(00)80198-7

10. Safavi-Hemami, H., Li, Q., Jackson, R.L., Song, A.S., Boomsma, W., Banyopadhyay, P.K., Gruber, C.W., Purcell, A.W., Yandell, M., Olivera, B.M. and Ellgaard, L. (2016) Rapid expansion of the protein disulfide isomerase gene family facilitates the folding of venom peptides. *Proc Natl Acad Sci USA* **113**, 3227-3232 doi: 10.1073/pnas.1525790113

11. Abskharon, R.N., Ramboarina, S., El Hassan, H., Gad, W., Apostol, M.I., Giachin, G., Legname, G., Stevaert J., Messens, J., Soror S.H. and Wohlkonig, A. (2012). A novel expression system for production of soluble prion proteins in E. coli. *Microb Cell Fact*, **11**, 6 doi: 10.1186/1475-2859-11-6

12. Abskharon, R., Dang, J., Elfarash, A., Wang, Z., Shen, P., Zou, L.S., Hassan, S., Wang, F., Gujioka, H., Stevaert, J., Mulaj, M., Surewicz, W.K., Castilla, J., Wohlkoing, A. and Zou, W.Q. (2017). Soluble polymorphic bank vole prion proteins induced by co-expression of quiescin sulfhydryl oxidase in E. coli and their aggregation behaviors. *Microb Cell Fact*, **16**, 170 doi: 10.1186/s12934-017-0782-x

13. Zhang, W., Zheng, W., Mao, M., & Yang, Y. (2014). Highly efficient folding of multi-disulfide proteins in superoxidizing escherichia coli cytoplasm. *Biotechnol Bioeng*, **111**, 2520-2527. doi:10.1002/bit.25309

14. Hatahet, F. and Ruddock, L.W. (2013) Topological plasticity of enzymes involved in disulfide bond formation allows catalysis in either the periplasm or the cytoplasm. *J Mol Biol* **425**, 3268-3276 doi: 10.1016/j.jmb.2013.04.034

15. Bulleid, N.J. and Freedman, R.B. (1988) Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes. *Nature* **335**, 649-651 doi: 10.1038/335649a0

16. Djender, S., Schneider, A., Beugnet, A., Crepin, R. Desrumeaux, K:E., Romani, C., Moutel S., Perez, F. and de Marco, A. (2014). Bacterial cytoplasm as an effective cell compartment

for producing functional VHH-based affinity reagents and camelidae IgG-like recombinant antibodies. *Microb Cell Fact,* **13**, 140. doi: 10.1186/s12934-014-0140-1

17. Nielsen, L.D., Foged, M.M., Albert, A., Bertelsen, A.B., Soltoft, C.L., Robinson, S.D. Petersen, S.V., Purcell, A.W., Olivera, B.M., Norton, R.S., Vasskog, T., Safavi-Hemami, H., Teilum, K. and Ellgaard, L. (2019). The three-dimensional structure of an H-superfamily conotoxin reveals a granulin fold arising from a common ICK cysteine framework. *J Biol Chem,* **294**, 8745-8759 doi: 10.1074/jbc.RA119.007491

18. Gaciarz, A., Veijola, J., Uchida, Y., Saaranen, M. J., Wang, C., Hörkkö, S. and Ruddock, L. W. (2016). Systematic screening of soluble expression of antibody fragments in the cytoplasm of E. coli. *Microl Cell Facts,* **15**, 22. doi: 10.10.1186/s12934-016-0419-5

19. Gaciarz, A., Khatri, N.K., Velez-Suberbie, M.L., Saaranen, M.J., Uchida, Y., Keshavarz-Moore, E. and Ruddock, L.W. (2017). Efficient soluble expression of disulfide bonded proteins in the cytoplasm of Escherichia coli in fed-batch fermentations on chemically defined minimal media. *Micro Cell Fact,* **16**, 108. doi: 10.1186/s12934-017-0721-x

20. Biterova, E., Esmaeeli, M., Alanen, H.I., Saaranen, M. and Ruddock, L.W. (2018). Structures of Angptl3 and Angptl4, modulators of triglyceride levels and coronary artery disease. *Sci Rep,* **8**, 7 doi: 10.1038/s41598-018-25237-7

21. Linton, D., Dorrell, N., Hitchen, P.G., Amber, A., Karyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, A. and Wren, B.W. (2005) Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. *Mol Microbiol* **55**, 1685-1703 doi: 10.1111/j.1365-2958-2005.04519.x

22. Valderrama-Rincon, J.D., Fisher, A.C., Merritt, J.H., Fan, Y.Y., Reading, C.A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., DeLisa, M.P. (2012) An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat Chem Biol* **8**, 434-436 doi: 10.1083/nchembio.921

23. Naegeli, A., Neupert, C., Fan, Y.Y., Lin, C.W., Poljak, K., Papini, A.M., Schwarz, F. and Aebi, M. (2014) Molecular analysis of an alternative N-glycosylation machinery by

functional transfer from Actinobacillus pleuropneumoniae to Escherichia coli. *J Biol Chem* **289**, 2170-2179 doi: 10.1074/jbc.M113.524462

24. Mohorko, E., Glockshuber, R. and Aebi, M. (2011) Oligosaccharyltransferase: the central enzyme of N-linked glycosylation. *J Inherit Metab Dis* **34**, 869-878 doi: 10.1007/s10545-011-9337-1

25. Mueller, P., Gauttam, R., Raab, N., Handrick, R., Wahl, C., Leptihn, S., Zorn, M., Kussmaul, M., Scheffold, M., Elkmanns, B., Elling, L. and Gaisser, S. (2018). High level in vivo mucin-type glycosylation in escherichia coli. *Micro Cell Fact* **17**, 168 doi:10.1186/s12934-018-1013-9

26. Bennett E.P., Mandel U., Clausen, H., Gerken, T.A., Fritz T.A. and Tabak, L.A. (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22** ,736-756 dor: 10-1093/glycob/cwr182

27. Harrus, D., Khoder-Agha, F., Peltoniemi, M., Hassinen, A., Ruddock, L., Kellokumpu, S. and Glumoff, T. (2018). The dimeric structure of wild-type human glycosyltransferase B4GalT1. *PloS One,* **13**, e0205571. doi: 10.1371/journal.pone.0205571

28. Ortiz-Soto, M.E. and Seibel. J (2016) Expression of functional human sialyltransferases ST3Gal1 and ST6Gal1 in escherichia coli. *PLoS One,* **11**, e0155410. doi: 10.1371/journal.pone.0155410

29. Lauber, J., Handrick, R., Leptihn, S., Duerre, P. and Gaisser, S. (2015) Expression of the functional recombinant human glycosyltransferase GalNAcT2 in escherichia coli. *Micro Cell Fact* **14**, 3. doi: 10.1186/s12934-014-0186-0

30. Schwab, K., Lauber, J. and Hesse, F. (2016). Fluorometric in situ monitoring of an escherichia coli cell factory with cytosolic expression of human glycosyltransferase GalNAcT2: Prospects and limitations. *Bioengineering (Basel)* **3**, E32 doi: 10.3390/bioengineering3040032.

31. Reimer J., Hansen, H.G., Appenzeller-Herzog C., Johansson L. and Ellgaard L. (2011). Identification of the PDI-family member ERp90 as an interaction partner of ERFAD. *PLoS One*, **6**, e17037. doi: 10.1371/journal.pone.0017037

32. Moilanen, A., Korhonen, K., Saaranen, M.J. and Ruddock, L.W. (2018) Molecular analysis of human Ero1 reveals novel regulatory mechanisms for oxidative protein folding. *Life Sci Alliance* **1**, e201800090 doi: 10.26508/lsa.201800090

33. Popovic, M., Mazzega, E., Toffoletto, B. and de Marco, A. (2018). Isolation of anti-extracellular vesicle single-domain antibodies by direct panning on vesicle-enriched fractions. *Micro Cell Fact* **17**, 9. doi: 10.1186/s12934-017-0856-9

34. Veggiani, G. and de Marco, A. (2011). Improved quantitative and qualitative production of single-domain intrabodies mediated by the co-expression of Erv1p sulfhydryl oxidase. *Protein Expr Purif* **79**, 111-114. doi: 10.1016/j.pep.2011.03.005

35. Holmberg, M.A., Gowda, N.K.C. and Andréasson, C. (2014). A versatile bacterial expression vector designed for single-step cloning of multiple DNA fragments using homologous recombination. *Protein Expr Purif* **98**, 38-45. doi: 10.1016/j.pep.2014.03.002

36. Khoshtinat Nikkhoi, S., Rahbarizadeh, F., Ahmadvand, D. and Moghimi, S.M. (2018). Multivalent targeting and killing of HER2 overexpressing breast carcinoma cells with methotrexate-encapsulated tetra-specific non-overlapping variable domain heavy chain anti-HER2 antibody-PEG-liposomes: In vitro proof-of-concept. *Eur J Pharm Sci* **122**, 42-50. doi: 10.1016/j.ejps.2018.06.019

37. Gąciarz A. and Ruddock, L.W. (2017). Complementarity determining regions and frameworks contribute to the disulfide bond independent folding of intrinsically stable scFv. *PLoS One*, **12**: e0189964. doi: 10.1371/journal.pone.0189964

38. Matos, C.F., Robinson, C., Alanen, H.I., Prus, P., Uchida, Y., Ruddock, L.W., Freedman, R.B. and Keshavarz.Moore, E. (2014) Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains. *Biotechnol Prog* **30**, 281-290 doi: 10.1002/btpr 1858

39. Jones, A.S., Austerberry, J.I., Dajani, R., Warwicker, J., Curtis, R., Derrick, J.P. and Robinson, C. (2016) Proofreading of substrate structure by the twin-arginine translocase is highly dependent on substrate conformational flexibility but surprisingly tolerant of surface charge and hydrophobicity changes. *Biochim Biophys Acta* **1863**, 3116-3124 doi: 10.1016/j.bbamcr.2016.09.006

40. Nativel, B., Figuester, A., Andries, J., Planesse, C., Couprie, J., Gasque, P., Viranaicken, W. and Iwema, T. (2016). Soluble expression of disulfide-bonded C-type lectin like domain of human CD93 in the cytoplasm of escherichia coli. *J Immunol Methods*, **439**, 67-73. doi: 10.1016/j.jim.2016.10.003

41. Roth, R., van Zyl, P., Tsekoa, T., Stoychev, S., Mamputha, S., Buthelezi, S. and Crampton, M. (2017). Co-expression of sulphydryl oxidase and protein disulphide isomerase in escherichia coli allows for production of soluble CRM197. *J Appl Microbiol* **122**, 1402-1411. doi: 10.1111/jam.13441

42. Ferreira, S.C.M., Torelli, F., Klein, S., Fyumagwa, R., Karesh, W.B., Hofer, H., Seeber, F. and East, M. L. (2018). Evidence of high exposure to toxoplasma gondii in free-ranging and captive african carnivores. *Int J Parasitol Parasites Wildl* **8**, 111-117. doi: 10.1016/j.ijppaw.2018.12.007

43. Auer, S., Azizi, L., Faschinger, F., Blazevic, V., Vesikari, T., Gruber, H. J. and Hytönen, V. P. (2017). Stable immobilisation of his-tagged proteins on BLI biosensor surface using cobalt. *Sensors & Actuators: B*, **243**, 104-113. doi: 10.1016/j.snb.2016.11.090

44. Griffo, A., Rooijakkers, B.J.M., Hähl, H., Jacobs, K., Linder, M.B. and Laaksonen, P. (2019). Binding forces of cellulose binding modules on cellulosic nanomaterials. *Biomacromolecules*, **20**, 769-777. doi: 10.1021/acs.biomac.8b01346

45. Rooijakkers, B.J.M., Martina S Ikonen, M.S. and Linder, M.B. (2018). Fungal-type carbohydrate binding modules from the coccolithophore emiliania huxleyi show binding affinity to cellulose and chitin. *PLoS One*, **13**, e0197875. doi: 10.1371/journal.pone.0197875

46. Schuetz, A., Radusheva, V., Krug, S.M. and Heinemann, U. (2017). Crystal structure of the tricellulin c-terminal coiled-coil domain reveals a unique mode of dimerization. *Ann N Y Acad Sci* **1405**, 147-159. doi:10.1111/nyas.13408

Figure legends

Figure 1: Mechanisms related to disulfide bond formation in the cytoplasm.

(A) Native disulfide bond formation in proteins occurs through two broad mechanisms, oxidation of a dithiol to a disulfide and isomerization of disulfide bonds, where the net number of thiols and disulfides in the protein does not change, but which cysteines are linked by disulfides does. In engineered systems Erv1p is usually the catalyst of oxidation and PDI the catalyst of isomerization. (**B**) Pathways for reduction in the cytoplasm of *E.coli*. Disruption of both pathways, for example Δ*trxB* / Δ*gor* strains, allows disulfide bond formation in POI, by the "reverse" of the usual function of thioredoxin. (**C**) Upper panel, the physiological function of thioredoxin (Trx) is to reduce disulfide bonds. In Δ*trxB* strains thioredoxin can still reduce physiological substrates e.g. ribonucleotide reductase, but needs to be re-reduced by transferring a disulfide bond to another protein e.g. a folding POI. Thioredoxin is not a sulfhydryl oxidase i.e. it cannot directly introduce disulfide bonds into folding proteins using molecular oxygen. (**D**) While reducing pathways (thioredoxin and reduced glutathione) can act in competition with oxidation for early stage folding intermediates, native proteins cannot be reduced by them. Hence catalyzed disulfide bond formation is possible even in wild-type *E.coli* strains with both reducing pathways intact, so long as the system for oxidation has sufficient activity. The reducing pathways can also aid native disulfide bond formation by reducing non-native disulfide bonds.

Figure 2: The mechanism of fusion proteins helping oxidative folding

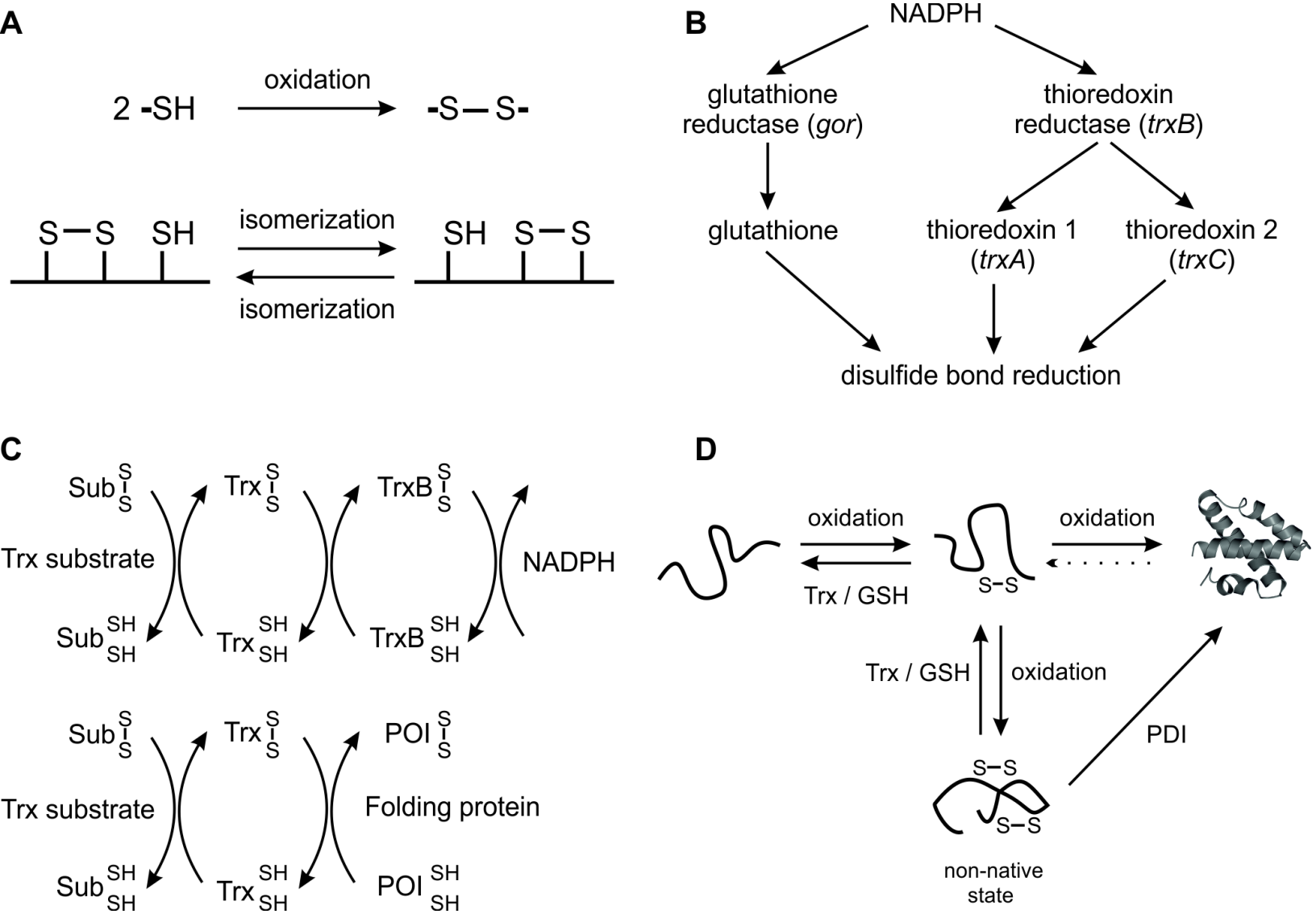
Oxidative protein folding is in competition with aggregation of folding intermediates i.e. non-native proteins. At its simplest, this can be thought of as a kinetic competition. Anything which increases the rate of folding e.g. the introduction of protein folding catalysts such as Erv1p and PDI, will increase the yield of folded protein. Similarly, anything which decreases the rate of aggregation will increase the yield of folded protein. This is the net effect of most fusion partners e.g. MBP. When a folding catalyst e.g. DsbC is used as the fusion partner, both effects may occur. The use of a fusion partner is a balance. Too little solubilization of folding intermediates and the fusion partner will not help obtain folded protein. However, too much solubilization of folding intermediates will result not only in folded protein being obtained, but also non-native folding intermediates i.e. false positives. Separation of different folded states of a soluble protein can be timely, complex and costly. The ideal fusion partner would give the maximum number of positives, while minimizing false positives

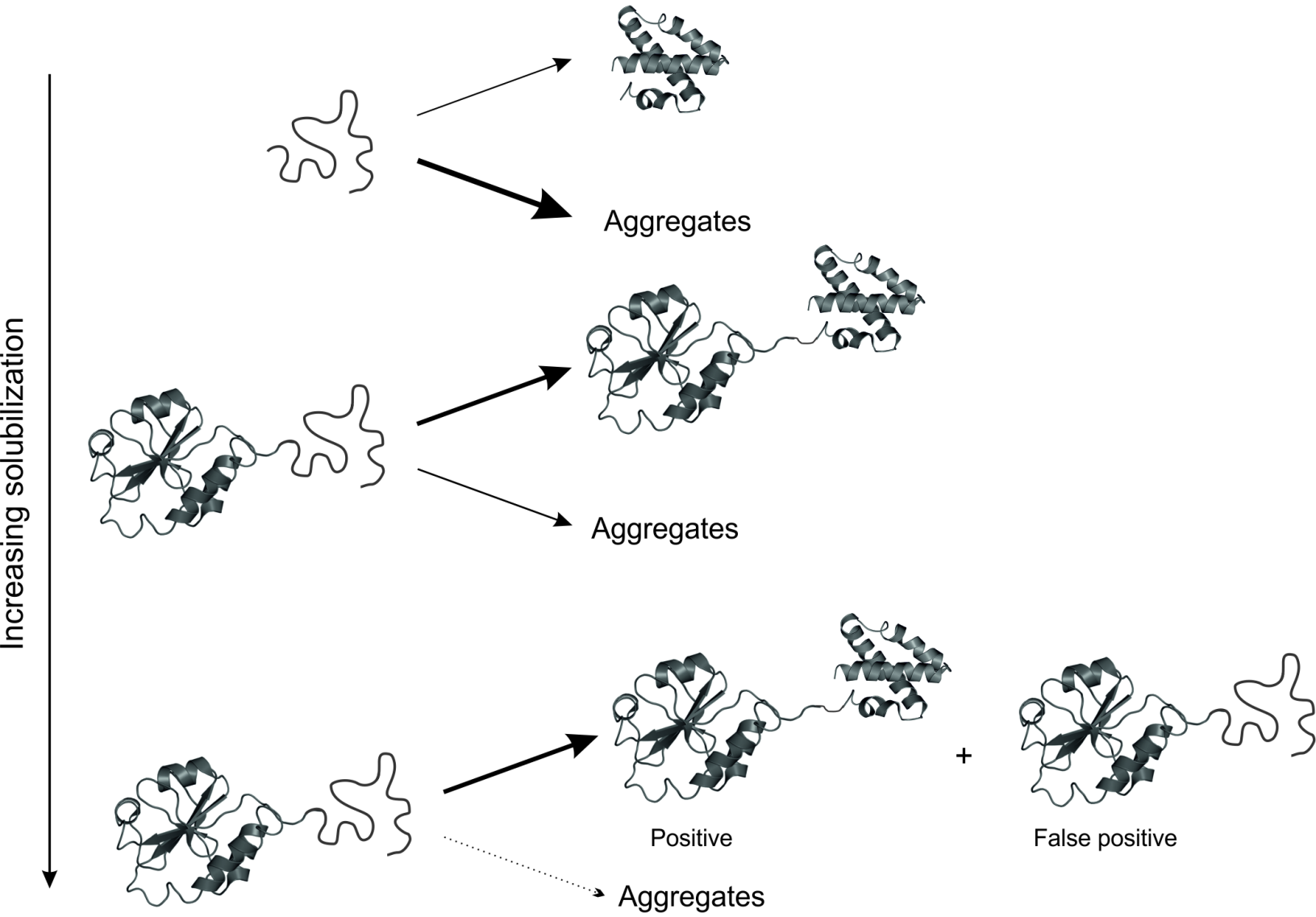
Figure 3 Combining PTMs in the cytoplasm of *E.coli*

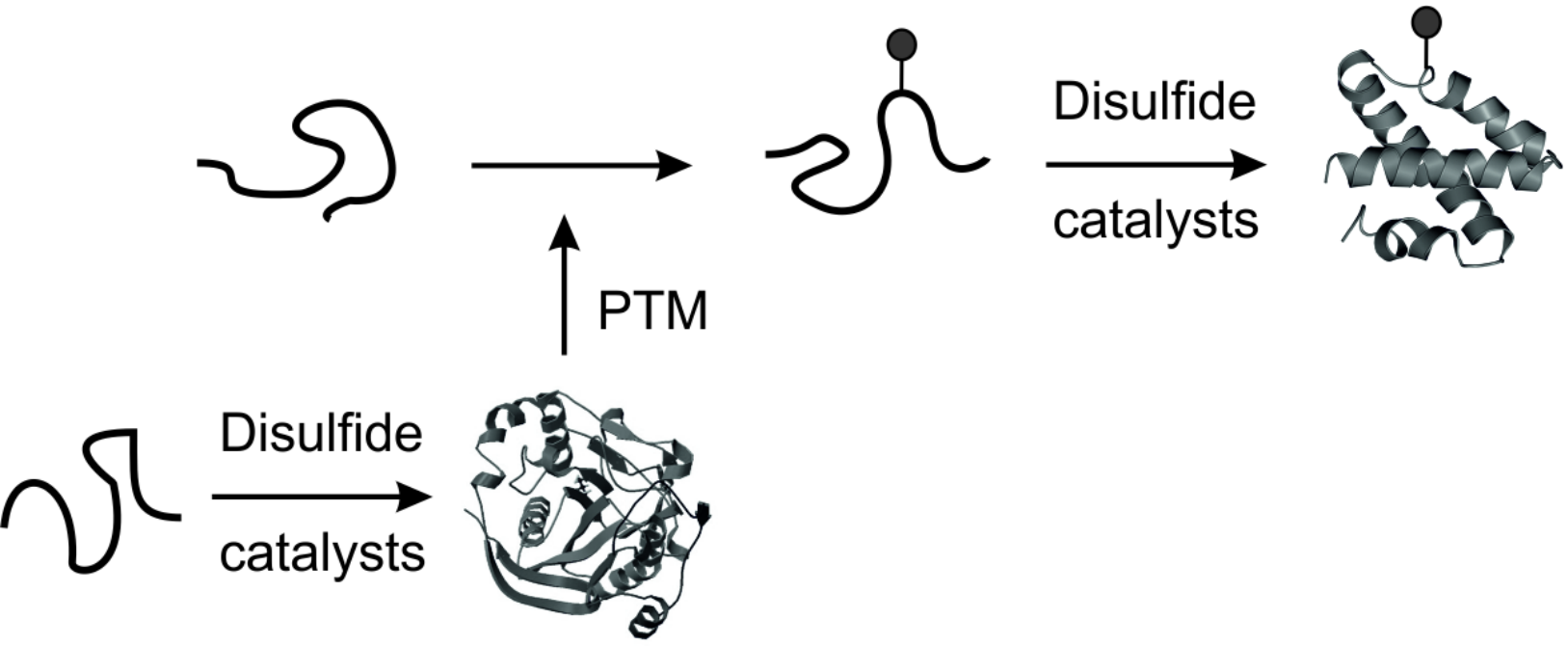
Many other PTMs are found in proteins that contain disulfide bonds. Often these are formed by enzymes that themselves contain disulfide bonds. By co-expressing the POI and PTM forming enzymes with catalysts of disulfide bond formation, it is possible to obtain folded POI which contain both disulfide bonds and other PTMs.

Table 1: Proteins successfully expressed with adding catalysts of disulfide bond formation including a sulfhydryl oxidase into the cytoplasm of *E.coli*. Proteins marked (I) have an inter-molecular disulfide linking a homodimer, while those marked (I*) have an inter-molecular disulfide linking a heterodimer

Protein	Disulfides	Reference
Antibody fragments		
VHH	1	[16, 33-36]
scFv	2	[18, 19, 37-39]
Fab	5 or 6 (I*)	[18]
Antigens / Vaccine development		
CD93 CTLD	4	[40]
CRM197	2	[41]
H-Vc7.2 (conotoxin)	3	[17]
SAG1	5	[42]
Cytokine / Growth Factor / Hormone		
BMP4	7 (I)	[8]
CSF3	2	[8]
HGH	2	[19]
IL6	2	[19]
IL17	2	[8]
Interferon alpha 2	2	[8]
Resistin	11 (I)	[8]
Enzymes		
AppA	4	[7, 38]
B4GalT1	2	[27]
Enterokinase fragment	4	[8]
Ero1α / Ero1β	6 or 7 (I*)	[8, 32]
GalNAcT2	5	[25, 29, 30]
Luciferase	5	[13]
PhoA	2	[7, 38]
ST3Gal1	3	[28]
ST6Gal1	3	[28]
vtPA	9	[8, 13]
Miscellaneous		
Angptl3/Angptl4 fragment	2	[20]
Avidin	1	[19, 43]
BPTI (and mutants)	2 or 3	[8]
CBM1	2	[44, 45]
ERp90 domain	1	[31]
Prion proteins	1	[11, 12]
Tricellulin	1 or 2	[46]


A
2 -SH
oxidation
-S—S-
S—S SH
isomerization
SH S—S
isomerization
B
NADPH
glutathione reductase (gor)
thioredoxin reductase (trxB)
glutathione
thioredoxin 1 (trxA)
thioredoxin 2 (trxC)
disulfide bond reduction
C
Sub
Trx
TrxB
Trx substrate
NADPH
Sub
Trx
TrxB
Sub
Trx
POI
Trx substrate
Folding protein
Sub
Trx
POI
D
oxidation
oxidation
Trx / GSH
S–S
Trx / GSH
oxidation
PDI
S–S
S–S
non-native state


Increasing solubilization
Aggregates
Aggregates
Positive
+
False positive
Aggregates


Disulfide
catalysts
PTM
Disulfide
catalysts