A novel in vivo screening system for protein folding and stability and investigations on variants of DsbA and DsbC from Escherichia coli

Author(s):
Philipps, Björn

Publication Date:
2002

Permanent Link:
https://doi.org/10.3929/ethz-a-004320423

Rights / License:
In Copyright - Non-Commercial Use Permitted
A novel *in vivo* screening system for protein folding and stability and investigations on variants of DsbA and DsbC from *Escherichia coli*

Björn Philipps

Swiss Federal Institute of Technology
Zürich

Diss. Nr. 14517
A novel in vivo screening system for protein folding and stability
and
investigations on variants of DsbA and DsbC from Escherichia coli

A dissertation submitted to the
Swiss Federal Institute of Technology Zürich
for the degree of
Doctor of Natural Science

presented by
Björn Philipps
Dipl. Chem. (Universität Freiburg im Breisgau, Germany)
born on March 12, 1973
Federal Republic of Germany

accepted on the recommendation of:
Prof. Dr. Glockshuber, examiner
Prof. Dr. Hilvert, co-examiner
2002
Table of contents

1. ABSTRACT

2. ZUSAMMENFASSUNG

3. INTRODUCTION

   3.1 Protein folding
       3.1.1 Interactions that contribute to the thermodynamic stability of proteins
       3.1.2 Levinthal’s paradox and folding pathways
       3.1.3 Protein folding in vivo

   3.2 Protein disulfide isomerases
       3.2.1 Protein disulfide isomerase from the endoplasmatic reticulum of eukaryotic cells
       3.2.2 The thioredoxin superfamily of proteins
       3.2.3 The thioredoxin fold and the Cys-Xaa-Xaa-Cys motif
       3.2.4 The intrinsic redox potential of thioredoxin-like proteins and the pKₐ value of the nucleophilic active-site cysteine
       3.2.5 Mechanism of disulfide bond formation
       3.2.6 The Dsb protein family from *Escherichia coli*

   3.3 DsbA, the main periplasmic oxidant of *Escherichia coli*
       3.3.1 The protein DsbA from *Escherichia coli*
       3.3.2 The three-dimensional structure of DsbA
       3.3.3 Reoxidation of DsbA by the inner membrane protein DsbB
       3.3.4 Catalytic properties of DsbA

   3.4 DsbC, the protein disulfide isomerase of *Escherichia coli*
       3.4.1 The protein DsbC from *Escherichia coli*
       3.4.2 The three-dimensional structure of DsbC
       3.4.3 Reduction of DsbC by the inner membrane protein DsbD
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3.5 In vivo and in vitro screening systems for protein folding and stability</strong></td>
</tr>
<tr>
<td>3.5.1 In vitro screening systems</td>
</tr>
<tr>
<td>3.5.2 In vivo screening systems</td>
</tr>
<tr>
<td>3.5.3 Flow cytometry screening systems</td>
</tr>
<tr>
<td><strong>3.6 The green fluorescent protein (GFP) from Aequorea victoria</strong></td>
</tr>
<tr>
<td>3.6.1 The three-dimensional structure of GFP</td>
</tr>
<tr>
<td>3.6.2 The formation of the GFP chromophore</td>
</tr>
<tr>
<td>3.6.3 Spectroscopic properties of GFP</td>
</tr>
<tr>
<td>3.6.4 Fluorescence resonance energy transfer (FRET)</td>
</tr>
<tr>
<td>3.6.5 FRET with different GFP variants</td>
</tr>
<tr>
<td><strong>3.7 Aim of the thesis</strong></td>
</tr>
<tr>
<td><strong>4. REFERENCES</strong></td>
</tr>
<tr>
<td><strong>5. RESULTS</strong></td>
</tr>
<tr>
<td>5.1 FRET-based in vivo screening system for intracellular protein folding and increased protein stability</td>
</tr>
<tr>
<td>Philipps, B., Hennecke, J. &amp; Glockshuber, R.</td>
</tr>
<tr>
<td>5.2 Randomization of the entire active-site helix α1 of the thiol-disulfide oxidoreductase DsbA from Escherichia coli</td>
</tr>
<tr>
<td>Philipps, B. &amp; Glockshuber, R.</td>
</tr>
<tr>
<td>5.3 Requirement for two active sites for efficient disulfide isomerase activity of disulfide isomerases</td>
</tr>
<tr>
<td>Frei, P., Philipps, B. &amp; Glockshuber, R.</td>
</tr>
<tr>
<td><strong>6. CURRICULUM VITAE</strong></td>
</tr>
<tr>
<td><strong>7. DANKSAGUNG</strong></td>
</tr>
</tbody>
</table>
1. Abstract

Since Anfinsen’s pioneering experiments (1973) it has been known that the final three-dimensional structure of a protein is exclusively determined by its primary sequence. However, it is still not possible to predict the amino acid sequence of a new protein with predetermined properties. This is caused, on the one hand, by the degeneracy of the folding code and, on the other hand, by the fact that no reliable methods are available to predict the catalytic properties or the thermodynamic stability of a protein, even if the three-dimensional structure is known.

In the first part of this thesis, a novel in vivo screening system for protein folding and stability was designed, which is independent of the function of the protein. In the second part, the folding and function of the thiol/disulfide oxidoreductases DsbA and DsbC from Escherichia coli were explored with different screening systems and biochemical studies.

Fluorescence resonance energy transfer (FRET) was used to establish a novel in vivo screening system that allows rapid detection of folded proteins and proteins with increased thermodynamic stability in the cytoplasm of Escherichia coli. The system is based on fusion of the green fluorescent protein (GFP) to the C-terminus of a protein “X” of interest, and fusion of blue-fluorescent protein (BFP) to the N-terminus of protein X. Efficient FRET from BFP to GFP in the ternary fusion protein is only observed in vivo when protein X is folded and brings BFP and GFP in close proximity, while FRET is lost when BFP and GFP are far apart due to unfolding or intracellular degradation of protein X. As models for protein X, four engineered immunoglobulin V L domains with different thermodynamic stabilities were inserted between BFP and GFP. These engineered V L domains do not require the conserved disulfide bond of antibody domains for folding and can thus be expressed in the reducing environment of the cytoplasm. After production of the ternary fusion proteins in E. coli, the in vivo FRET intensities correlate well with the thermodynamic stabilities of the inserted V L domains. Specifically, analysis of the cell extracts shows that the extent of intracellular proteolysis correlates exactly with V L stability. This can be explained by preferred proteolytic degradation of the unfolded state, which is populated to a higher extent at equilibrium in the case of a protein with lower stability.

The screening system was validated by identification of novel antibody V L domains with increased thermodynamic stability from expression libraries with randomized V L domains. After enrichment with a bacterial cell sorter, the selected variants indeed
Abstract

proved to be about 4 kJ mol\(^{-1}\) more stable than the original V\(\text{L}\) domain. The established in vivo screening system should thus be generally applicable for generation of proteins with improved stabilities.

In the second part of the thesis the thiol/disulfide oxidoreductases DsbA and DsbC from \textit{E. coli} were investigated. DsbA is a soluble, monomeric protein consisting of 189 amino acids and is essential for disulfide bond formation in the periplasmic space. Disulfide bond formation by DsbA is very rapid but fairly unspecific. This is compensated in vivo by the homodimeric disulfide isomerase DsbC, consisting of 2 x 215 amino acids, which catalyses reshuffling of non-native disulfide bridges. DsbA is the most oxidizing member of the thiol/disulfide oxidoreductase family, with a redox potential of -122 mV. The active site is located at the N-terminal end of the active-site helix \(\alpha_1\) with the sequence Cys\(^{30}\)-Pro\(^{31}\)-His\(^{32}\)-Cys\(^{33}\)-Tyr\(^{34}\)-Gln\(^{35}\)-Phe\(^{36}\)-Glu\(^{37}\). Previous mutagenesis studies on the dipeptide between the active-site cysteines have shown that the dipeptide sequence strongly influences the redox properties of the enzyme. Here, we have extended these random mutagenesis studies to the complete active-site helix such that all non-cysteine residues of the helix were randomized (amino acids 31, 32 and 34-37). Surprisingly, the vast majority (66\%) of the resulting variants proved to be biologically active and complemented the DsbA deficiency in a \(\text{dsbA}\) null strain. A large number of non-conservative replacements were found in active variants, even at well-conserved positions. This indicates that tertiary structure context determines \(\alpha\)-helical secondary structure formation of the randomized sequence. A subset of active and inactive variants was further characterized. All these variants were more reducing than wild type DsbA, but the redox potentials of biologically active variants did not drop below -210 mV. All inactive variants had redox potentials lower than -210 mV, although some of the inactive proteins could still be reoxidized efficiently by DsbB. This demonstrates that rapid oxidation of substrate polypeptides is the crucial property of DsbA in vivo.

All disulfide isomerases described so far contain at least two catalytically active thioredoxin domains. In the third part of the thesis it was investigated whether the catalytic properties of DsbA can be changed from a dithiol oxidase into a disulfide isomerase by artificial dimerization of DsbA. It was possible to demonstrate in vitro that a fusion protein consisting of DsbA and the dimerization domain of DsbC is indeed dimeric and can, in contrast to monomeric DsbA wild type, catalyze the isomerization of wrong disulfide bridges in proteins with multiple disulfide bonds.
2. Zusammenfassung


In dieser Doktorarbeit wurden zwei verschiedene Teilaspekte der Proteinfaltung betrachtet: Im ersten Teil konnte ein neues in vivo Screening-System für Proteinfaltung und Proteinstabilität entwickelt werden, das unabhängig von der Funktion des Proteins ist. Im zweiten Teil der Arbeit wurde mit Screening-Systemen und biochemischen Experimenten die Faltung und Funktion der Thiol/Disulfid Oxidoreduktasen DsbA und DsbC aus *Escherichia coli* untersucht.

Zusammenfassung


Redoxpotential, das kleiner als -210 mV war. Obwohl alle inaktiven Varianten ein Redoxpotential unterhalb von -210 mV aufwiesen, konnten einige von ihnen immer noch von DsbB als Oxidationsmittel regeneriert werden. Dies zeigt, daß die schnelle Oxidation der sich faltenden Polypeptide die entscheidende Funktion von DsbA in vivo ist.

3. Introduction

3.1 Protein folding

One of the fundamental questions in life science is how genetically encoded primary structure information is transformed into biological activity, i.e., how a newly synthesized polypeptide folds into an active protein. It is now generally accepted that the final three-dimensional structure and thus the function of a protein is exclusively determined by its amino acid sequence and that the native tertiary structure represents the state of lowest free energy of a polypeptide chain in aqueous solution. This was first shown in a classical in vitro refolding experiment of reduced, unfolded ribonuclease A (Anfinsen, 1973). The folding code however still remains mysterious. First, this is because the protein folding code is degenerate, meaning that identical peptide segments up to eleven residues may assume entirely different structures in proteins (Minor & Kim, 1996). Second, practically identical tertiary structures may be encoded by very different protein sequences so that about 25% sequence identity between two proteins is sufficient for an identical globular fold. Biophysical studies of protein folding have focused on structure (Perutz, 1951; Perutz et al., 1960), stability (Jaenicke, 1991c; Privalov, 1979; Tanford, 1968; Tanford, 1970) and folding mechanisms of proteins (Baldwin, 1995; Baldwin, 1996; Ikai et al., 1973; Kim & Baldwin, 1982; Kim & Baldwin, 1990; Schmid, 1992). Clearly, the solution of the protein folding problem would give us tools to predict the structure and possibly the function of a protein from its amino acid sequence and also allow the design of proteins with prescribed structures, functions and stabilities.

3.1.1 Interactions that contribute to the thermodynamic stability of proteins

The thermodynamic stability of a protein is defined as the difference in the Gibbs free energy ($\Delta G^\circ$) between the native and the unfolded state (Jaenicke, 1987):

$$\Delta G^\circ_{\text{Stab}} = \Delta G^\circ_{\text{Folded}} - \Delta G^\circ_{\text{Unfolded}} = \Delta H^\circ_{\text{Stab}} - T \cdot \Delta S^\circ_{\text{Stab}} \tag{1}$$

Proteins exhibit marginal stabilities equivalent to only a small number of weak interactions (Dill, 1990a; Jaenicke, 1991b). Average values for the Gibbs free energy of stabilization ($\Delta G^\circ_{\text{Stab}}$) of medium size globular proteins under physiological conditions are in the range of -10 to -100 kJ mol$^{-1}$ (Pfeil, 1998). The average
stabilization per residue is less than the random thermal energy, \( R \cdot T \), which is around 2.5 kJ mol\(^{-1}\) at room temperature. For a long time it was assumed that the ideal unfolded state of a polypeptide chain in aqueous solution is a random coil in which all rotation angles are independent of each other and all different conformations have comparable free energy (Privalov, 1979). Recently, however, theoretical calculations have shown that the angles of rotation from one amino acid are not independent of the conformation of other amino acids (Pappu et al., 2000), and measurements of NMR coupling constants of proteins in polyacrylamide gels also suggested a decreased conformational freedom of the unfolded state (Shortle & Ackerman, 2001). Despite such local conformational restrictions, a high conformational entropy (\( \Delta S_{\text{conf}}^0 \)) of the unfolded state is observed. Upon protein folding the decrease in chain entropy is not compensated by the increase in water entropy from the release of water of the hydrophobic core. (Bieri & Kiefhaber, 1999; Jaenicke, 1991a; Jaenicke, 1996; Jaenicke & Böhm, 2001). Significant contributions to the decrease of the chain enthalpy upon folding come from electrostatic forces between ionized, polar and non polar groups, such as ion pairs, hydrogen bonds, dipole interactions and dispersion forces. The physical nature of the hydrophobic effects therefore tends to be both entropic due to water release and enthalpic due to contributions of dispersion forces (Baldwin, 1986; Dill, 1990b; Kauzmann, 1959; Makhatadze & Privalov, 1996; Privalov & Gill, 1989; Schellman, 1997).

In secretory proteins, covalent disulfide bonds between cysteine residues may additionally contribute to protein stability. Disulfide bonds are usually buried in the folded conformation (Srinivasan et al., 1990). The decrease of conformational freedom results in a loss of entropy in the unfolded state (Flory, 1956; Matsumura et al., 1989). The gain in Gibbs free energy of folding from disulfide bridges (\( \Delta G_{\text{SS}}^0 \)) is mainly derived from the loss of entropy in the unfolded state and is suggested to correlate with the length of the loop formed by the disulfide bridge (\( N \) corresponds to number of residues in the loop) (Pace et al., 1988):

\[
\Delta S_{\text{SS}}^0 = \left[ -8.7 - \left( \frac{3}{2} \right) \cdot R \cdot \ln(N) \right] \text{J mol}^{-1} \text{K}^{-1} \tag{2}
\]
3.1.2 Levinthals’ paradox and folding pathways

If protein folding were based on a random search through the huge number of possible conformations, proteins would require an astronomical amount of time to find the single, most stable native structure (Levinthal, 1968). As protein folding in practice may occur at the submillisecond time scale, however it has been argued that defined pathways must exist.

The ‘hydrophobic collapse model’ assumes that the folding process is initiated by rapid burial of hydrophobic side chains, thus decreasing the conformational search space. The rearrangement to the native structure would then be the rate-limiting step in protein folding (de Gennes, 1985; Dill, 1985; Thirumalai, 1995). The classical ‘nucleation model’ proposes that the rate-limiting step in protein folding is the formation of a nucleated species, followed by formation of the native state on a much faster time scale. Thus, the nucleus does not become significantly populated (Ristow & Wetlauffer, 1973; Tsong et al., 1971; Wetlauffer, 1973; Wetlauffer & Ristow, 1973).

In the ‘framework model’ it is assumed that elements of secondary structure are formed prior to any tertiary structure (Kim & Baldwin, 1990; Ptitsyn, 1973; Ptitsyn, 1991). These mobile elements diffuse until they collide effectively and stabilize each other by local docking rearrangements. The rate-limiting step in this model has been proposed to be diffusion (Karplus & Weaver, 1976; Karplus & Weaver, 1994) or docking (Kim & Baldwin, 1990; Ptitsyn, 1973; Ptitsyn, 1991) of the preformed elements. Finally, small one-domain proteins with less than 100 amino acids may fold without any intermediates through a single transition state (Jackson & Fersht, 1991; Radford, 2000).

All these models have been developed using computer simulations and in vitro measurements of protein folding, with the unfolded, full-length polypeptide chain as a starting point.

3.1.3 Protein folding in vivo

Protein folding in vivo may start either after release of the synthesized polypeptide chain from the ribosome or may occur cotranslationally on the ribosome. In the latter case the N-terminal part of the polypeptide would already fold while the C-terminal protein segment is still being synthesized. Although protein folding is a spontaneous
process *in vitro*, different enzymes catalyze the slow steps of folding *in vivo* i.e., disulfide bond formation is catalyzed by protein disulfide isomerases and Xaa-Pro cis-trans isomerization is catalyzed by peptidyl prolyl cis-trans isomerases (PPI’s) (Bardwell *et al.*, 1991; Schmid, 1992).

Nonspecific aggregation of unfolded and misfolded proteins is prevented by molecular chaperones *in vivo*. Molecular chaperones are defined as proteins that recognize exposed, hydrophobic surfaces in unfolded or partially unfolded polypeptides and make non-covalent interactions with these proteins to prevent non-specific aggregation. The molecular chaperones can be divided into ATP-independent and ATP-dependent chaperones, in which the substrate binding and the release are regulated by ATP-dependent conformational changes (Bukau & Horwich, 1998). Molecular chaperones, many of which belong to the family of the heat shock proteins (hsp), are also involved in other cellular processes, such as translocation of proteins across membranes, conformational regulation of signal transduction molecules, and protein degradation (Martin & Hartl, 1997). The heat shock proteins are classified into five different chaperone families: Hsp100, Hsp90, Hsp70, Hsp60 and small Hsps (Bukau & Horwich, 1998).

Peptidyl prolyl peptide bonds are different from all other normal peptide bonds, because the proline ring makes the trans conformation energetically almost as unfavorable as the cis conformation. Therefore, both cis and trans conformations of the peptidyl prolyl bonds can be observed in native proteins. In unfolded proteins, rotation about the Xaa-Pro peptide bonds occurs with an equilibrium fraction of 10 to 40% of the cis-isomer (Schmid, 1992). The Xaa-Pro cis-trans isomerization also contributes significantly to the conformational heterogeneity of the unfolded state due to its high activation energy of about 90 kJ/mol, corresponding to half-life times of 10 to 100 s at 25 °C. This isomerization is always the rate-limiting folding step in proteins with cis-Xaa-Pro peptide bonds in the native state and is catalyzed by the enzymes of the peptidyl prolyl cis-trans isomerase (PPI) family (Schmid, 1992). These ubiquitous enzymes are found in all organism and most cellular compartments. To date, four different classes of PPI are known: the cyclophilins (Fischer *et al.*, 1989), the FK500 binding proteins (FKBP’s) (Harding *et al.*, 1989), the parvulins (Rahfeld *et al.*, 1994) and the trigger factor of E. coli (Stoller *et al.*, 1995).
3.2 Protein disulfide isomerasers

Proteins containing inter- or intramolecular disulfide bridges are present in most organisms. These disulfide bonds are a typical feature of secretory proteins formed in oxidative compartments such as the endoplasmatic reticulum (ER) in eukaryotic cells and the periplasm in prokaryotic cells. Disulfide cross-links are often crucial for the function and stability of folded proteins, and reduction of disulfide bridges frequently causes complete unfolding of the protein. Disulfide bond formation is rate-limiting for folding because the disulfide bridge between two cysteines cannot form spontaneously and requires a bimolecular reaction between the polypeptide and an oxidant. In the case of proteins with multiple disulfide bonds, where the number of possible disulfide cross links increases by around one order of magnitude with each additional cysteine pair, isomerization of wrong disulfides is often another rate-limiting step for folding and may even become slower than disulfide bond formation.

3.2.1 Protein disulfide isomerase from the endoplasmatic reticulum of eukaryotic cells

The ER of eukaryotic cells is an oxidizing environment as judged by the ca. 100 fold higher level of oxidized glutathione compared to the cytosol (Hwang et al., 1992). Since the formation of disulfide bridges is a slow process, it is catalyzed by the enzyme protein disulfide isomerase (PDI), which is present in the ER in a concentration of 1 mM (Noiva & Lennarz, 1992). Many studies have shown that PDI is crucial for disulfide bond formation in the ER in vivo. Microsomes that are deficient in all soluble ER proteins (including PDI) can still transport proteins but are not longer able to form disulfide bonds. This defect can be completed by addition of PDI (Bulleid & Freedman, 1988). The importance of the PDI protein is indicated by the fact that disruption of its encoding gene is lethal in yeast cells (LaMantia & Lennarz, 1993). Recently, it was shown that the reoxidation of PDI in yeast is performed by the flavin adenine dinucleotide (FAD) dependent membrane protein Ero1p (Tu et al., 2000). PDI catalyzes both the oxidation and the isomerization of disulfide bonds, but the most important function of the enzyme is presumably the rearrangement of incorrect disulfide bonds (Darby et al., 1994; Ruoppolo et al., 1996).

PDI is a homodimeric 55 kDa protein consisting of five domains termed a, b, b’, a’ and c (Darby et al., 1996). The a and the a’ and the b and the b’ domains are...
homologous to each other, and all domains were suggested to share the thioredoxin fold (Kemmink et al., 1997), which was confirmed by the structures of the a and the b domain (Kemmink et al., 1996; Kemmink et al., 1999). The a and a’ domains are catalytically active and contain the active site sequence Cys-Gly-His-Cys (Edman et al., 1985). Importantly, PDI is only an efficient disulfide isomerase when both catalytic domains (a and a’) and the b’ domain are present (Darby et al., 1998a).

3.2.2 The thioredoxin superfamily of proteins

Thioredoxin (Trx) was first described as a hydrogen donor for ribonucleotide reductase, the enzyme that converts ribonucleotides to deoxyribonucleotides (Reichard, 1962). In 1964, thioredoxin was isolated from the cytoplasm of *Escherichia coli* and named (Laurent et al., 1964). Thioredoxin is reduced by the NADPH-dependent flavoenzyme thioredoxin reductase (Holmgren, 1977). A second hydrogen donor enzyme for ribonucleotide reductase was found in a thioredoxin deficient *E. coli* strain and named glutaredoxin (Grx) for its use of the tripeptide glutathione as reductant (Holmgren, 1976). Because *E. coli* strains lacking both Trx and Grx are still viable, a third hydrogen donor for ribonucleotide reductase was postulated and two proteins with thioredoxin-like activity in vivo have been reported: NrdH (Jordan et al., 1997) and TrxC (Miranda-Vizuete et al., 1997).

Other in vivo functions of thioredoxin include hydrogen donation to methionine sulfoxide reductase (Black et al., 1960), sulfate reduction (Russel & Holmgren, 1988), participation in the light and dark cycle of chloroplasts (Wolosiuk & Buchanan, 1977), and redox control of the transcription factors NF κB and OxyR (Hayashi et al., 1993; Holmgren & Bjornstedt, 1995; Zheng et al., 1998).

3.2.3 The thioredoxin fold and the Cys-Xaa-Xaa-Cys motif

The basic tertiary structure of all thioredoxin proteins was first observed in the X-ray structure of *E. coli* thioredoxin and is called the thioredoxin fold (Holmgren et al., 1975; Katti et al., 1990). This fold is shared by all thiol/disulfide oxidoreductases like Trx, Grx, DsbA, DsbC, the a, a’, b and b’ domains of PDI and a few other proteins (Holmgren et al., 1975; Kemmink et al., 1996; Martin, 1995; Martin et al., 1993; McCarthy et al., 2000; Sodano et al., 1991). The basic fold is exemplified by the
structure of prokaryotic glutaredoxin, the smallest member of the family (Sodano et al., 1991) and consists of N-terminal βαβ and C-terminal ββα units connected by an α-helix (Martin, 1995). The central four-stranded β-sheet, in which β3 is antiparallel to the other strands, is flanked by α1 and α3 running parallel to the β-strands (Figure 1) (Martin, 1995).

Figure 1: Architecture of the thioredoxin fold (Martin, 1995). β-Sheet strands are drawn as arrows and α-helices as rectangles. The dashed lines indicate the separation of the fold into N-terminal (βαβ) and C-terminal (ββα) motifs connected by α-helix 2.

Enzymes from the thioredoxin family possess an active-site disulfide bond. It is located at the N-terminus of α-helix 1 and characterized by two cysteine residues separated by two amino acids (Martin, 1995). For example in thioredoxin the relevant sequence is Cys\textsuperscript{32}-Gly\textsuperscript{33}-Pro\textsuperscript{34}-Cys\textsuperscript{35} (Holmgren, 1968). This Cys-Xaa-Xaa-Cys (Xaa corresponds to any amino acid) motif has subsequently been found in all other cytoplasmic disulfide oxidoreductases, such as DsbA, DsbC, DsbG and Grx, and PDI (Andersen et al., 1997; Bardwell et al., 1991; Edman et al., 1985; Höög et al., 1983; Missiakas et al., 1994).

3.2.4 The intrinsic redox potential of thioredoxin-like proteins and the pK\textsubscript{a} value of the nucleophilic active-site cysteine

Although all members of the thioredoxin superfamily share the Cys-Xaa-Xaa-Cys motif, their intrinsic redox potentials vary strongly. The ability of the enzymes to act as oxidants follows the order DsbA > DsbG > DsbC > PDI > Grx > Trx. The redox potentials range from -122 mV for DsbA (Wunderlich & Glockshuber, 1993; Zapun et
Introduction

al. et al., 1993) to -270 mV for Trx (Krause et al., 1991; Lin & Kim, 1989). The equilibrium constants of the enzymes with glutathione thus vary by a factor of 100,000. Recent biophysical studies have revealed that the Xaa-Xaa dipeptide between the active-site cysteines strongly influences the redox properties of the enzyme (Chivers et al., 1996; Grauschopf et al., 1995; Holst et al., 1997; Huber-Wunderlich & Glockshuber, 1998; Kortemme et al., 1996; Krause et al., 1991; Mössner et al., 1998; Rossmann et al., 1997).

The redox potential of thioredoxin-like proteins is usually determined by measuring the equilibrium constant with glutathione (keq), which yields $E_o'$ of the enzyme through the Nernst equation (3), assuming a standard redox potential of -240 mV for the redox couple of oxidized glutathione (GSSG) and reduced glutathione (GSH) (Rost & Rapoport, 1964).

$$E_o' = -240 \text{ mV} - \frac{RT}{zF} \cdot \ln K_{eq}$$

(3)

3.2.5 Mechanism of disulfide bond formation

In the enzymes of the thioredoxin superfamily, the N-terminal active-site cysteine (PSH) is generally solvent exposed. It is a strong nucleophile and forms mixed disulfides with substrates. The more C-terminal cysteine (PSh) is always buried and only reacts with the N-terminal active-site cysteine (Freedman et al., 1994; Kallis & Holmgren, 1980; Wunderlich et al., 1993a; Zapun et al., 1994). The equilibrium between a disulfide oxidoreductase (P) and glutathione (G) comprises four individual disulfide exchange reactions with individual rate constants $k_1$ to $k_4$.

$$P^{SH} + GSSG \xrightleftharpoons[k_2]{k_1} P^{SSG} + GSH \xrightleftharpoons[k_4]{k_3} P^S + 2 \text{GSH}$$

(4)

The values of the rate constants $k_1$, $k_2$ and $k_4$ can be calculated from the Brønsted theory of general acid/base catalysis and the pKₐ values of the thiols involved (Szaejewski & Whitesites, 1980) using equation (5), where $k^{obs}$ is the observed rate constant at a given pH and $pK_a^{nuc}$, $pK_a^c$ and $pK_a^{lg}$ are the pKₐ values of the corresponding nucleophilic, central and leaving group thiol, respectively. The only uncertainty is the prediction of the intramolecular rate constant $k_3$, which additionally depends on the effective concentration of the buried, C-terminal thiolate attacking the mixed disulfide. Simulations have revealed that the apparent concentration of the buried active-site cysteine is in the range of 1 M (Mössner et al., 2000).
Knowing the pKₐ values of all thiols involved in the reaction and the effective concentration of the buried active-site cysteine, it is possible to predict the forward and reverse reaction rates and, consequently, the equilibrium constants (Kₑq) for the reaction between the thiol/disulfide oxidoreductase and a low-molecular-weight thiol/disulfide pair such as GSH/GSSG (equation 6). This overall linkage between redox potential and pKₐ value is a very useful tool for understanding the redox properties of thiol/disulfide oxidoreductases.

\[
\log(K_{eq}) = \left( 6.3 + 0.59 \cdot pK_{a}^{\text{obs}} - 0.40 \cdot pK_{a}^{\text{cys}} - 0.59 \cdot pK_{a}^{\text{eq}} \right) - \log\left( 1 + 10^{pK_{a}^{\text{cys}} - pH} \right)
\]  

(5)

Combination of equations (5) and (6), assuming standard conditions (pH = 7.0) and glutathione as the reference thiol compound with a pKₐ value of 8.7, yields equation (7). The further assumption that the buried cysteine (pKₐ of Cys, bur) has an invariant pKₐ of 11.7 yields the equation (8). With these equations it is possible to predict the dependence of Eₐ' of disulfide oxidoreductases on the pKₐ of the nucleophilic active-site cysteine (pKₐ of Cys, nuc) (Mössner et al., 2000).

\[
\log K_{eq} = \log\left( \frac{k_{1}^{\text{obs}} \cdot k_{3}^{\text{obs}}}{k_{2}^{\text{obs}} \cdot k_{4}^{\text{obs}}} \right)
\]

(6)

\[
\log K_{eq} = 1.18 \cdot (pK_{a}^{\text{Cys,nuc}} + pK_{a}^{\text{Cys,bur}} - 17.4) + 3.417 - \log\left( 1 + 10^{pK_{a}^{\text{Cys,nuc}} - 7} \right) - \log\left( 1 + 10^{pK_{a}^{\text{Cys,bur}} - 7} \right)
\]  

(7)

\[
\log K_{eq} = 1.18 \cdot (pK_{a}^{\text{Cys,nuc}} - 5.73) - \log\left( 1 + 10^{pK_{a}^{\text{Cys,bur}} - 7} \right) - 1.253
\]

(8)

### 3.2.6 The Dsb protein family from Escherichia coli

The properties of the periplasm of *E. coli*, such as its chemical composition, ionic strength, pH and redox potential, are largely determined by the surrounding medium, since the outer membrane of *E. coli* is permeable to molecules smaller than 500 Da. Since disulfide bond formation and isomerization are strongly pH dependent and inherently slow compared to folding itself (Shaked et al., 1980; Szajewski & Whitesites, 1980), both reactions need to be catalyzed in the periplasm.

The formation of correct disulfide bridges in the periplasm is catalyzed by the Disulfide bond formation (Dsb) proteins. In contrast to PDI in the ER, different enzymes DsbA and DsbC, catalyze disulfide bond formation and isomerization in *E. coli*. The most efficient oxidation catalyst in the periplasm is the protein DsbA (Bardwell et al., 1991). DsbA is reoxidized by the inner membrane protein DsbB.
Introduction

(Bardwell et al., 1993). DsbC is the thiol-disulfide isomerase of E. coli responsible for the reshuffling of the nonnative disulfide bonds randomly formed by DsbA (Missiakas et al., 1994). DsbC must be kept in the reduced state for catalytic activity, which is accomplished by electron flow from cytoplasmic NADPH through the inner membrane protein DsbD (Figure 2) (Rietsch et al., 1997).

DsbG has been discovered as a further member of the periplasmic Dsb family and is homologous to DsbC (Andersen et al., 1997; van Straaten et al., 1998). Its overexpression allows E. coli cells to grow at otherwise lethal conditions of reduced 1,4-dithio-DL-threitol (DTT) and null mutants of DsbG are only viable in the presence of low molecular weight disulfides. DsbG appears to have a more limited substrate specificity compared to DsbC (Bessette et al., 1999; Shao et al., 2000).

Figure 2: Disulfide oxidoreductases involved in disulfide formation and isomerization in the periplasm of Escherichia coli (Glockshuber, 1999).

DsbE (also called CcmG) is another reductase in the periplasm that can suppress phenotypic defects of dsbD− strains, but its function is much more specialized towards biogenesis of cytochrome c (Fabianek et al., 1997; Missiakas & Raina, 1997; Monika et al., 1997). In contrast to other Dsb proteins, DsbE does not reduce insulin
in vitro and does not influence the folding of alkaline phosphatase. This suggests that DsbE (CcmG) acts specifically in the cytochrome c biogenesis pathway.

The functions of the periplasmic Dsb protein family are summarized in Figure 2 (Glockshuber, 1999).

3.3 DsbA, the main periplasmic oxidant of *Escherichia coli*

DsbA was first discovered in a mutant of *E. coli* defective for disulfide bond formation in periplasmic proteins (Bardwell *et al.*, 1991; Kamitani *et al.*, 1992). A DsbA deficient strain shows a series of characteristic phenotypes that are directly or indirectly related to this defect. DsbA mutants are more sensitive towards DTT. They lack certain disulfide-bonded proteins, such as alkaline phosphatase, and lack disulfide bonds in other proteins, such as β-lactamase and the outer membrane protein A (OmpA) (Bardwell *et al.*, 1991; Kamitani *et al.*, 1992; Missiakas *et al.*, 1993). DsbA mutants are also immotile because they fail to assemble functional flagella (Dailey & Berg, 1993). This defect can be used to screen variants of DsbA for disulfide oxidase activity in vivo.

As a second in vivo test for activity of DsbA a fusion protein of the inner membrane protein MalF and β-galactosidase has been used in a strain lacking both dsbA and cytoplasmatic lacZ (Bardwell *et al.*, 1993; Bardwell *et al.*, 1991). In the fusion protein the β-galactosidase is exposed to the periplasm. In the presence of active DsbA, two non-natural disulfide bridges are introduced into the β-galactosidase domain of the fusion protein. These cross links prevent the β-galactosidase from being internalized into the cytoplasm, the only compartment where β-galactosidase is active. Activity of DsbA can then be recognized using a blue/white test with the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X–Gal).

3.3.1 The protein DsbA from *Escherichia coli*

DsbA is a soluble, monomeric 21 kDa protein with 189 amino acids. It contains a single, catalytic disulfide with the active-site sequence Cys$^{30}$-Pro$^{31}$-His$^{32}$-Cys$^{33}$ (Bardwell *et al.*, 1991; Kamitani *et al.*, 1992). DsbA is essentially found in the oxidized state in vivo (Joly & Swartz, 1997; Kishigami *et al.*, 1995). However, the disulfide bond is 4 to 10 orders of magnitude less stable than structural disulfide
bonds. The redox equilibrium constant with glutathione is $1.2 \cdot 10^{-4}$ M (Wunderlich & Glockshuber, 1993; Zapun et al., 1993). This makes DsbA the strongest oxidant in the thioredoxin family, which is in accordance with its in vivo function. Generally, disulfide bonds increase the stability of folded proteins. However, the stability of the oxidized form of DsbA is lower than that of the reduced form (Wunderlich et al., 1993a; Zapun et al., 1993). Early hypothesis that the difference in conformational stability could be explained by a strained conformation of the oxidized form have not been confirmed by comparison of the three-dimensional structures of oxidized and reduced DsbA (Guddat et al., 1998; Guddat et al., 1997b; Martin et al., 1993; Schirra et al., 1998; Wunderlich et al., 1993a). It is now accepted that the lower stability of oxidized DsbA is based on additional stabilizing interactions, which are only present in the reduced state. The most important factor is the stabilization of the thiolate anion of cysteine 30, which results in a drop of the $pK_a$ from the normal value of about 9.5 to a value of 3.5 (Nelson & Creighton, 1994; Zapun et al., 1994). The thiolate anion is stabilized by the partial positive charge resulting from the helix dipole at the N-terminus part of the active-site helix (Hol, 1985; Kortemme & Creighton, 1995) and the positive charge of histidine 32 (Guddat et al., 1997a; Kortemme et al., 1996).

Mutagenesis studies of the Cys-Xaa-Xaa-Cys motif have shown that the dipeptide between the active-site cysteines is largely responsible for the $pK_a$ value of DsbA (Grauschopf et al., 1995; Huber-Wunderlich & Glockshuber, 1998). Indeed, replacement of the Pro-His dipeptide by the active-site sequence of other thiol/disulfide oxidoreductases shifts the $pK_a$ value towards the corresponding value of the related enzyme.

### 3.3.2 The three-dimensional structure of DsbA

The three-dimensional structure of DsbA, which has been determined by X-ray crystallography in both redox states (Guddat et al., 1998; Guddat et al., 1997b; Martin et al., 1993) and in the reduced state by NMR spectroscopy as well (Schirra et al., 1998), consists of two domains. The catalytic domain (residues 1 to 62 and 139 to 189) shows a thioredoxin-like fold (coloured in blue in Figure 3) containing a mixed five-stranded β-sheet flanked by three α-helices (α1, C-terminal part of α6 and α7). The second domain (residues 63 to 138) is unique for DsbA. The latter is inserted
into the catalytic domain and consists of five α-helices (α2 to α5 and the N-terminal part of α6), it is called the α-helical domain (coloured in red in Figure 3). The two domains are connected by a loop between β-strand 3 and α-helix 2 and by the interdomain α-helix 6. The function of the α-helical domain is presently unknown, but a role in substrate binding and interaction with DsbB has been suggested (Guddat et al., 1997b; Martin et al., 1993). Circular permutation studies on DsbA also showed that the α-helical domain is required for the stability of DsbA. Helices α2, α3, α5 and α6 must be intact for formation of the native ternary structure of DsbA (Hennecke et al., 1999).

Figure 3: Ribbon diagram of the three-dimensional structure of oxidized DsbA from *Escherichia coli* (Guddat et al., 1997b; Martin et al., 1993). The thioredoxin-like domain and the α-helical domain are shown in blue and red, respectively. The active-site cysteines are indicated in a ball-and-stick representation. The figure was created with the program MOLMOL (Koradi et al., 1996).
3.3.3 Reoxidation of DsbA by the inner membrane protein DsbB

Due to the permeability of the outer bacterial membrane cellular compounds smaller than 500 Da are not retained in the periplasmic space. Hence, reoxidation of DsbA by a low molecular weight disulfide is not possible in *E. coli*. Instead, DsbA is reoxidized by the inner membrane protein DsbB (Figure 2). DsbB has been found in *E. coli* strains, which are not able to oxidize DsbA and show a phenotype comparable to *dsbA* strains, i.e., an inability to form disulfide bonds in periplasmic proteins and sensitivity towards DTT (Bardwell *et al*., 1993; Dailey & Berg, 1993; Missiakas *et al*., 1993). DsbB is a 21 kDa inner membrane protein consisting of four transmembrane helices and two small periplasmic domains of approximately 20 to 40 residues each (Jander *et al*., 1994; Kobayashi & Ito, 1999). DsbB possesses six cysteine residues, of which four are essential. Two of them are located in the first periplasmic domain, also within a Cys-Xaa-Xaa-Cys motif. However, it is unlikely that the structure of this domain shows a thioredoxin-like fold. The reoxidation of DsbA occurs via a mixed disulfide between Cys\(^{104}\) of DsbB and Cys\(^{30}\) of DsbA (Guilhot *et al*., 1995; Kishigami & Ito, 1996; Kishigami *et al*., 1995). The reoxidation of DsbB is then accomplished by transfer of electrons via ubiquinone, cytochrome bd or cytochrome bo from the respiratory chain to molecular oxygen, which serves as the final electron acceptor. Under anaerobic conditions, DsbB passes the electrons to menaquinone, which then donates electrons to anaerobic electron acceptors, such as fumarate reductase or nitrate reductase (Figure 2) (Bader *et al*., 1999; Bader *et al*., 1998; Bader *et al*., 2000; Kobayashi & Ito, 1999; Kobayashi *et al*., 1997).

3.3.4 Catalytic properties of DsbA

The active-site disulfide bond of DsbA is very reactive and oxidizes low molecular weight thiols, such as DTT or glutathione, several orders of magnitude faster than a normal alkyl disulfide or a structural protein disulfide (Wunderlich *et al*., 1993b; Zapun *et al*., 1993). This high activity, which can again be explained by the low pK\(_a\) value of cysteine 30, is especially pronounced at acidic pH where normal thiol/disulfide exchange reactions no longer occur (Nelson & Creighton, 1994; Wunderlich *et al*., 1995; Wunderlich *et al*., 1993b). Oxidation of reduced polypeptide substrates by DsbA is even one order of magnitude faster than oxidation of small thiols (Wunderlich *et al*., 1993b). This property has been explained by an additional non-covalent
interaction between DsbA and the polypeptide substrate and suggests the existence of a polypeptide binding site (Darby & Creighton, 1995; Frech et al., 1996b; Wunderlich et al., 1995; Zapun & Creighton, 1994). The peptide binding site is probably located in the long, deep groove between $\beta 5/\alpha 7$ and $\alpha 1$ (Figure 3). This groove harbors a hydrophobic pocket directly below the active-site that is large enough to accommodate a small hydrophobic side chain (Guddat et al., 1997b). In addition, amino acid side chains from both domains contribute to a hydrophobic patch just above the active-site.

DsbA oxidizes folding polypeptides extremely rapidly, with second-order rate constants close to the diffusion limit. *In vitro* studies on oxidation of polypeptides by stoichiometric amounts of DsbA have, however, shown that the disulfide bridges introduced by DsbA are formed randomly (Bader et al., 2000; Wunderlich et al., 1995; Wunderlich et al., 1993b; Zapun & Creighton, 1994). The efficiency of DsbA-catalyzed isomerization of incorrect disulfide bonds strongly depends on the substrate polypeptide (Akiyama & Ito, 1993; Akiyama et al., 1992; Joly & Swartz, 1994; Wunderlich et al., 1995; Wunderlich et al., 1993b; Zapun & Creighton, 1994). The disulfide isomerase activity of DsbA is, however, low compared to other thiol/disulfide oxidoreductases like DsbC and eukaryotic PDI (Darby & Creighton, 1995; Zapun et al., 1995). The very rapid formation of disulfide bridges at an early stage of the folding process appears necessary, as cysteine thiols would otherwise become buried in the folding intermediates and inaccessible to DsbA (Frech et al., 1996a).

### 3.4 DsbC, the protein disulfide isomerase of *Escherichia coli*

Because DsbA oxidizes folding polypeptides extremely rapidly but randomly, the disulfide isomerase DsbC is needed to reshuffle of the nonnative disulfides *in vivo* (Missiakas et al., 1994). DsbC was discovered in *E. coli* mutants, which showed lower DTT resistance than wild type strains (Missiakas et al., 1994). Nevertheless, the *dsbC* deficiency shows a milder phenotype than the lack of DsbA and does not affect cell motility or disulfide formation in OmpA (Rietsch et al., 1996).
3.4.1 The protein DsbC from Escherichia coli

DsbC is a soluble, homodimeric protein. The monomer consists of 215 amino acids with a molecular weight of 23 kDa (Zapun et al., 1995). Each subunit possesses four cysteine residues. Two of them are catalytic and have the active-site sequence Cys$^{98}$-Gly$^{99}$-Tyr$^{100}$-Cys$^{101}$. The other two cysteines of DsbC form a structural disulfide bond (Zapun et al., 1995). The active-site of DsbC has a redox potential of -131 mV (Zapun et al., 1995) and is therefore only slightly less oxidizing than DsbA. However, DsbC is mainly found in the reduced state in vivo and acts as a disulfide isomerase (Joly & Swartz, 1997; Maskos, 1995; Raina & Missiakas, 1997; Rietsch et al., 1996; Sone et al., 1997; Zapun et al., 1995). DsbC also shows a stronger tendency to bind polypeptides, which may be related to its dimeric state and a larger peptide binding site with higher affinity for unfolded polypeptides (Darby et al., 1998b). For this reason it has been suggested, that DsbA acts as the major oxidant in the functional periplasm, whereas DsbC mainly serves as a disulfide isomerase. For example, the periplasmic production of urokinase-type plasminogen activator, a protein with 12 disulfide bonds, is more than 100-fold lower in dsbC strains, whereas the production of alkaline phosphatase, which possesses only two disulfide bridges, is only slightly affected (Rietsch et al., 1996).

3.4.2 The three-dimensional structure of DsbC

The three-dimensional structure of DsbC in the oxidized form has been determined by X-ray crystallography and shows two domains per subunit (McCarthy et al., 2000). The N-terminal dimerization domain (residues 1 to 61) (shown in orange and red in Figure 4) is connected to the C-terminal catalytic domain (shown in green and blue via a linker α–helix (shown in yellow and violet). The catalytic domain is divided into two subregions: the thioredoxin domain (indicated by dark green and dark blue in Figure 4) (residues 78 to 128 and 167 to 216) and the α-helical domain (indicated by light green and light blue) (residues 123 to 166). The thioredoxin subdomain consists of a mixed four-stranded β-sheet, flanked by a single α-helix on one side and two α-helices on the other side. The dimer interface is formed by the antiparallel β-sheet (β1 to β4) of the dimerization domain from one monomer together with an antiparallel β-sheet (β5 and β6) from the dimerization domain of the other monomer. This dimerization yields a V-shaped DsbC-dimer, where the active-sites of the catalytic
domains are fairly flexible with an average distance of 38 Å (McCarthy et al., 2000). The dimerization of DsbC is necessary for the disulfide isomerase activity of DsbC (Sun & Wang, 2000).

Figure 4: Ribbon diagram of the homodimeric oxidized DsbC from *Escherichia coli* (McCarthy et al., 2000). The dimerization domains are coloured in gold and red and the linker helices are shown in yellow and violet. The catalytic domains are coloured in green and blue; the thioredoxin subdomain is depicted in dark colours and the α-helical subdomain in light colours. The active-site cysteines are indicated by ball-and-stick representation. The figure was created with the program MOLMOL (Koradi et al., 1996).

### 3.4.3 Reduction of DsbC by the inner membrane protein DsbD

DsbC is mainly found in the reduced state *in vivo*. It is reduced primarily by the inner membrane protein DsbD (Rietsch et al., 1997), which was identified on a null mutant that suppresses the phenotypic defects of the *dsbA* strain. This effect requires the *dsbC* genotype (Missiakas et al., 1995) and causes accumulation of oxidized DsbC in the periplasm (Andersen et al., 1997; Bessette et al., 1999). These results indicate that DsbD is required to keep DsbC in the reduced state *in vivo*. DsbD (also called DipZ and CutA2) is a 59 kDa transmembrane protein consisting of three domains
(Chung et al., 2000; Gordon et al., 2000): a N-terminal periplasmic domain (α-domain), a hydrophobic domain with eight transmembrane segments (β-domain) and a C-terminal, periplasmic domain (γ-domain), which shows about 40% sequence identity with the α domain of eukaryotic PDI. Each domain of DsbD has two conserved cysteine residues, where the two cysteines of the C-terminal domain (γ-domain) form the Cys-Xaa-Xaa-Cys motif.

Deletion of the genes encoding cytoplasmic thioredoxin or thioredoxin reductase lead to the accumulation of oxidized DsbC (Rietsch et al., 1996; Rietsch et al., 1997). The phenotypes shared by mutants of DsbD, thioredoxin and thioredoxin reductase indicate that these enzymes are members of the same pathway (Figure 2). A model for the action of DsbD has recently been proposed, which involves electron transfer from thioredoxin to the β-domain of DsbD and then successively to the γ and the α domains and finally to DsbC (Goldstone et al., 2001; Katzen & Beckwith, 2000; Krupp et al., 2001).

3.5 In vivo and in vitro screening systems for protein folding and stability

Since the advent of recombinant DNA technology, engineering of proteins for improved binding specificities, ligand affinities and stability has been of keen interest. In this context, rational design of proteins with improved stability appears to be a particularly difficult task because it is still not understood how sequence changes affect protein production, function and biophysical properties. Therefore, even sophisticated structure-based engineering approaches often fail in producing novel proteins without additional experimentation (Dougan et al., 1998; Yelton et al., 1995). A major challenge for combinatorial and evolutionary approaches is the screening for thermodynamically stabilized protein variants in complex libraries independent of protein function. Important for every screening system is the linkage between phenotype and genotype.
3.5.1 **In vitro screening systems**

The advantage of available *in vitro* selection systems for biological macromolecules with new function is the size of the library that can be screened, potentially up to \(10^{14}\) different sequences, whereas the library size of *in vivo* screening systems is normally restricted to about \(10^{10}\) different sequences (Plückthun et al., 2000). The most common *in vitro* selection systems are SELEX (selection for DNA and RNA aptamers) and ribosome display. The latter relies on formation of a stable, ternary complex between ribosomes, mRNA and the nascent polypeptide. Complexes containing folded proteins with a desired specificity are enriched by selective adsorption to immobilized ligands. The RNA is then reverse transcribed, followed by amplification of the DNA encoding the corresponding protein. (Hanes et al., 1998; He & Taussig, 1997; Jermutus et al., 1998; Mattheakis et al., 1994; Plückthun et al., 2000). Alternatively, a covalent link can be established directly between the DNA and the encoded protein (Doi & Yanagawa, 1999; Roberts & Szostak, 1997).

3.5.2 **In vivo screening systems**

Currently, the most often used method for *in vivo* protein library screening/selection is based on protein display on the surface of filamentous bacteriophage. In this technique, a gene of interest is fused to the gene3 protein at the tip of the fd phage, which simultaneously bears the genetic information of the fusion protein on its tip. Several rounds of affinity panning and reinfection allows identification of proteins with desired binding specificities. Usually four to six rounds of panning and amplification are required to select specific binders (Atwell & Wells, 1999; Cwirla et al., 1990; Demartis et al., 1999; Forrer et al., 1999; Pedersen et al., 1998; Spada et al., 1997).

Detection of interacting proteins and peptides is possible using the yeast two-hybrid system. The two domains of the yeast transcriptional activator protein GAL4 are separated and fused to two different proteins. Interaction of these two proteins leads to transcriptional activation of a reporter gene containing a binding site for GAL4 (Chien et al., 1991; Drees, 1999; Fields & Song, 1989). Recently, a three-hybrid system for the isolation of novel protein catalysts has been developed (Firestine et al., 2000). A similar phenomenon is used for *E. coli* cells and *in vitro* studies with the protein complementation assay (Pelletier et al., 1999; Pelletier et al., 1998).
3.5.3 Flow cytometry screening systems

Flow cytometry is a one-dimensional implementation of fluorescence microscopy, where different particles or cells flow at high dilution in a stream through an optical cell and a collection device. Flow cytometers measure light scattering and the fluorescence profile of the individual cells and can be programmed to collect cells with desirable properties (Shapiro, 1995).

In the context of in vivo screening of protein libraries flow cytometry has been used to sort proteins displayed on the surface of cells. For example, cells displaying antibodies on their surface can be sorted after a fluorescently labeled antigen has been bound. As host cells, both E. coli (Daugherty et al., 2000; Daugherty et al., 1998; Daugherty et al., 1999; Georgiou et al., 1993) and yeast (Boder & Wittrup, 1997; Kieke et al., 1997) have been used.

3.6 The green fluorescent protein (GFP) from Aequorea victoria

The green fluorescent protein (GFP) from the jellyfish Aequorea victoria was first discovered as a companion protein to aequorin, a chemoluminescent protein from Aequorea victoria (Shimomura et al., 1962). The GFP protein converts the blue emission from aequorin to the green glow of the intact cells (Morise et al., 1974). The fluorescence spectrum of GFP is probably responsible for the bioluminescence of living Aequorea (Johnson et al., 1962), but the biochemistry of the fluorescence in vivo is not yet fully understood.

3.6.1 The three-dimensional structure of GFP

GFP is a soluble, monomeric protein consisting of 238 amino acids with a molecular weight of 27 kDa. The structure of GFP contains an antiparallel 11-stranded β-barrel forming a unique β-can structure (Figure 5) (Ormö et al., 1996; Yang et al., 1996). This β-barrel structure is very rigid and responsible for the high resistance of GFP to denaturation. Unfolding of GFP cannot be achieved with 8 M urea and requires treatment with 6 M GdmCl or other extreme conditions (Bokman & Ward, 1981; Chiang et al., 2001; Yang et al., 1996). The interior of the protein harbors a coaxial α-
helical segment that also contains the GFP fluorophore (Figure 5), which is formed autocatalytically as shown below (Figure 6).

**Figure 5: Ribbon diagram of GFP from Aequorea victoria** (Yang et al., 1996). The atoms of the chromophore consisting of the tripeptide Ser$^{65}$-Tyr$^{66}$-Gly$^{67}$ are coloured in green. The figure was created with the program MOLMOL (Koradi et al., 1996).

GFP can be fused via its N- or C-terminus to other proteins (Chalfie, 1995) because both termini of GFP are rather flexible and accessible at the top of the $\beta$-can. The flexibility of the termini was shown with deletion variants of GFP, lacking the first six or the last nine residues, that are still fluorescent (Dopf & Horiagon, 1996; Li et al., 1997).

### 3.6.2 The formation of the GFP chromophore

The chromophore of the GFP protein is a p-hydroxybenzylidene-imidazolidone ring formed by autocatalytic intramolecular cyclization of the tripeptide segment Ser$^{65}$-Tyr$^{66}$-Gly$^{67}$ (Figure 6). First, GFP folds into a nearly native conformation, that allows autocatalytic cyclization of the polypeptide backbone by nucleophilic attack of the amide of Gly$^{67}$ on the carbonyl of Ser$^{65}$. After release of a water molecule to form the
imidazolinone ring, molecular oxygen oxidizes the Cα-Cβ bond of Tyr66 and generates a conjugated π-system between the tyrosine ring and the imidazolinone (Figure 6). Only at this stage does the chromophore absorb and fluoresce in the visible region. It is evident that air oxidation is the rate-limiting step of fluorophore formation. Atmospheric oxygen is required for developing fluorescence, and fluorescence of anaerobically produced GFP develops with a simple pseudo-first order time course only after air is readmitted to the system (Heim et al., 1994; Inouye & Tsuji, 1994).

![Figure 6: The proposed mechanism for the intramolecular biosynthesis of the GFP chromophore formed by the tripeptide Ser65-Tyr66-Gly67 (Cubitt et al., 1995). The amino acids Ser65, Tyr66 and Gly67 are shown in green, red and blue, respectively. The rate constants are estimated from experiments on the GFP variant Ser65Thr (Heim et al., 1995; Reid & Flynn, 1997).](attachment:figure6.png)

### 3.6.3 Spectroscopic properties of GFP

The absorption and fluorescence spectra of GFP are quite complex. GFP has a major absorbance peak at 395 nm that is about three times higher in amplitude than a minor peak at 475 nm. If GFP is excited at 395 nm it shows an emission peak at 508 nm, while emission occurs at 503 nm following excitation at 475 nm (Heim et al., 1994). The fact that the shape of the emission spectrum is dependent on the excitation wavelength leads to the conclusion that the chromophore can exist in at least two chemically different states. One assumes that these states correspond to
the neutral and the negatively charged chromophore (with the aromatic hydroxyl group deprotonated). The 475 nm peak arises from molecules harboring the anionic chromophore, whereas the 395 nm peak represents the neutral chromophore (Cubitt et al., 1995; Heim et al., 1994). Both forms of the chromophore are stabilized by a complex network of hydrogen bonds with site chains and backbone atoms from amino acids in five β-strands and the central helical segment and solvent molecules (Ormò et al., 1996).

The autocatalytic formation of the chromophore from atoms of the polypeptide chain itself is the reason why GFP and GFP variants have become an extremely important tool in biotechnology. In contrast to GFP other known bioluminescent proteins possess external cofactors such as lumazines (Lee et al., 1992) or flavins (Macheroux et al., 1987). Likewise, phycobiliproteins (Glazer, 1989) and peridinin chlorophyll-a protein (Song et al., 1976) use tetrapyrroles as pigments. Since availability and correct insertion of these cofactors into the apoproteins is not guaranteed in all organisms their attractiveness as spectroscopic tags and probes is limited. Thus far, aside from GFP there is only one fluorescent protein known that does not need an external cofactor: the red fluorescent protein drFP583, commercially known as DsRed (Matz et al., 1999). It is a homolog of GFP from a Discosoma species. GFP and DsRed share the same overall β-can fold, but have less than 30% sequence identity (Gross et al., 2000; Matz et al., 1999; Yarbrough et al., 2001). Unfortunately, DsRed is only fluorescent as a tetramer, which limits biotechnological applications of this protein (Baird et al., 2000).

For this reasons, GFP has become the most important reporter molecule for studies of protein localization and protein-protein interactions in vivo (Bastiaens & Pepperkok, 2000; Margolin, 2000). Many different variants of GFP with different spectral characteristics are available, such as the enhanced green fluorescent protein (EGFP), the blue fluorescent protein (BFP), the cyan fluorescent protein (CFP) and the yellow fluorescent protein (YFP). The absorption and fluorescence properties of these GFP variants and DsRed are shown in Table 1 (Tsien, 1998). Despite the variety of GFP variants, to date no red fluorescent variant of GFP has been created by rational or random mutagenesis of GFP.
3.6.4 Fluorescence resonance energy transfer (FRET)

Fluorescence resonance energy transfer (FRET) is one of the few tools available for measuring nanometer scale distances and changes in distances, both in vitro and in vivo. FRET is observed when a donor fluorophore is excited by light and transfers the excited state energy to an acceptor fluorophore in close proximity. This leads to a reduction in fluorescence intensity of the donor and an increase in emission intensity of the acceptor. Förster showed that the efficiency of transfer ($E_T$) of a given donor/acceptor couple depends on the inverse sixth power of the distance ($R$) between donor and acceptor (Förster, 1948):

$$E_T = \frac{R_o^6}{R^6 + R_o^6} \quad (9)$$

$R_o$ is the distance at which transfer efficiency is 50%. This value is dependent on the characteristics of every donor/acceptor couple. The values for the $R_o$ of a given donor/acceptor pair can either be measured directly or calculated from the overlapping integral of the fluorescence of the donor and the absorption of the acceptor. In turn, if $R_o$ is known, $E_T$ measurements allow the determination of the distance between donor and acceptor. FRET can therefore be used as a spectroscopic ruler (Stryer & Haugland, 1967). Since donor/acceptor pairs typically have $R_o$ values in the range of 20 to 60 Å, distances and distance changes in this range can be measured. In general FRET is better suited for detecting changes in distance than measuring absolute distances, because the energy transfer also depends on the relative orientation of the dyes, which is often unknown.

### Table 1: Spectroscopic characteristics of DsRed and the most important variants of GFP

<table>
<thead>
<tr>
<th>Protein</th>
<th>Excitation maximum</th>
<th>Emission maximum</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>395 nm &amp; 475 nm</td>
<td>508 nm &amp; 503 nm</td>
<td>a, b</td>
</tr>
<tr>
<td>EGFP</td>
<td>488 nm</td>
<td>508 nm</td>
<td>a - c</td>
</tr>
<tr>
<td>BFP</td>
<td>384 nm</td>
<td>448 nm</td>
<td>c</td>
</tr>
<tr>
<td>CFP</td>
<td>436 nm</td>
<td>485 nm</td>
<td>c, d</td>
</tr>
<tr>
<td>YFP</td>
<td>512 nm</td>
<td>522 nm</td>
<td>a, c</td>
</tr>
<tr>
<td>DsRed</td>
<td>558 nm</td>
<td>583 nm</td>
<td>e</td>
</tr>
</tbody>
</table>

Table 1: Spectroscopic characteristics of DsRed and the most important variants of GFP (Matz et al., 1999; Tsien, 1998); a: (Patterson et al., 1997); b: (Ward, 1997); c: (Cubitt et al., 1995); d: (Heim et al., 1994); e: (Matz et al., 1999).
3.6.5 FRET with different GFP variants

Different GFP variants, mainly the two couples BFP/EGFP and CFP/YFP are used for measurements of FRET in vitro and in vivo. The $R_o$ values of these couples have been calculated from the spectroscopic properties of the proteins (Table 1) to be 41.4 and 49.2 Å, respectively (Patterson et al., 2000).

Intramolecular FRET in fusion proteins containing a GFP donor and a GFP acceptor variant can be used to monitor polypeptide cleavage, calcium signaling and phosphorylation in vivo. The first demonstration of this technique was performed by fusion of BFP and EGFP, separated by a flexible linker containing a Factor Xa protease cleavage site. The cleavage of the linker was followed by a decrease in FRET signal (Mitra et al., 1996). Subsequently, proteolytic activity was also monitored in linkers containing either a trypsin cleavage site (Heim & Tsien, 1996) or a caspase-3 cleavage site (Luo et al., 2001; Xu et al., 1998).

In another type of experiment, the Ca$^{2+}$ dependence of the binding of calmodulin to the calmodulin binding peptide (CBP) was investigated by GFP-based FRET. BFP and EGFP were fused to the termini of CBP and the binding of calmodulin was detected by a decrease in FRET in vitro (Romoser et al., 1997) and in vivo (Persechini et al., 1997). In a similar experiment CFP and YFP were fused to calmodulin linked to CBP. The binding of calcium to calmodulin is followed by an increase in FRET in vitro and in vivo (Miyawaki et al., 1997; Truong et al., 2001).

Intermolecular FRET can also visualize protein-protein interactions. This form of FRET is much more difficult to observe, since unbound donor and acceptor molecules decrease the signal to noise ratio. Additionally, no FRET occurs if the distance or the orientation between the chromophores is unfavorable after complex formation. One example of FRET for measuring protein-protein interactions is the formation of the soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) complexes that are involved in synaptic vesicle formation (Xia et al., 2001). Also, the binding of hen egg lysozyme (HEL) to the variable domains of an anti-HEL antibody could be monitored by FRET. Each isolated variable antibody domain was fused to either BFP or EGFP and the increase in FRET was monitored upon association of the ternary complex formed by HEL and the two antibody domains (Arai et al., 2000).
3.7 Aim of the thesis

The first part of this thesis focuses on the development of a new *in vivo* screening system for protein folding and stability. The system is based on FRET from BFP to GFP in a ternary fusion protein. This FRET is only observed if protein X, which is located between BFP and GFP, is folded and brings the fluorescence donor/acceptor pair in close proximity. FRET is absent if protein X is unfolded or degraded.

The second part of the thesis addresses the importance of the amino acids from the active-site helix of the disulfide oxidoreductase DsbA on folding and catalytic activity of the enzyme. To probe sequence space for the active-site helix, all non-cysteine were randomized and biologically active and inactive variants were identified.

In the final part of the thesis artificial dimerization of DsbA was investigated as a means of changing the catalytic properties of DsbA from a dithiol oxidase into a disulfide isomerase *in vitro* and *in vivo*. This experiment was based on the empirical finding that all disulfide isomerases described so far contain at least two catalytically active thioredoxin domains.
4. References


References


Guddat, L. W., Bardwell, J. C., Glockshuber, R., Huber-Wunderlich, M., Zander, T. & Martin, J. L. (1997a). Structural analysis of three His32 mutants of DsbA:


Guddat, L. W., Bardwell, J. C., Zander, T. & Martin, J. L. (1997b). The uncharged surface features surrounding the active site of *Escherichia coli* DsbA are conserved and are implicated in peptide binding. *Protein Sci* **6**(6), 1148-56.


5. Results

5.1 FRET-based in vivo Screening for Intracellular Protein Folding and Increased Protein Stability

B. Philipps, J. Hennecke & R. Glockshuber
FRET-based in vivo Screening for Intracellular Protein Folding and Increased Protein Stability

Björn Philipps, Jens Hennecke and Rudi Glockshuber

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Hönggerberg, CH-8093 Zürich, Switzerland

To whom correspondence should be addressed:

Rudi Glockshuber
Institut für Molekularbiologie und Biophysik
Eidgenössische Technische Hochschule Hönggerberg
CH-8093 Zürich
Switzerland
Phone: +41-1-633 6819
Fax: +41-1-633 1036
E-mail: rudi@mol.biol.ethz.ch
Abstract

Fluorescence resonance energy transfer (FRET) was used to establish a novel in vivo screening system that allows rapid detection of protein folding and protein variants with increased thermodynamic stability in the cytoplasm of Escherichia coli. The system is based on the simultaneous fusion of the green fluorescent protein (GFP) to the C-terminus of a protein “X” of interest, and of blue-fluorescent protein (BFP) to the N-terminus of protein X. Efficient FRET from BFP to GFP in the ternary fusion protein is only observed in vivo when protein X is folded and brings BFP and GFP in close proximity, while FRET is lost when BFP and GFP are far apart due to a unfolding or intracellular degradation of protein X. The screening system was validated by identification of antibody VL intradomains with increased thermodynamic stabilities from expression libraries after random mutagenesis, bacterial cell sorting, and colony screening.
Introduction

Engineering of proteins with novel properties is one of the greatest challenges of modern protein chemistry, in particular since present theoretical prediction methods for protein function and stability, as well as structure-based engineering alone are often not sufficient for generating novel proteins with desired functions. For this reason, powerful random mutagenesis techniques in conjunction with screening and selection methods have been developed to complement rational engineering with molecular evolution. Combination of both techniques indeed appears the most promising approach for obtaining polypeptides with new properties. Besides genetic complementation methods, powerful screening and selection techniques have been developed in the last years for identification of proteins with new functions, such as the phage display technique, the yeast two-hybrid system, ribosome display, fragment complementation, and cell surface display on yeast or bacteria.

Besides the need for evolutionary methods for obtaining polypeptide with new functions, there is also an independent, strong need for techniques to evolve proteins towards higher thermodynamic stabilities and expression yields. However, practically none of the above mentioned screening and selection techniques can be applied for improving protein stability. The only exception is a recently developed, phage-based in vitro selection method that links thermodynamic stability with protease resistance of protein domains inserted between the domains N2 and CT of the gene-3 protein of the fd phage, where phage infectivity is lost when N2 and CT are disconnected through degradation of the inserted domain. Another method of screening for protein stability has been described recently, in which protein libraries are fused to the C-terminus of the green fluorescent protein (GFP). This method essentially screens for proteins with a reduced tendency to aggregate and form inclusion body as judged by higher amounts of soluble GFP. The system thus screens for improved expression yield, but has not yet been shown to generate proteins with increased stability.

Here, we present the first selection system for optimizing protein folding and improving protein stability in the cytoplasm of living cells. The system is based on detection of folding in vivo, independent of protein function. The method makes use of the distance-dependent fluorescence resonance energy transfer (FRET) from the
green fluorescent protein (GFP) from *Aequorea victoria* to BFP, a variant of GFP with blue-shifted spectroscopic properties. Fluorescence donor-acceptor pairs of GFP variants have successfully been used for *in vivo* measurements of conformational changes in calmodulin caused by calcium binding, and detection of protein-protein interactions\(^{17-20}\). The general feature of our selection system is a ternary fusion protein in which BFP and GFP are fused to the termini of a protein of interest (protein X), where only tertiary structure formation of protein X brings BFP and GFP in sufficient proximity for *in vivo* detection of a FRET signal, while the signal is absent when BFP and GFP are far apart due to an unfolded form or disconnection by intracellular degradation of protein X. The present screening/selection system is compatible with FRET-based bacterial cell sorting, completely independent of protein function and should therefore be generally applicable for identification of protein variants with improved thermodynamic stabilities from complex libraries.
Results

Thermodynamic stability correlates with intracellular FRET intensity.

The FRET-based in vivo screening system described here is based on the assumption that simultaneous fusion of a protein X to BFP at its N-terminus and to GFP at its C-terminus will only result in FRET from BFP to GFP when protein X adopts native tertiary structure in vivo and brings BFP and GFP in close proximity. Moreover, we expected that protein X variants with decreased thermodynamic stability should yield a lower FRET signal. This is because the fraction of unfolded molecules at equilibrium is higher in a protein with lower thermodynamic stability. Therefore, if the unfolded fraction were significantly populated (i.e., if $\Delta G_{\text{fold}}$ is close to or greater than zero) one would expect a loss of FRET intensity that is proportional to the fraction of unfolded proteins. Second, unfolded molecules, which are principally always present at equilibrium, are much more sensitive to intracellular proteolysis than native proteins. Proteolytic degradation in vivo should consequently increase with decreasing stability and should also decrease the total amount of produced protein by removing the unfolded state from the folding equilibrium. In the context of a ternary fusion protein with BFP and GFP, degradation of protein X would lead to a total loss of FRET due to disruption of the covalent linkage between BFP and GFP (Fig. 1).

To test these assumptions experimentally, we inserted four engineered immunoglobulin V L domains with different thermodynamic stabilities between BFP and GFP (Fig. 2B). We selected these engineered V L domains because they do not require the conserved disulfide bond of antibody domains for folding and can thus be produced as soluble, native proteins in the reducing environment of the cytoplasm\textsuperscript{21}. The free energies of folding of the reduced V L variants range from -13.3 to -24.4 kJ mol\textsuperscript{-1}\textsuperscript{21}. Here, we have called these V L variants V L1 to V L4 according to their increasing thermodynamic stability (V L4 is the most stable V L variant). In addition, we constructed a plasmid for expression of a direct fusion of BFP and GFP with a connecting 15 amino acid peptide linker as a positive control (termed BFP-GFP) (Fig. 2A), and a plasmid for coexpression of isolated BFP and GFP as a negative control (Fig. 1, Fig. 2C).

All constructs were expressed in the cytoplasm of E. coli under control of the lac promoter - operator system. Fig. 3A shows a direct FRET analysis of soluble
fractions of extracts from equal amounts of IPTG-induced *E. coli* cells (selective excitation of BFP at 387 nm and recording of fluorescence above 400 nm). The fluorescence data exhibited the expected increase in FRET intensity with increasing $V_L$ stability. Moreover, the direct fusion of BFP to GFP showed the highest fluorescence signal, while practically no FRET was observed when isolated BFP and GFP were coexpressed (Fig. 3A). When the same cell extracts were analyzed for the presence of GFP by selective excitation at 450 nm, identical amounts of GFP were detected, except for the coexpression construct, which yielded somewhat lower intracellular GFP concentrations (Fig. 3B). As GFP is the C-terminal fusion partner in the ternary fusion proteins, these data demonstrate that the total amount of translated ternary fusion protein per cell is the same for all four $V_L$ domains, indicating that the different FRET intensities result from different extents of intracellular proteolysis. This could be confirmed by Western Blot analysis of the extracts after SDS-PAGE separation and GFP/BFP-specific immunostaining, where the fraction of ternary fusion protein increased with increasing $V_L$ stability, and the fraction of the proteolytically degraded ternary fusion protein, evident by bands corresponding to isolated BFP and GFP, simultaneously decreased (Fig. 3C). The yields of the four different ternary fusion proteins after purification with ion exchange, hydrophobic and gel filtration chromatography varied between 2 mg per liter of bacterial culture for the $V_{L1}$ construct and 57 mg per liter for the $V_{L4}$ construct, and also completely reproduced the order of thermodynamic $V_L$ stabilities (Fig. 3D). Overall, the expression data proved that intracellular proteolysis of $V_L$ domains with limited stability is responsible for the observed correlation between stability and FRET intensity, and not a significantly populated unfolded $V_L$ fraction in the ternary fusions. This correlation is also guaranteed by the extraordinary stability of GFP and BFP against intracellular proteolysis.

The possibility that loss of FRET intensity in the ternary fusion proteins can arise due to unfolding of protein X and increased average distance between BFP and GFP could, however, also be demonstrated by selective unfolding of the $V_L$ domains in the purified, ternary fusion proteins with 6 M urea. At this denaturant concentration, BFP and GFP are still completely native, while all $V_L$ domains are unfolded. Fig. 4 shows that FRET from BFP to GFP in the ternary fusions is completely lost in 6 M urea, and fluorescence is diminished to that of a 1:1 mixture of BFP and GFP. We conclude
that loss of FRET will also occur \textit{in vivo} when the polypeptide that is inserted between BFP and GFP is unfolded, but not degraded by cellular proteases.

\textit{Application of intracellular FRET to screening and selection of bacteria producing proteins with improved stability}

To test whether the FRET-based detection of protein folding and stability can be used to identify proteins with increased thermodynamic stability at the level of single bacterial colonies and individual cells, we performed colony screening and cell sorting experiments. Fig. 5A shows a 1:1 mixture of IPTG-induced \textit{E. coli} cells either expressing the binary BFP-GFP fusion (green colonies) or coexpressing BFP and GFP (blue colonies), which can easily be discriminated on an agar plate. Visual inspection also allowed discrimination of cells producing ternary fusions with \textit{V\textsubscript{L}1} and \textit{V\textsubscript{L}4}, but proved difficult for discrimination of \textit{V\textsubscript{L}} constructs with similar stabilities, e.g. \textit{V\textsubscript{L}1} and \textit{V\textsubscript{L}2} (data not shown). We therefore investigated FRET-based bacterial cell sorting as a tool for more selective enrichment of cells with increased FRET intensity. Moreover, bacterial cell sorting allows automated screening of a much larger number of bacteria producing complex libraries of protein variants compared to single-colony screening, which has its practical limitations at more than about $10^5$ different clones. We used a FACStar\textsuperscript{PLUS} cell sorter, which is capable of sorting 3000 bacterial cells per second based on GFP fluorescence intensity of individual bacteria. BFP in the ternary fusions was selectively excited at 351 nm with an argon laser, and GFP fluorescence at 530 ± 15 nm was monitored. Fig. 5B shows that the cell sorter could discriminate well between bacteria producing different ternary fusion constructs, except for the \textit{V\textsubscript{L}1} and \textit{V\textsubscript{L}2} constructs, which show the weakest and most similar FRET intensities (see Fig. 3A). Bacterial cell sorting and definition of the fluorescence threshold for selection also proved more efficient than colony screening with respect to the time of induction with IPTG required for development of sufficient FRET intensity for \textit{in vitro} selection. Induction for 24 h at 30 °C proved sufficient for cell sorting compared to two days for reliable colony screening. We reproducibly achieved an enrichment factor of about 50 by bacterial cell sorting when we tested mixtures of cells producing BFP-VL4-GFP and BFP-VL1. We used mixing ratios between 1:1 and 1:1000, and the enrichment factor proved to be independent of the mixing ratio after counting blue and green fluorescent clones on agar plates.
Validation of the system by selection of $V_L$ variants with improved stability from an expression library

We tested the FRET-based system for identification of proteins with increased thermodynamic stability with a directed random mutagenesis of the antibody domain $V_L2$. As the two domains $V_L2$ and $V_L4$ only differ in the two amino acids at position 33 and 51 (Table I) (sequence numbering according to reference 21), we simultaneously randomized $V_L2$ in the BFP-$V_L2$-GFP construct at the corresponding positions with Kunkel mutagenesis. We then applied one round of bacterial cell sorting, followed by manual screening for bright green single colonies on agar plates. Analysis of cells before and after cell sorting showed an enrichment factor of about 50. Bright green clones were then selected and sequenced (Table I). The two brightest green clones ($V_L21$ with Ser$^{33}$ and Asp$^{51}$ and $V_L22$ with Ser$^{33}$ and Glu$^{51}$) were then chosen for biochemical characterization. The two corresponding $V_L$ genes were cloned into a cytoplasmic expression plasmid, produced as isolated proteins with a C-terminal (His)$_6$ tag in the $E. coli$ cytoplasm and purified to homogeneity. Fig. 6 shows the GdmCl-dependent equilibrium folding transition of the reduced $V_L2$ domain compared to the transitions of the selected, reduced variants $V_L21$ and $V_L22$. Both selected variants indeed proved to be about 4 kJ mol$^{-1}$ more stable than “wild type” $V_L2$, (Fig. 6, Table II), clearly demonstrating the general applicability of FRET-based screening of bacterial colonies producing complex libraries for generating protein variants with increased stability.
Discussion

The screening system for protein stability and protein folding established in this study is, to the best of our knowledge, the only method available so far for testing protein folding in the cytoplasm of living cells that is entirely independent of protein function. The only other, comparable system is phage display-based in vitro selection for protease-resistant proteins inserted into the gene 3 protein of the fd phage\textsuperscript{15}. Each method has distinctive advantages, depending on the protein to be stabilized. Our cytoplasmic screening system will presumably be the method of choice if a protein with multiple thiol groups, which is sensitive towards oxidation and misfolding through formation of wrong disulfide bonds in the oxidizing environment of the periplasm, should be stabilized. In contrast, the phage-based system would have to be used in the case of disulfide-bonded proteins, which normally cannot fold in the reducing cytoplasm. Both methods are nearly comparable in terms of the complexity of protein libraries that can be handled and the enrichment factors that are achieved per round of selection. The availability of a cytoplasmic screening system for protein stability is of general practical interest, as there is a substantial need for proteins like functional antibodies that are sufficiently stable without the conserved disulfide bridge of antibody domains and can be applied in the cytoplasm. In this context, as shown by successful stabilization of an antibody VL intradomain, the present method appears to be particularly useful.

The applicability of the present screening system for increased protein stability is strongly dependent on the distance between the natural termini in the folded protein. Unfortunately, there are presently no experimental data available for the $R_o$ value (distance between donor and acceptor where FRET-efficiency is half-maximal) of the BFP-GFP donor-acceptor couple. It is therefore difficult to judge to what extent the distance between the termini restricts general application of the method. There is, however, a theoretically predicted $R_o$ value of 41 Å for the BFP-GFP donor-acceptor couple\textsuperscript{23}. Due to the $R^{-6}$ dependence of FRET on the distance (R) between the chromophores\textsuperscript{24}, one would predict that FRET-efficiency drops below 10%, the presumed detection limit of the method, at a distance of more than 65 Å. If one takes the minimum value of about 7-8 Å for the distance between the buried chromophores and the surface of the proteins (i.e., the “top” of the β-barrel where the GFP termini are located), this would leave about 45-50 Å maximal distance between the termini of
the protein X to be stabilized. In the case of a globular protein with termini at opposite ends in the three-dimensional structure, this would mean that the method could still be applied to proteins of about 30 kDa in size. This is also about the upper size limit for one-domain proteins. In the case of larger, multidomain proteins, one most likely would encounter additional problems independent of the location of the termini, mainly due to preferred proteolytic cleavage at domain boundaries, which would immediately abolish any relationship between thermodynamic stability and protease resistance. Finally, analysis of protein structures has shown that the natural termini of proteins are, on the average, in closer proximity than one would predict from a random distribution\textsuperscript{25}. Therefore, we predict that the screening system established in this study will be suitable for most one-domain proteins smaller than 30 kDa. In principle, one also worry about the lower size limit of the protein to be stabilized since, at some point, BFP and GFP will remain within FRET distance even when protein X is unfolded. Assuming a very small protein of only 50 residues, the average distance between the termini in the unfolded state should still be 70-80 Å (as calculated from \(R_o^2)^{0.5} = (130 \cdot \text{number of residues})^{0.5}\). Thus, no significant FRET signal would be expected if the small intervening protein X were unfolded, suggesting that our system can also be applied to improve the stability of very small polypeptides.

Interestingly, analysis of ternary fusion proteins in cell extracts has shown that our screening systems works because unstable proteins are more susceptible to proteolysis than stable proteins. Moreover, direct observation of a folding equilibrium (mixtures of unfolded and native states) of the protein inserted between BFP and GFP is most likely not possible due to degradation of the unfolded protein by cytoplasmic proteases. In this context, it is important to note that rather long incubation times of induced cells producing the ternary fusion proteins are required for fluorescence and FRET development (about 24 hours). This is most likely due to slow folding of BFP and GFP, which requires a rate-limiting air oxidation step for fluorophore development\textsuperscript{27,28}. This of course gives cellular proteases plenty of time to degrade proteins with marginal stability to completion.

We assume that the present system can also be easily modulated according to the specific requirements of a given screening experiment. For instance, selective pressure for stability could be modulated \textit{in vivo}, for example by use of protease-
deficient strains, coexpression of proteases, change in growth temperature and prolonged expression times. Control is also possible \textit{in vitro}, in particular through definition of the FRET-intensity threshold during bacterial cell sorting. Overall, we believe that the FRET-based system introduced in this study will be a useful tool for engineering novel proteins with improved properties.

\textbf{Acknowledgements:}

Help during flow cytometry measurements on the FACStar\textsuperscript{PLUS} from the Institute of Biomedical Engineering of the Swiss Federal Institute of Technologie (ETH) and the University of Zürich, Switzerland, is gratefully acknowledged. This project was supported by the ETH Zürich.
Results

FRET-based in vivo screening for protein folding and stability

Experimental protocol

Materials

DE52-cellulose was purchased from Whatman (Maidstone, United Kingdom), Ni-NTA-cellulose was from Quiagen (Hilden, Germany) and Phenyl Superose HR and SourceQ was obtained from Amersham-Pharmacia (Uppsala, Sweden). Isopropyl-β-D-thiogalactoside (IPTG) and 1,4-dithio-l-D,L-threitol (DTT) were from AGS GmbH (Heidelberg, Germany), reduced glutathione (GSH) and oxidized glutathione (GSSG) were from Sigma (Deisenhofen, Germany) and lysozyme was obtained from Roche Diagnostics (Rotkreuz, Switzerland). All other chemicals were from Merck (Darmstadt, Germany) and of the highest purity available. Plasmids with the genes of BFP and GFP were obtained from Quantum (Montreal, Quebec, Canada) and oligonucleotides from Microsynth (Balgach, Switzerland). DNA-modifying enzymes were from MBI Fermentas (Vilnius, Lithuania), Roche Diagnostics (Rotkreuz, Switzerland), AGS GmbH (Heidelberg, Germany) or New England Biolabs (Beverly, MA, USA). Polyclonal rabbit antibodies raised against GFPuv (Clontech, Palo Alto, CA, USA) were obtained from Eurogentec (Seraing, Belgium). Expression plasmids for the V_L domains V_L1 to V_L421 were kindly provided by Dr. B. Steipe (Genzentrum, University of Munich, Germany).

Plasmid construction

For the production of isolated BFP and GFP the expression plasmids pBFP1 and pGFP1 were constructed. The genes derived from pQBI67-BFP and pQBI63 (Quantum) respectively were cut and cloned into the plasmid pRBI via XbaI and BamHI.

The oligonucleotides 5’-ACG AAT TTC TAG AAA TAA-3’ (XbaI restriction site underlined) and 5’-ATG GAT GAA CTA TAC AAA TCC GGA GGC TCG AGG GAT CCC CT-3’ (XhoI and BamHI sites underlined) and the plasmid pBFP1 were used for amplification of the BFP gene and the oligonucleotides 5’-TCG TCT AGA CTG AGG CGT CCC CT-3’ (XbaI, XhoI and NsiI sites underlined) and 5’-AGG GCG ATG GCC CAC T-3’ (XbaI, XhoI and NsiI sites underlined) and 5’-AGG GCG ATG GCC CAC T-3’ and the plasmid pGFP1 were used to amplify the GFP gene. The PCR fragments were cloned via XbaI and BamHI into pRBI, yielding pBFP2 and pGFP2.
These plasmids were cleaved with the restriction enzymes *XhoI* and *DraIII* and the vector fragment of pBFP2 was ligated with the GFP fragment of pGFP2, yielding the expression vector pBFP-GFP for production of the dimeric fusion protein.

For coexpression of BFP and GFP via a dicistronic operon, a first PCR with the oligonucleotides 5′-ACG AAT TTC **TAG AAA** TAA-3′ (*XbaI* site underlined) and 5′-ATG GAT GAA CTA TAC AAA TAA CTC GAG GGA TCC CCT-3′ (*XhoI* and *BamHI* sites underlined) with pBFP1 as template and a second PCR with the oligonucleotides 5′-TCG TCT AGA CTC GAG ATC ACA CAG GAA ACA GAC CAT GGC TAG CAA AGG AGA AGA ACT C-3′ (*XbaI* and *XhoI* sites underlined) and 5′-AGG GCG ATG GCC CAC T-3′ with pGFP1 as a template was performed. The PCR fragments were cloned via *XbaI* and *BamHI* into pRBI, yielding pBFP3 and pGFP3. After cleavage of the plasmids with *XhoI* and *DraIII*, the vector fragment of the pBFP3 and the GFP fragment of pGFP3 were ligated, yielding the coexpression vector pBFP+GFP.

The genes of the antibody domains V\textsubscript{L1}, V\textsubscript{L2}, V\textsubscript{L3} and V\textsubscript{L4} were amplified via PCR with the oligonucleotides 5′-CTC **CTC GAG** AGC CAT GGA TAT CGT TAT GAC CCA-3′ (*XhoI* site underlined) and 5′-ACC AAA CTG GAA CTG AAG GGA GGA TCA GGC GGA AAT GCA TAC-3′ (*NsiI* site underlined) using the vectors pTET-T\textsubscript{M}500, pTET-T\textsubscript{M}601, pTET-T\textsubscript{M}705 and pTET-T\textsubscript{M}703\textsubscript{21} as templates. The PCR fragments were cloned via *XhoI* and *NsiI* into the vector pBFP-GFP, yielding the ternary fusion expression vectors pBFP-V\textsubscript{L1}-GFP, pBFP-V\textsubscript{L2}-GFP, pBFP-V\textsubscript{L3}-GFP and pBFP-V\textsubscript{L4}-GFP. All amplified genes were verified by dideoxynucleotide sequencing.

**Random mutagenesis**

Random mutagenesis of residues 33 and 51 of V\textsubscript{L2} in the ternary fusion protein was performed according to Kunkel\textsuperscript{30} using single stranded, uridinylated DNA of the vector pBFP-V\textsubscript{L2}-GFP with the oligonucleotides 5′-CTG ATA CCA CGC CAA NNN ATT AGA GAC AGA TTG AGA AGC TCT GCA GCT GAT G-3′ and 5′-TTG GTC CAG GGT GGA CTC TCT CCT-3′.
Colony screening and bacterial cell sorting

*E. coli* JM83 cells were transformed with the VL2 mutant library from the Kunkel mutagenesis and grown on LB-agar plates containing 100 µg/ml ampicillin and 1 mM IPTG for 24 hours. For bacterial cell sorting, about 20'000 colonies were resuspended in PBS buffer (20 mM sodium phosphate, pH 7.4, 115 mM NaCl) and diluted with PBS buffer to a concentration of 3 \cdot 10^6 cells per ml. The bacterial cells were sorted on a FACStar\textsuperscript{PLUS} (Becton Dickingson, San Jose, CA, USA) cell sorter. The selection for bright green bacterial cells was performed by excitation at 351 nm and detection of fluorescence at 530 ± 15 nm using a flow rate of 1500 events/sec. After one round the resulting bacterial cells were plated on agar plates containing 100 µg/ml ampicillin and 1 mM IPTG. After 24 h at 30 °C bright green clones were selected and the mutant VL genes were sequenced. The two brightest green clones were chosen for further investigations. For production of the isolated VL variants, the fusion vectors were digested with *Sac*I and *Bgl*II, and the gene fragments of the VL domains were ligated into the expression vector pTET-Tμ601\textsuperscript{21}, which allows production of the isolated VL variants with a C-terminal (His)\textsubscript{6}-tag in the *E. coli* cytoplasm.

The enrichment factor of one round of the cell sorter was determined by sorting different mixtures of cells expressing either BFP-GFP or coexpressing isolated BFP and GFP and counting the percentage of the green colonies after selection and growth on LB/amp/IPTG-agar plates.

Recombinant protein expression and purification

Expression and purification of BFP, GFP and BFP-GFP using the vectors pBFP1, pGFP1 and pBFP-GFP was performed essentially as described previously\textsuperscript{31}. After cell growth in 2xYT medium containing 100 µg/ml ampicillin (2xYT/amp) to an OD\textsubscript{600} of 1.0 and induction with 1 mM IPTG, the cells were grown for 16 hours at 30 °C and harvested by centrifugation. Bacteria were resuspended in 20 mM Tris/HCl, pH 8.0 containing lysozyme (1 mg/ml), incubated at 30 °C for 30 min and disrupted by sonification. After centrifugation, the supernatant was incubated for 30 min at 50 °C. The heat-precipitated proteins were removed by centrifugation and the supernatant was applied to a DE52 anion exchange column equilibrated with 20 mM Tris/HCl, pH 8.0. The proteins were eluted using a linear gradient from zero to 0.5 M NaCl in
the same buffer and the elute was mixed with 4 M ammonium sulfate to a final concentration of 0.8 M ammonium sulfate. The solution was applied to a phenyl sepharose column equilibrated with the same buffer and the proteins were eluted with a linear gradient from 0.8 to zero M ammonium sulfate. After dialysis against PBS buffer the proteins were concentrated to approximately 5 ml and applied to a HiLoad 26/60 Superdex 75 gel filtration column equilibrated with PBS buffer. Fractions with pure protein were pooled and dialysed against distilled water. The same protocol was used for the production and purification of the ternary fusion proteins, except that the thermal denaturation step was omitted and an additional SourceQ anion exchange chromatography in 20 mM Tris/HCl, pH 8.0 was introduced before the gel filtration.

Cytoplasmic production, refolding and purification of isolated VL domains from cytoplasmic inclusion bodies was performed as described\(^2\). The cells were grown in 2xYT/amp medium to an OD\(_{600}\) of 0.9 and induced with 0.2 µg/ml anhydrotetracycline. After further growth at 20 °C over night, the cells were harvested by centrifugation, resuspended in 67 ml of PBS buffer per liter of culture and lysed by sonification. After centrifugation the inclusion body pellet was washed three times with 23 % (w/v) sucrose, 0.5 % (v/v) Triton X-100 in 20 mM sodium phosphate, pH 8.0. The inclusion bodies were solubilized in 6 M guanidine hydrochloride (GdmCl) and 100 mM sodium phosphate, pH 8.0 and the cysteines were kept reduced by 10 mM GSH. The solution was applied to a Ni-NTA column equilibrated with the same buffer. For on-column refolding of the VL domains a linear gradient from 6 to zero M GdmCl in 100 mM sodium phosphate, pH 8.0 was performed and on-column oxidation of VL domains was achieved by washing with five column volumes of 5 mM GSSG in 100 mM sodium phosphate, pH 8.0. After washing with 100 mM sodium phosphate, pH 8.0, the proteins were eluted by a linear gradient from zero to 300 mM imidazole in 100 mM sodium phosphate, pH 8.0. The elute was concentrated to 10 ml and applied to an HiLoad 16/60 Superdex 75 gel filtration column equilibrated with PBS buffer. The fractions containing pure, monomeric VL domains were pooled, concentrated to 10 ml and dialyzed against distilled water. The final yield varied between 5 to 10 mg of oxidized VL domains per liter of bacterial culture.
**Expression analysis of the ternary fusion proteins**

For production and degradation analysis of the ternary fusion proteins the soluble fractions of cell extracts were analyzed by fluorescence and immunostaining. JM83 *E. coli* cells were grown to an OD$_{600}$ of 1.0 in 2xYT/amp medium and induced with 1 mM IPTG. After further growth at 30 °C for 24 hours the cells were harvested by centrifugation, resuspended in 8 ml lysis buffer (20 mM sodium phosphate, pH 7.4, 115 mM NaCl, 10 mM DTT, 1 mM EDTA, 0.1% lysozyme) per gram of bacteria and lysed by sonification. After centrifugation the supernatant was diluted 1:10 with 20 mM sodium phosphate, pH 7.4, 115 mM NaCl, 10 mM DTT, 1 mM EDTA. For FRET analysis, emission spectra were recorded from 400 to 600 nm (excitation wavelength: 387 nm), and from 470 to 600 nm (excitation wavelength: 450 nm) for selective detection of soluble GFP. Samples from the soluble fraction of cell extracts were applied to an SDS-Gel and blotted onto PVDF-membrane. The GFP-specific immunostaining was performed with polyclonal rabbit anti-GFPuv3 antibodies as described$^{32}$.

**$V_L$ domain unfolding and equilibrium transitions**

Denaturant-induced unfolding of the $V_L$ domains in the context of the ternary fusion protein was measured at 25 °C and a protein concentration of 1 μM in 20 mM sodium phosphate, pH 7.4, 115 mM NaCl, 10 mM DTT, 1 mM EDTA, containing zero or 6 M urea. Unfolding of $V_L$ was detected by the decrease of FRET intensity between 400 and 600 nm using an excitation wavelength of 387 nm. Control experiments verified that BFP and GFP are not unfolded in the presence of 6 M urea.

Refolding transitions of the reduced $V_L$ domains were measured fluorimetrically at 20 °C and pH 7.4. Proteins were unfolded in 5 M GdmCl, 10 mM HEPES/NaOH, pH 7.4, 5 mM DTT and diluted into 10 mM HEPES/NaOH, pH 7.4, 2.5 mM DTT, 1 mM EDTA, containing zero to 2.5 M GdmCl (final protein concentration: 3.2 μM). Refolding was measured by the fluorescence increase at 320 nm (excitation wavelength: 280 nm) as described$^{33}$ and the data were fitted according to a two-state transition$^{34}$. 
Figure legends

**Figure 1.** Concept of in vivo screening for protein folding and stability: The system is based on a ternary fusion in which a protein of interest (X) is fused to the green fluorescent protein (GFP) at its C-terminus and BFP, a variant GFP with blue-shifted spectroscopic properties, at its N-terminus. Due to the strong dependence of FRET on the distance between fluorescence donor and acceptor, FRET from BFP to GFP is only observed when both proteins are in close proximity due to tertiary structure formation of protein X (A). In contrast, FRET from BFP to GFP is lost when protein X is unfolded (B) or degraded by cellular proteases (C). Direct fusion of BFP with GFP with a short linker (see Fig. 2) was used as positive control (D), and coexpression of isolated BFP and GFP with a dicistronic operon was used as negative control (E).

**Figure 2.** Genetic constructs used for expression of BFP-GFP fusion proteins. (A) Segment of the genetic sequence of the construct BFP-GFP, containing a flexible linker sequence with a thrombin cleavage site and single XhoI and NsiI sites for subsequent insertion of antibody VL genes. (B) Sequences of the flexible linkers used to connect different antibody VL domains (VL1-VL4) to BFP and GFP. The methionine residue (asterisk) at the N-terminus of the VL domain is not contained in the natural antibody domain, but was previously shown to stabilize VL domain and was therefore included in the construct. (C) Section of the sequence of the plasmid for coexpression of BFP and GFP via a dicistronic operon und lac operator / promoter control. The region between the BFP and GFP gene with the second ribosomal binding site (RBS 2) for GFP translation is shown. The genetic sequence between the XhoI site and the ATG start codon of GFP is from the strong expression vector pTrc99a (accession number U13872).

**Figure 3.** Fluorescence and Western blot analysis demonstrating the inverse correlation between thermodynamic stability of VL domains and proteolytic degradation in the E. coli cytoplasm. (A) Soluble fractions of equal amounts of extracts of E. coli JM83 cells producing the respective ternary fusion proteins in the cytoplasm were diluted with buffer (pH 7.4), and FRET efficiency between BFP and GFP was measured by recording of GFP fluorescence spectra after excitation of BFP at 387 nm. (B) Same samples as in (A), but selective excitation of GFP in cell
extracts. The equal concentrations of GFP in all extracts demonstrate identical total fusion protein production, but different proteolytic processing of the $V_L$ domains in the ternary fusion proteins. (C) GFP and BFP-specific immunostaining with polyclonal rabbit anti-GFP antibodies of Western blots after separation of equal amounts of soluble fractions of cell extracts by SDS-PAGE. The antibodies equally recognize BFP and GFP. (D) Yields of ternary fusion proteins per liter of bacterial culture after purification from soluble fractions of cell extracts. (C) and (D): (1) BFP-GFP: direct fusion of BFP and GFP; (2) coexpression of isolated BFP and GFP; (3)-(6) BFP-$V_L$1-GFP to BFP-$V_L$4-GFP: ternary fusion proteins containing antibody $V_L$ domains with increasing thermodynamic stability.

Figure 4. Loss of FRET from BFP to GFP upon selective unfolding of the $V_L$ domains in purified, ternary fusion proteins with 6 M urea. Fusion proteins were excited at the absorbance wavelength of BFP (387 nm), and fluorescence spectra were recorded at 25 °C between 400 and 600 nm. Solid lines: proteins (1 $\mu$M) at pH 7.4; dashed lines: proteins (1 $\mu$M) in 6 M urea, pH 7.4.

Figure 5. Screening and *in vitro* selection of *E. coli* cells producing different BFP-$V_L$-GFP fusion proteins. (A) Screening for green fluorescent, single colonies on agar plates. As an example, a 1:1 mixture of bacteria producing BFP-GFP (green) or coexpressing isolated BFP and GFP (blue) is shown (excitation at 350 nm). (B) Analysis of *E. coli* cultures producing the domains $V_L$1 to $V_L$4 as ternary fusion proteins with BFP and GFP with a FACStarPLUS cell sorter. Different FRET intensities due to different degrees of intracellular $V_L$ degradation can be well separated and enriched. Cells producing BFP-GFP and coexpressing BFP and GFP are shown as positive and negative controls, respectively. (C) Efficient discrimination by cell sorting between cells expressing BFP-$V_L$2-GFP and cells expressing the selected, stabilized variants BFP-$V_L$21-GFP or BFP-$V_L$22-GFP. The vertical line at 225 fluorescence units represents the fluorescence threshold that was defined for selection of $V_L$21 and $V_L$22 after random mutagenesis of $V_L$2. The plot shows cellular FRET intensity distributions of bacterial clones expressing BFP-$V_L$2-GFP, BFP-$V_L$21-GFP or BFP-$V_L$22-GFP, and 1:1 mixtures of cells expressing BFP-$V_L$2-GFP and BFP-$V_L$21-GFP,
or BFP-V\textsubscript{L}2-GFP and BFP-V\textsubscript{L}22-GFP. Cells were excited at 351 nm, and fluorescence intensity at 530 ± 15 nm was recorded.

**Figure 6.** Normalized equilibrium transitions of the reduced, isolated V\textsubscript{L} domains V\textsubscript{L}21 and V\textsubscript{L}22, which were selected after site-directed random mutagenesis of the domain V\textsubscript{L}2. Transitions were measured in 10 mM HEPES/NaOH, pH 7.4, 2.5 mM DTT, 1 mM EDTA with zero to 2.5 M GdmCl, and followed by the increase in fluorescence intensity at 320 nm upon folding (excitation at 280 nm, protein concentrations: 3.2 \mu M). Data were fitted according to a two-state equilibrium (solid lines).
References


17. Romoser, V.A., Hinkle, P.M. & Persechini, A. Detection in living cells of Ca2+-dependent changes in the fluorescence emission of an indicator composed of


Table I: DNA and amino acid sequences of the resulting \( \text{V}_L \) variants after random mutagenesis

<table>
<thead>
<tr>
<th>Clon-No.</th>
<th>amino acid 33</th>
<th>amino acid 51</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA</td>
<td>AS</td>
</tr>
<tr>
<td>BVL2G</td>
<td>TAC</td>
<td>Tyr</td>
</tr>
<tr>
<td>BVL21G</td>
<td>TCC</td>
<td>Ser</td>
</tr>
<tr>
<td>BVL21aG</td>
<td>TCC</td>
<td>Ser</td>
</tr>
<tr>
<td>BVL22G</td>
<td>TCC</td>
<td>Ser</td>
</tr>
<tr>
<td>BVL4G</td>
<td>CAC</td>
<td>His</td>
</tr>
</tbody>
</table>
Table II: Thermodynamic stability of the resulting $V_L$ domains

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\Delta G$ (kJ/mol)</th>
<th>$m$ (kJ/mol M)</th>
<th>$\Delta \Delta G$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_L 2$</td>
<td>-10.5 (± 1.8)</td>
<td>16.8 (± 3.1)</td>
<td>-</td>
</tr>
<tr>
<td>$V_L 21$</td>
<td>-14.0 (± 0.6)</td>
<td>16.7 (± 0.8)</td>
<td>-3.5</td>
</tr>
<tr>
<td>$V_L 22$</td>
<td>-14.5 (± 1.2)</td>
<td>18.7 (± 1.7)</td>
<td>-4.0</td>
</tr>
</tbody>
</table>
Results

FRET-based in vivo screening for protein folding and stability

Figure 1

A

B

C

D

E

387 nm

X

GFP

509 nm

387 nm

BFP

GFP

387 nm

BFP

GFP

387 nm

BFP

GFP

387 nm

BFP

GFP

387 nm

BFP

GFP
Figure 2

A

\[
\begin{align*}
\text{BFP} & \quad \text{linker sequence} & \quad \text{GFP} \\
\text{XhoI} & \quad \text{NsI} \\
\text{---GTGAACTATACAAAATCCGAGGCTGAGTTCCCGAGGATCCAGGATCTGACAAAAAGGAGAA---} \\
\text{---CTACTGATCTTTTATGGCCTTCGGAGGCTGAGTTCCCGAGGATCCAGGATCTGACAAAAAGGAGAA---} \\
\text{---DELYK SGGSSVPRSGTGTNAASKGE---} \\
\end{align*}
\]

B

\[
\begin{align*}
\text{BFP} & \quad \text{linker 1} & \quad \text{V\textsubscript{\text{e}} domain} & \quad \text{linker 2} & \quad \text{GFP} \\
\text{XhoI} & \quad \text{NsI} \\
\text{---GTGAACTATACAAAATCCGAGGCTGAGTTCCCGAGGATCCAGGATCTGACAAAAAGGAGAA---} \\
\text{---CTACTGATCTTTTATGGCCTTCGGAGGCTGAGTTCCCGAGGATCCAGGATCTGACAAAAAGGAGAA---} \\
\text{---DELYK SGSRAMDIVMITKTELKGSGGNNASKGE---} \\
\end{align*}
\]

C

\[
\begin{align*}
\text{BFP} & \quad \text{RBS2} & \quad \text{GFP} \\
\text{XhoI} \\
\text{---GTGAACTATACAAAATCCGAGGCTGAGTTCCCGAGGATCCAGGATCTGACAAAAAGGAGAA---} \\
\text{---CTACTGATCTTTTATGGCCTTCGGAGGCTGAGTTCCCGAGGATCCAGGATCTGACAAAAAGGAGAA---} \\
\text{---DELYK Stop MASKGGE---} \\
\end{align*}
\]
Results

FRET-based in vivo screening for protein folding and stability

Figure 3

A

B

C

D

Fluorescence intensity (arbitrary units)

Wavelength (nm)

A

BFP-GFP
BFP-VL4-GFP
BFP-VL3-GFP
BFP-VL2-GFP
BFP-VL1-GFP

Coexpression of isolated BFP and GFP

Fluorescence intensity (arbitrary units)

Wavelength (nm)

B

BFP-GFP
BFP-VL4-GFP
BFP-VL3-GFP
BFP-VL2-GFP
BFP-VL1-GFP

Coexpression of isolated BFP and GFP

Fluorescence intensity (arbitrary units)

Wavelength (nm)

C

D

Purified protein (mg/liter of culture)

BFP-GFP

BFP + GFP

BFP + GFP

BFP-GFP

BFP-VL4-GFP

BFP-VL3-GFP

BFP-VL2-GFP

BFP-VL1-GFP

Coexpression of isolated BFP and GFP

Purified protein (mg/liter of culture)

3 4 5 6

2.2 mg/l
4.1 mg/l
46.7 mg/l
57.8 mg/l
Figure 4

![Graph showing relative fluorescence vs. wavelength for different constructs.]
Figure 5

A

B

Fluorescence intensity (arbitrary units)

Number of counts

Coexpression of isolated BFP and GFP

BFP-V_1-1-GFP
BFP-V_2-GFP
BFP-V_4-GFP
BFP-V_3-GFP
BFP-GFP

C

Fluorescence intensity (arbitrary units)

Number of counts

BFP-V_21-GFP
BFP-V_22-GFP
BFP-V_21-GFP + BFP-V_22-GFP
BFP-V_22-GFP + BFP-V_2-GFP
BFP-V_2-GFP
Figure 6

Fraction of folded protein vs. [GdmCl] (M) for proteins V\textsubscript{L} 2, V\textsubscript{L} 21, and V\textsubscript{L} 22.
5.2 Randomization of the entire active-site helix α1 of the thiol-disulfide oxidoreductase DsbA from Escherichia coli

B. Philipps & R. Glockshuber
Randomization of the entire active-site helix $\alpha_1$ of the thiol-disulfide oxidoreductase DsbA from Escherichia coli

Björn Philipps and Rudi Glockshuber

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Hönggerberg, CH-8093 Zürich, Switzerland

To whom correspondence should be addressed:

Rudi Glockshuber
Institut für Molekularbiologie und Biophysik
Eidgenössische Technische Hochschule Hönggerberg
CH-8093 Zürich
Switzerland
Phone: +41-1-633 6819
Fax: +41-1-633 1036
E-mail: rudi@mol.biol.ethz.ch
Abstract

DsbA from the periplasm of *Escherichia coli* is the most oxidizing member of the thiol-disulfide oxidoreductase family (\(E_o = -122\) mV) and required for efficient disulfide bond formation in the bacterial periplasm. The reactivity of the catalytic disulfide bond (Cys\(^{30}\)-Pro\(^{31}\)-His\(^{32}\)-Cys\(^{33}\)) is primarily due to an extremely low pK\(_a\) value (3.4) of the nucleophilic active-site cysteine Cys\(^{30}\). The Cys\(^{30}\) thiolate is assumed to be stabilized to a large extent by the partial positive dipole charge of the active-site helix \(\alpha_1\), comprising residues 30-37. Previous mutagenesis studies on the dipeptide Pro-His between the active site cysteines had shown that the dipeptide sequence strongly influences the redox properties of the enzyme. Here, we have extended these random mutagenesis studies to the complete active-site helix such that all non-cysteine residues of the helix were randomized (amino acids 31, 32, and 34-37). We find that about two thirds of the resulting variants complement DsbA deficiency in a *dsbA* deletion strain. Sequencing of 98 variants revealed a large number of non-conservative replacements in active variants, even at well-conserved positions. This indicates that tertiary structure context strongly determines \(\alpha\)-helical secondary structure formation of the randomized sequence. A subset of active and inactive variants was further characterized. All purified variants were more reducing than wild type DsbA, but the redox potentials of biologically active variants did not drop below -210 mV. All inactive variants had redox potentials lower than -210 mV, although some of the inactive proteins could still be reoxidized efficiently by DsbB. This demonstrates that efficient oxidation of substrate polypeptides is the crucial property of DsbA *in vivo*. 
**Introduction**

The thiol-disulfide oxidoreductase DsbA from *Escherichia coli* is a monomeric protein of 189 amino acids, which is required for efficient disulfide bond formation in the bacterial periplasm (Bardwell *et al*., 1991; Kamitani *et al*., 1992). The enzyme contains a catalytic disulfide bridge, which is located at the N-terminus of the active-site helix α1 with the sequence Cys$_{30}$-Pro$_{31}$-His$_{32}$-Cys$_{33}$-Tyr$_{34}$-Gln$_{35}$-Phe$_{36}$-Glu$_{37}$. DsbA-mediated disulfide bond formation in the periplasm involves oxidation of reduced, newly translocated substrate polypeptides by DsbA, followed by reoxidation of DsbA by the inner membrane protein DsbB. DsbB is in turn reoxidized by molecular oxygen through ubiquinone and terminal cytochrome oxidases (Bader *et al*., 1999; Bader *et al*., 1998).

The three-dimensional structure of both oxidized and reduced DsbA has been solved by X-ray crystallography and NMR spectroscopy (Guddat *et al*., 1998; Martin *et al*., 1993; Schirra *et al*., 1998). The enzyme possesses a thioredoxin fold, which is common to most thiol/disulfide oxidoreductases and contains the active-site disulfide (Martin, 1995). In addition to the thioredoxin domain, DsbA possesses a second, purely α-helical domain, which is inserted into the thioredoxin domain (Martin *et al*., 1993). The structural differences between oxidized and reduced DsbA are locally restricted to the active-site region, and there is no major domain motion related to the redox state of DsbA (Guddat *et al*., 1998; Schirra *et al*., 1998).

Biophysical studies on DsbA have shown that the protein is the strongest known oxidant within the disulfide oxidoreductase family (Wunderlich & Glockshuber, 1993) and oxidizes thiol compounds extremely rapidly (Darby *et al*., 1994; Wunderlich *et al*., 1993b; Zapun *et al*., 1993; Zapun *et al*., 1994). The pK$_a$ of the nucleophilic, active-site thiol of Cys$_{30}$ has an extremely low value of 3.4 (compared to 9 to 10 for normal cysteine side chains) (Nelson & Creighton, 1994). The low pK$_a$ can, at least qualitatively, explain the high redox potential of the enzyme, as well as the fact that the reduced state of DsbA is thermodynamically more stable than the oxidized state (Wunderlich *et al*., 1993a; Zapun *et al*., 1993). The X-ray and NMR structures of reduced DsbA indicate that it is mainly the partial positive charge from the dipole of the active-site helix α1, which stabilizes the thiolate anion of Cys$_{30}$, which is the most N-terminal residue of α-helix 1. Moreover, His$_{32}$ contributes about one pK$_a$ unit to the lowered pK$_a$ of Cys$_{30}$ (Hol, 1985; Kortemme & Creighton, 1995; Nelson & Creighton,
Results

Randomization of the active-site helix of DsbA

1994), and 3 possible hydrogen bonds from the thiolate to the thiol of Cys$^{33}$, and/or the amine hydrogens of His$^{32}$ and Cys$^{33}$ further stabilize the Cys$^{30}$ thiolate (Guddat et al., 1998).

The redox potentials of the different members of the thiol/disulfide oxidoreductase family vary over a wide range, from -121 mV for the most oxidizing member DsbA to -270 mV for thioredoxin, the most reducing member (Krause et al., 1991; Lin & Kim, 1989; Wunderlich & Glockshuber, 1993; Zapun et al., 1993). Previous mutagenesis studies have shown that the Xaa-Xaa dipeptide sequence between the active-site cysteines is a critical determinant of the redox properties of thiol-disulfide oxidoreductases. The first experiment in this direction was the replacement of the Gly-Pro dipeptide in the *E. coli* thioredoxin with that of eukaryotic protein disulfide isomerase (PDI) (Gly-His). The mutation shifted the redox properties of thioredoxin by 35 mV towards the redox potential of PDI (Krause et al., 1991). Analogous results were reported for the catalytic a-domain of PDI, in which the reverse exchange of the dipeptide Gly-His against the thioredoxin dipeptide Gly-Pro leads to a decrease in redox potential by 66 mV (Kortemme et al., 1996). For thioredoxin variants carrying the dipeptide Gly-His (like PDI), Pro-His (like DsbA) and Pro-Tyr (like glutaredoxin), the redox potential increases by 49 mV, 66 mV and 75 mV (Mössner et al., 1998), and lowered redox potentials were obtained for analogous dipeptide variants of DsbA (Huber-Wunderlich & Glockshuber, 1998). Random mutagenesis of the dipeptide in DsbA, thioredoxin or PDI, combined with functional screening or selection methods has also clearly demonstrated the functional importance of the Xaa-Xaa dipeptide (Chivers et al., 1996; Grauschopf et al., 1995; Holst et al., 1997; Huber-Wunderlich & Glockshuber, 1998).

In the present study, we have extended previous randomization experiments on the active-site region of DsbA to gain insights into the sequence requirements for the entire active-site helix of the enzyme. To this end, we have simultaneously randomized all six non-cysteine residues of the active-site helix α1, i.e., residues 31, 32 and 34-36 (Fig. 1), and screened overproducing bacterial clones for biologically active and inactive variants. From the pool of 98 different sequences (Table I), nine different variants were chosen for further *in vitro* characterization (Table II). The redox potentials, pK$_a$ values of Cys$^{30}$, reactivities towards the reduced polypeptide substrate hirudin and the reoxidation by DsbB were measured for all variants and compared with the properties of wild type DsbA.
Results

Active-site helix randomization and screening for active and inactive variants

Fig. 1A shows the active-site helix $\alpha_1$ of DsbA comprising residues 30-37 in the context of the tertiary structure of the oxidized protein. In both oxidized and reduced DsbA, the first turn of the active-site helix is a $3_{10}$ helix, which then continues as a regular $\alpha$-helix to residue 37 (Guddat et al., 1998; Martin et al., 1993; Schirra et al., 1998). The $3_{10}$-helical segment contains Pro$^{31}$, which is invariant in the DsbA protein family (Fig. 1B). Besides the conserved His$^{32}$, the residues in the C-terminal position of the helix are also rather well conserved. Residue 34 is usually either Tyr or Ala, residues 35 and 37 are always polar or charged, and residue 36 is exclusively a hydrophobic or aromatic amino acid (Fig. 1B).

To explore the available sequence space for the active-site helix, we randomized all-non-cysteine residues of segment 30-37 by Kunkel mutagenesis and analysed the resulting variants for biological activity in an E. coli dsbA deletion mutant. For this purpose, we used two independent DsbA complementation assays based on i) oxidative inactivation of periplasmically oriented $\beta$-galactosidase through active DsbA (Bardwell et al., 1991; Grauschopf et al., 1995) and ii) DsbA-dependent flagellar assembly, which results in immotile cells in the absence of active DsbA (Dailey & Berg, 1993).

As we intended to look for both active and inactive DsbA variants, randomization of segment 30-37 was performed both starting from the wild type background and from a catalytically inactive variant in which both active-site cysteines were replaced by alanine residues. This allowed quantification of the mutagenesis yield and determination of the fraction of active and inactive variants after randomization (see Materials and Methods). Despite the quite well conserved sequence of the active-site helix $\alpha_1$, 70% of all mutant proteins with randomized residues 31, 32, and 34-37 proved to be biologically active in complementation assays. On the basis of this result, we decided to characterize 98 mutant proteins further with respect to primary structure and periplasmic expression levels (assayed by immunoblots after SDS-PAGE of periplasmic extracts). Based on the identified fraction of active and inactive variants, we selected 69 variants, which were active in both DsbA complementation assays (variants 1-69), and 25 variants that proved inactive in both assays (variants 74-98) (Table I). Moreover, we identified a third class of DsbA variants (variants 70-
which we termed semi-active. These variants did not reconstitute motility of the dsbA deletion strain, but were nevertheless active according to the β-galactosidase assay (Table I).

Previous studies on circularly permutated DsbA variants had shown a correlation between thermodynamic stability of DsbA variants and periplasmic expression levels, most likely caused by a higher degree of proteolysis of less stable proteins by periplasmic proteases (Hennecke et al., 1999). Analysis of the expression levels of the 69 active variants showed that about two thirds of the proteins were still produced at levels similar to wild type DsbA (Table I). In the case of the inactive variants, the majority of mutant proteins (variants 82-98) were obtained at lower levels, albeit 8 inactive variants were still produced at wild type levels. Overall, the periplasmic expression levels qualitatively indicate that many of the active variants have stabilities similar to that of the wild type protein, while most of the inactive variants are thermodynamically less stable (Table I).

Sequence analysis of active and inactive variants

Table I shows the active-site helix sequences of all 98 DsbA variants selected for this study. In the case of the 69 active variants, the most unexpected result is the fact that practically any residue except proline is tolerated at positions 34-37, although the segment is well conserved. The same holds true for position 32, where proline is additionally tolerated. These data indicate that tertiary structural context strongly controls folding of helix α1 to a biologically active conformation, independent of helix propensity, degree of solvent accessibility of individual residues, and size and charge of the residues introduced by mutagenesis. Most amino acids also appear to be tolerated at position 31. However, we never found arginine or lysine residues at this position. In contrast, basic residues were specifically enriched at position 31 in the set of inactive variants. Interestingly, all semi-active variants investigated had a basic residue at position 31 as well (Table I). The variability of the dipeptide at positions 31 and 32 in the active variants is in accordance with previous random mutagenesis studies on these two residues (Grauschopf et al., 1995).

Eighteen out of the 25 inactive DsbA variants had one or two proline residues within segment 34-37 (Table I). Here, incorporation of the α-helix breaker proline is certainly the most prominent mutation leading to a inactive (and possibly folding-incompetent)
enzymes. Nevertheless, two variants (20 and 47) from the set of active DsbA proteins also contained prolines in segment 34-37. Consequently, even larger conformational changes caused by prolines in this segment can be accommodated by the DsbA protein, and do not necessarily lead to loss of biological activity.

**Spectroscopic characterization of selected DsbA variants**

For further characterization of the mutant proteins, we selected well-expressed members of each group, i.e., active variants 1-3, semi-active variants 70 and 71, and inactive variants 82-85 (Tables I and II). All nine variants were produced as soluble proteins in the cytoplasm, which gives much higher yields of DsbA than periplasmic production (Hennecke *et al.*, 1999). The variants were purified to homogeneity by conventional chromatography (see Materials and Methods).

The far-UV circular dichroism (CD) spectra of the oxidized and reduced variants proved to be extremely similar to the spectra of oxidized and reduced wild type DsbA (Fig. 2). Consequently, the overall fold of the variants, even of the inactive variants, is essentially unchanged. The overall wild type-like tertiary structure of the mutant proteins was also confirmed by fluorescence spectroscopy. Specifically, the strong tryptophan fluorescence increase of DsbA upon reduction of the catalytic disulfide bond (Wunderlich *et al.*, 1993a; Zapun *et al.*, 1993) was retained in all variants. This is a good indication for the intactness of the tertiary structure, as fluorescence quenching by the catalytic disulfide does not occur via direct contact between a tryptophan and the disulfide, but through a complex through-space energy transfer mechanism that requires the correct relative orientations and local environments of the two tryptophan residues in the protein (Hennecke *et al.*, 1997a; Sillen *et al.*, 1999). Overall, both the CD and fluorescence data indicate that even the biologically inactive DsbA variants 82-85 have wild type-like tertiary structures.

**Redox potentials of the variants and pKₐ values of Cys³⁰**

There is a theoretical and experimentally established correlation between the redox potentials of disulfide oxidoreductases, the pKₐ of their active-site cysteine, and the difference in thermodynamic stability between the oxidized and reduced state of the enzymes (Grauschopf *et al.*, 1995; Huber-Wunderlich & Glockshuber, 1998; Lin & Kim, 1989; Mössner *et al.*, 1998; Mössner *et al.*, 2000; Nelson & Creighton, 1994;
Results

Randomization of the active-site helix of DsbA

Szajewski & Whitesites, 1980). Moreover, the redox potential of the catalyst of disulfide bond formation in the bacterial periplasm and the rate of substrate oxidation have proved to be critical factors for the function of DsbA in vivo (Jonda et al., 1999). We first determined the intrinsic redox potentials of the purified variants at pH 7.0 and 25°C by measuring their equilibrium constants with glutathione (Fig. 3A, Table II). As observed for practically all DsbA variants investigated so far (Grauschopf et al., 1995; Hennecke et al., 1999; Hennecke et al., 1997b; Huber-Wunderlich & Glockshuber, 1998; Jacobi et al., 1997), all nine active-site helix variants were more reducing than wild type DsbA, with redox potentials between -165 and -222 mV. For comparison, a value of -123 mV was determined for the wild type protein (Table II). In contrast, all biologically inactive variants had redox potentials below -210 mV. This is in good agreement with DsbA complementation studies with periplasmically produced thioredoxin variants, where only thioredoxins with redox potentials above -220 mV could complement the DsbA deficiency (Jonda et al., 1999).

We also determined the pK$_a$ values of the nucleophilic, active-site cysteine (Cys$_{30}$) thiols for all variants, using the loss of thiolate anion absorbance at 240 nm upon protonation (Mössner et al., 1998; Nelson & Creighton, 1994). As expected from the lowered redox potentials of the mutant proteins compared to the wild type, all variants showed increased pK$_a$ values for Cys$_{30}$ (Fig. 3B, Table II), ranging between 4.9 and 6.6 (compared to 3.3 for wild type DsbA). The Cys$_{30}$ pK$_a$ values are thus still at least 3 pK$_a$ units below the pK$_a$ of a normal cysteine (9.5), again indicating that all variants have an intact tertiary structure. However, the observed Cys$_{30}$ pK$_a$ values of the variants correlate only weakly with the corresponding equilibrium constants with glutathione, with the variants 1, 70 and 82 deviating most strongly from theory (Fig. 3B). A possible reason for this weak correlation are electrostatic effects on the pK$_a$ value of Cys$_{30}$ at pH 7.0 resulting from the charged residues during mutagenesis of helix $\alpha$1 (cf. Table I).

Stoichiometric oxidation of the substrate polypeptide hirudin

To probe the oxidative capabilities of the purified active-site helix variants of DsbA towards polypeptide substrates, we investigated their ability to oxidize the reduced thrombin inhibitor hirudin (65 amino acids, 3 disulfide bonds in the native state), a well-established DsbA substrate (Wunderlich et al., 1993b). Reduced hirudin was
mixed with 3 molar equivalents of the oxidized DsbA variants at 25 °C at pH 7.0, reactions were acid-quenched after different incubation times, folding intermediates were separated by reversed-phase HPLC, and the half-lives for refolding were estimated from the HPLC profiles (Hennecke et al., 1999; Otto & Seckler, 1991; Wunderlich et al., 1993b) (Fig. 4, Table II). In the case of wild type DsbA, this reaction is characterized by extremely rapid, random oxidation of hirudin, followed by slow isomerization of non-native disulfide bonds catalyzed by reduced DsbA (Hennecke et al., 1999; Otto & Seckler, 1991; Wunderlich et al., 1993b). All active and semi-active variants yielded refolding rates between 10 and 90% of wild type activity and quantitative refolding; and the active variant 1 and the semi-active variant 70 proved to be even more efficient than DsbA wild type (Fig. 4) (see Discussion for details). In contrast, the inactive variants 83-85 were not capable of oxidizing hirudin efficiently, and the inactive variant 82 showed only about 10% of wild type activity (Table II, Fig. 4). These data suggest that one of the reasons for the lack of biological activity of variants 82-84 is inefficient oxidation of polypeptide substrates.

Interactions between the active-site variants of DsbA and DsbB

Finally, we investigated the ability of the purified active-site helix variants of DsbA to be reoxidized by DsbB, using DsbB-containing membrane fractions of *E. coli* for catalysis of air oxidation of the reduced variants. Reactions were followed by the decrease in DsbA fluorescence, as established (Bader et al., 1998). All active and semi-active DsbA variants had DsbB substrate properties similar to wild type DsbA, with $k_{cat}$ and $K_M$ values within a factor of 2 of the wild type values (Fig. 5, Table II). The only exceptions were active variant 1 and semi-active variant 70, which showed 5.4 and 3.0-fold higher $k_{cat}$ values than wild type DsbA (Fig. 1, Table II). Among the inactive proteins, variants 84 and 85 were reoxidized very slowly by DsbB such that we could not determine reliable catalytic parameters under the applied conditions. In contrast, inactive variants 82 and 83 were oxidized with wild type efficiency (Table II, Fig. 5). Thus, insufficient substrate oxidation activity is the probable reason for the lack of biological activity of variants 82 and 83, while both low oxidative power and slow reoxidation by DsbB cause the lack of activity in the case of variants 84 and 85.
Discussion

Numerous mutagenesis studies in the last few years have demonstrated that proteins can be very tolerant towards multiple amino acid replacements, and that tertiary structure context plays a major role in defining the eventual fold of a polypeptide chain (DeGrado et al., 1999; Gustafsson et al., 2001; Minor & Kim, 1996). For example, randomization of all amino acids from the hydrophobic core of barnase leads to a surprisingly high tolerance towards substitutions (Axe et al., 1996). Moreover, extensive site-directed substitutions of amino acids in T4 lysozyme suggest that about half of the natural primary structure would be sufficient for defining the three-dimensional structure of the protein (Matthews, 1996). In the case of the small SH3 domain, it was even possible to reduce the amino acid alphabet to only five different amino acids (Riddle et al., 1997). Nevertheless, the fact that about 70% of all variants of DsbA obtained after complete randomization of six residues from the active-site helix α1 retained native tertiary structure and biological activity is remarkable and unexpected. In particular, because helix α1 is not only a well-conserved element of regular secondary structure in the DsbA family, but also a very critical element for catalytic activity of the enzyme, as the dipole of helix α1 is assumed to play a crucial role in maintaining the low pKa value of the nucleophilic, active-site cysteine. The low pKa of Cys30 can indeed be considered the key factor underlying the enormous oxidative force of the enzyme and the extreme reactivity of its active-site disulfide bond (Grauschopf et al., 1995; Nelson & Creighton, 1994). Although we cannot make conclusive statements about the exact local structures of the randomized segments in the mutant proteins from our present data, all spectroscopic techniques applied so far, as well as the abnormally low pKa values of Cys30 observed for all mutant proteins, indicate that even the biologically inactive DsbA variants obtained after randomization of α1 have essentially wild type-like tertiary structures. Altogether, these results demonstrate that tertiary structure context almost alone determines folding of segment 30-37 to a conformation guaranteeing a functional enzyme. This is in agreement with the view that tertiary structure context is more important for folding of polypeptide segments within a protein than, for example, secondary structure propensities or even steric factors (Minor & Kim, 1996). Our data also provide further evidence for the extremely degenerate code for protein folding, i.e., that a large number of sequences fulfill the requirements for the same tertiary structure.
The high sequence variability of the active-site helix of DsbA is also in agreement with results from a systematic circular permutation study on DsbA, which demonstrated that the α-helical domain of DsbA, but not its catalytic thioredoxin domain, is critical for folding and stability and the most stable part of the enzyme (Hennecke et al., 1999). The α-helical domain of DsbA, which is inserted into the thioredoxin motif, could thus have provided a scaffold for the evolving high oxidative activity of DsbA’s thioredoxin domain by intramolecular stabilization. Our present data are also in agreement with previous mutagenesis studies on the DsbA segment 38-40, which connects α1 to the second part of the active-site helix α1’ (residues 41-50). Deletion of residues 38-40, as well as other mutations that eliminate all charged residues in the vicinity of the catalytic disulfide bond, had practically no effect on the functional properties of DsbA (Hennecke et al., 1997b; Jacobi et al., 1997).

The measured redox potentials of the active-site helix variants generated in this study also provide interesting information on the functional requirements for the catalyst of disulfide bond formation in the bacterial periplasm. Together with previous data obtained for periplasmically produced thioredoxin variants, our data suggest that the redox potential of a biologically active catalyst that can complement DsbA deficiency in vivo must be above -210 mV. In the case of DsbA, it appears that the general consequence of a lowered redox potential in DsbA variants is less efficient oxidation of polypeptide substrates. In this context, active variant 1 is particularly interesting, as its redox potential is already at the threshold value of -210 mV. The low redox potential of variant 1 may be compensated by its fast reoxidation through DsbB, which occurs 5.4-fold faster at DsbB saturation than wild type DsbA (Fig. 5, Table II). Rationally engineered, more reducing analogs of variant 1 could therefore be used to search for biologically active DsbA variants with even more reducing redox potentials.

The fact that some of the DsbA variants (variants 1 and 70) appeared to be more efficient oxidants in the hirudin oxidation and refolding experiments than the wild type, despite their lower redox potentials (Fig. 4), should not be misinterpreted. One of the main disadvantages of DsbA is its extremely rapid, random oxidation of substrate proteins with multiple cysteine residues. This leads initially to a high fraction of fully oxidized but scrambled molecules in the case of the substrate hirudin (Wunderlich et al., 1993b). As scrambled, oxidized molecules cannot isomerize spontaneously to the native conformation without reduction of at least one of the
disulfide bonds, isomerization of nonnative disulfide bonds is rate limiting for hirudin folding after oxidation with wild type DsbA. A slightly slower oxidation of hirudin, allowing intramolecular disulfide rearrangements of partially oxidized species, would thus lead to higher apparent folding rates. This appears to be exactly the case for the DsbA variants 1 and 70 (Fig. 4), which cause a slower disappearance of the HPLC peak corresponding to fully reduced hirudin but faster formation of native hirudin than wild type DsbA (Fig. 4).

Another interesting aspect of the present study with regard to the *in vivo* function of DsbA is identification of the semi-active variants 70-73 (Tables I and II), which all successfully inactivated periplasmically oriented β-galactosidase by unspecific disulfide bond formation, but failed to restore motility to the *dsbA* deletion strain. The two purified variants 70 and 71 had sufficiently high redox potentials for biological activity (-166 and -185 mV, respectively) proved to be fully active in oxidizing hirudin, and were efficiently reoxidized by DsbB *in vitro*. Thus, there is no obvious reason why these variants should not be fully biologically active. We believe that the simplest explanation for the failure to reconstitute motility in the *dsbA* deletion strain is a more restricted substrate specificity of the semi-active variants. As mentioned above, all sequenced semi-active variants have a basic residue at position 31, which is normally not found in active variants (Table I; only exception: variant 47). A basic residue at position 31 may prevent efficient oxidation of subunits of the flagellar P-ring, whose functional assembly is probed by the motility reconstitution assay. This result suggests that broad substrate specificity is a critical functional property of DsbA, which is supposed to be the common oxidant of all 148 periplasmic *E. coli* proteins (Blattner *et al.*, 1997).
Materials and methods

Materials

1,4-Dithio-DL-threitol (DTT), reduced glutathione (GSH), oxidized glutathione (GSSG), ampicillin, 5,5´-dithio-bis-(2-nitrobenzoic acid) (DTNB), 5-bromo-4-chloro-3-β-D-galacto-pyranoside (X-Gal) and polymyxin B sulfate were purchased from Axon Labs AG (Baden, Switzerland). DE52-cellulose were purchased from Whatman (Maidstone, United Kingdom) and the Phenyl Superose HR 10/10 column was obtained from Pharmacia (Uppsala, Sweden). All other chemicals were from Merck (Darmstadt, Germany) and of the highest purity available. DNA oligonucleotides were from Microsynth (Balgach, Switzerland) and DNA-modifying enzymes from MBI Fermentas (Vilnius, Lithuania), Roche Diagnostics (Rotkreuz, Switzerland), Axon Lab AG (Baden, Switzerland) or New England Biolabs (Beverly, MA, USA). Polyclonal rabbit anti-DsbA antibodies were obtained from Drs. Rosskopf and Fraefel (Zürich, Switzerland). *E. coli* THZ2 (dsbA::kan, recA::cam, λmalF-lacZ102) was kindly provided by Dr. J. Bardwell (University of Michigan).

Random mutagenesis and screening

The plasmid pDsbA3-CC30/33AA for periplasmic production of the inactive DsbA variant Cys30Ala-Cys33Ala was derived after amplification of the mutant gene from a previously described plasmid (Wunderlich *et al.*, 1995) and replacement of the *dsbA* wild type gene in pDsbA3 (Hennecke *et al.*, 1999) via the restriction sites NdeI and BamHI. Plasmids pDsbA3 and pDsbA3-CC30/33AA thus both contain the trc promoter - lac operator sequence for inducible gene expression and a f1 replication origin. Random mutagenesis of the active-site helix was performed according to Kunkel (Kunkel *et al.*, 1987), using single-stranded, uridinylated vector DNA. For identification of active variants after random mutagenesis, pDsbA3-CC30/33AA was used as template vector to eliminate active wild type background. For analogous reasons, we used the wild type expression plasmid pDsbA3 as mutagenesis template for identification of biologically inactive DsbA variants. In both experiments, the mutagenesis primer 5´-AGA AAT ATG CAG AAC TTC NNN NNN NNN GCA NNN NNN GCA GAA GAA AGA GAA AAA CTC-3´ was used.

Screening for DsbA⁺ and DsbA⁻ phenotypes was performed in the absence of IPTG with i) a blue-white screening assay based on oxidative inactivation of periplasmically
Results

Randomization of the active-site helix of DsbA

... orientated β-galactosidase (Bardwell et al., 1991) and ii) by motility assays based on DsbA-dependent assembly of intact flagella (Dailey & Berg, 1993) as described (Hennecke et al., 1997b) using *E. coli* THZ2 (dsbA::kan, recA::cam, λmalF-lacZ102) as expression strain. Periplasmic production of DsbA variants was analyzed by SDS-PAGE of periplasmic extracts, followed by immunoblotting and specific staining with polyclonal rabbit anti-DsbA antibodies (Hennecke et al., 1999). The sequences of the active and inactive variants were determined by dideoxynucleotide sequencing after restriction analysis of the corresponding expression plasmids.

The mutagenesis yield for generation of active variants was quantified after growth of the entire pool of DsbA-producing *E. coli* cells in LB-amp medium, preparation of the plasmid library and complete digestion of the library with *Nsi*I. The mutagenesis primer eliminates a single *Nsi*I site in pDsbA3-CC30/33AA. The yield of mutagenesis could thus be determined after separation of the restriction digest products on an agarose gel and densitometric analysis of ethidium bromide-stained bands.

Cloning of cytoplasmic expression plasmids and purification of selected DsbA variants

As previous results have shown that cytoplasmic rather than of periplasmic expression of DsbA results in much higher production yields (Hennecke et al., 1999), the genes encoding the nine different DsbA variants selected for further characterization were amplified by PCR with the oligonucleotides 5'-GCG ACT GGA ATT CCA T AT GGC GCA GTA TGA AGA TG-3' (*Nde*I site is underlined) and 5'-GGC GCG TGG GGA TCC-3' (*Bam*HI site is underlined) and cloned via *Nde*I and *Bam*HI into the vector pDsbAcyto (Hennecke et al., 1999).

The cytoplasmic production of the DsbA variants was performed essentially as described previously (Hennecke et al., 1999). *E. coli* BL21(DE3) harboring the corresponding derivative of pDsbAcyto were grown at 37 °C in rich medium (20 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl, 20 ml/l glycerol, 50 mM K₂HPO₄, 10 mM MgCl₂, 10 g/l glucose, 100 μg/l ampicillin) to an OD₆₀₀ of 1.0 and induced with 1 mM IPTG. After further growth for 5 h at 37 °C the cells were harvested by centrifugation, suspended in 12 ml of extraction buffer (50 mM Tris/HCl, pH 8.0, 1 mM MgCl₂, 1 mg/l lysozyme) per liter of bacterial culture at 0 °C and disrupted by sonification. The soluble fraction of the extract was dialyzed against 10 mM MOPS/NaOH, pH 7.3,
applied to a DE52 cellulose anion exchange column, and bound proteins were eluted by a linear gradient from zero to 400 mM NaCl. Fractions containing the DsbA variants were pooled, mixed with 4 M ammonium sulfate to a final concentration of 0.8 M ammonium sulfate and applied to a phenyl sepharose column equilibrated with 0.8 M ammonium sulfate, 10 mM MOPS/NaOH pH 7.3. The proteins were eluted with a linear gradient from 0.8 to zero M ammonium sulfate. Fractions with pure DsbA variants were pooled, dialyzed against distilled water and concentrated.

Protein concentrations were measured by the specific absorbance at 280 nm (Gill & von Hippel, 1989) and the correct mass of the variants was verified with MALDI-TOF mass spectrometry. All variants were obtained in the oxidized form after purification, as shown by the lack of free thiol groups (Ellman, 1959). The only exception was variant 71, which has an additional free cysteine at position 32 (see Table I). Far- and near-UV CD spectra were measured at 25 °C on a Jasco 715 CD spectropolarimeter as described (Hennecke et al., 1997a).

**Redox potentials of active-site helix variants**

The redox potentials of the variants were determined by measuring their equilibrium constants with glutathione, using the redox state-dependent fluorescence of the DsbA variants at 325 nm and assuming no significant equilibrium concentrations of DsbA-glutathione mixed disulfides (Wunderlich & Glockshuber, 1993). Measurements were performed in 100 mM sodium phosphate, pH 7.0, 1 mM EDTA, containing 0.1 mM GSSG and zero to 2 mM GSH at 25 °C, and a value of -240 mV was used for the standard redox potential of glutathione (Rost & Rapoport, 1964) to deduce \( E_0' \) values.

**\( pK_a \) values of the active site variants**

The \( pK_a \) value of Cys\(^{30}\) in the different DsbA variants was measured by the decrease in absorbance at 240 nm due to protonation of the Cys\(^{30}\) thiolate (Jacobi et al., 1997; Mössner et al., 1998; Nelson & Creighton, 1994). Titrations were performed at protein concentrations of 3.5 \( \mu \)M in 10 mM sodium phosphate, 10 mM sodium citrate, 10 mM boric acid containing 200 mM KCl and 3 mM DTT by stepwise addition of small portions of 0.1 M HCl. Absorbance values were corrected for the volume
increase. Identical measurements were performed with the corresponding oxidized proteins in the absence of DTT and used for baseline correction.

_Hirudin refolding_

Refolding of reduced hirudin by stoichiometric amounts of the oxidized DsbA variants was performed at pH 7.0 and 25 °C. Reduced, unfolded hirudin (28 µM) was prepared and mixed with oxidized DsbA (84 µM) as described (Hennecke et al., 1999; Otto & Seckler, 1991; Wunderlich et al., 1993b). Aliquots of 120 µl were removed after different reaction times and quenched with formic acid (final concentration: 10% (v/v), pH < 2). The reaction products were separated by reversed-phase HPLC on a 218TP54 C18 column (Vydac, Hesperia, CA, USA) at 55 °C with a 20 to 24% (v/v) acetonitrile gradient in 0.1% (v/v) trifluoroacetic acid, and hirudin folding intermediates were detected by their absorbance at 230 nm.

_Oxidation of the reduced variants by DsbB in vitro_

Membranes containing DsbB and terminal oxidases were prepared from *E. coli* JCB819 cells overexpressing DsbB as described (Bader et al., 1998; Jonda et al., 1999). The DsbB-catalyzed oxidation of reduced DsbA was followed by the decrease in specific DsbA fluorescence at 330 nm (excitation at 280 nm). The reaction was performed at 25 °C in a volume of 1 ml in 50 mM Tris/Cl pH 8.0, 300 mM NaCl containing different concentrations of the reduced DsbA variants (1-40 µM) as described (Mössner et al., 1998). The reaction was started by addition of DsbB-containing membranes to reduced DsbA.
Randomization of the active-site helix of DsbA

Results

Figure legends

Figure 1. (A) Ribbon representation of oxidized DsbA (Guddat et al., 1997; Martin et al., 1993). The thioredoxin-like domain is shown in blue and the α-helical domains in red. The segment comprising the active-site helix α1 (residues 30-37) is depicted in green, and the side chains of Cys30 and Cys33 are shown in ball-and-stick representation. The figure was created with the program MOLMOL (Koradi et al., 1996). (B) Alignment of the active-site helices α1 of known DsbA proteins. Basic residues are depicted with blue background (light blue for His), acidic residues with red background, polar residues with green and hydrophobic residues with brown background, respectively. For proteins from the same organism with identical active-site sequences only one representative is shown.

Figure 2. Comparison of the near-UV (top) and far-UV (bottom) circular dichroism spectra of the oxidized (left) and reduced (right) variants of DsbA. The spectra were recorded at pH 7.4 and 25 °C.

Figure 3. Redox properties of DsbA variants after active-site helix randomization. (A) Determination of the intrinsic redox potentials of purified variants through equilibration of the proteins at different [GSH]2/[GSSG] ratios (cf. Table II). The fractions of oxidized and reduced protein were measured through the redox state-dependent tryptophan fluorescence (cf. Table II), and the redox transitions were normalized. As an example, the redox equilibria with glutathione of the active variant 3, the semi-active variant 70, and the inactive variant 83 are shown. For numbering and primary structure of the variants, see Table I. (B) Plot of measured equilibrium constants with glutathione (K_{eq}) against the pK_a of Cys30 for the purified active-site variants of DsbA. The solid and dashed lines indicate theoretically expected correlations based on the Brønsted theory of disulfide exchange reactions (Szajewski & Whitesites, 1980). The solid line represents a model that assumes a constant pK_a of 11.7 for the buried Cys33 thiol, whereas the dashed line is based on a model which assumes pK_a (Cys30) + pK_a (Cys33) = 16.9 (Mössner et al., 2000). Variants 1 and 70 deviate most strongly from the theoretical models.
Figure 4. Kinetics of the stoichiometric oxidation of reduced hirudin (28 μM) with 3 molar equivalents of purified DsbA variants at pH 7.0 and 25°C. Reactions were quenched with acid after different reaction times and disulfide folding intermediates of hirudin were separated by reversed-phase HPLC using a water-acetonitrile gradient. R and N denote the HPLC retention times of completely reduced hirudin (6 SH groups) and oxidized, native hirudin (3 disulfide bonds), respectively.

Figure 5. Michaelis-Menten kinetics of the DsbB-catalyzed oxidation of reduced DsbA variants with molecular oxygen at pH 8.0 and 25°C. DsbA concentration was varied between zero and 40 μM, and the reactions were followed by the decrease in DsbA fluorescence at 330 nm. A membrane preparation from an *E. coli* strain overproducing DsbB was used to catalyze the reaction as described previously (Bader *et al.*, 1998; Jonda *et al.*, 1999).
References


Randomization of the active-site helix of DsbA


Randomization of the active-site helix of DsbA

Table I: DsbA variants with randomized active-site helix

<table>
<thead>
<tr>
<th>variant number</th>
<th>amino acid number</th>
<th>% buried</th>
<th>expres. level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>WT</td>
<td>C P P C Y C F P</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>• 1</td>
<td>C P P C R L T T</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 2</td>
<td>C P E C S G F S</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 3</td>
<td>C P H A C G R Y T</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 4</td>
<td>C P H C O V T T R</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 5</td>
<td>C P D C R G G R R</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 6</td>
<td>C P T C C Q V T H</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 7</td>
<td>C P F C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 8</td>
<td>C P T C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 9</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 10</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 11</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 12</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 13</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 14</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 15</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 16</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 17</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 18</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 19</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 20</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 21</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 22</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 23</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 24</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 25</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 26</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 27</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 28</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 29</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 30</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 31</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 32</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 33</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>• 34</td>
<td>C T S C C Q V F A</td>
<td>+++</td>
<td></td>
</tr>
</tbody>
</table>

a DsbA variants were classified after determined in vivo activity in a β-galactosidase (Bardwell et al., 1991) and a motility assay (Dailey & Berg, 1993): active in both assays (class I), active in the β-galactosidase assay but inactive in the motility assay (class II) and inactive in both assays (class III).

b the purified variants for in vitro measurements are marked (*).

c % buried of the side chain of the amino acid was calculated with the program MOLMOL (Koradi et al., 1996).

d expression level of the DsbA variants was classified in: strong overexpression (+++), overexpression (++), expression only visible with immunostaining (+) and no expression detectable even with immunostaining (-).
### Class Ib: active DsbA variants

<table>
<thead>
<tr>
<th>variant number</th>
<th>amino acid number</th>
<th>expres. level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>% buried</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>WT</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>36</td>
<td>C</td>
<td>S</td>
</tr>
<tr>
<td>37</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>38</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>39</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>40</td>
<td>C</td>
<td>S</td>
</tr>
<tr>
<td>41</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>42</td>
<td>C</td>
<td>S</td>
</tr>
<tr>
<td>43</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>44</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>45</td>
<td>C</td>
<td>M</td>
</tr>
<tr>
<td>46</td>
<td>C</td>
<td>S</td>
</tr>
<tr>
<td>47</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>48</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>49</td>
<td>C</td>
<td>Y</td>
</tr>
<tr>
<td>50</td>
<td>C</td>
<td>Y</td>
</tr>
<tr>
<td>51</td>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td>52</td>
<td>C</td>
<td>S</td>
</tr>
<tr>
<td>53</td>
<td>C</td>
<td>L</td>
</tr>
<tr>
<td>54</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>55</td>
<td>C</td>
<td>S</td>
</tr>
<tr>
<td>56</td>
<td>C</td>
<td>V</td>
</tr>
<tr>
<td>57</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>58</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>59</td>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td>60</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>61</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>62</td>
<td>C</td>
<td>S</td>
</tr>
<tr>
<td>63</td>
<td>C</td>
<td>V</td>
</tr>
<tr>
<td>64</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>65</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>66</td>
<td>C</td>
<td>I</td>
</tr>
</tbody>
</table>

---

**a** DsbA variants were classified after determined *in vivo* activity in a β-galactosidase (Bardwell *et al.*, 1991) and a motility assay (Dailey & Berg, 1993): active in both assays (class I), active in the β-galactosidase assay but inactive in the motility assay (class II) and inactive in both assays (class III).

**b** The purified variants for *in vitro* measurements are marked (*•*).

**c** % buried of the side chain of the amino acid was calculated with the program MOLMOL (Koradi *et al.*, 1996).

**d** Expression level of the DsbA variants was classified in: strong overexpression (+++), overexpression (++), expression only visible with immunostaining (+) and no expression detectable even with immunostaining (-).
### Class II: semi-active DsbA variants

<table>
<thead>
<tr>
<th>Variant number&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Amino acid number</th>
<th>% buried&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Expression level&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>WT</td>
<td>C</td>
<td>P</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>33</td>
<td>30</td>
</tr>
<tr>
<td>70</td>
<td>C</td>
<td>R</td>
<td>G</td>
</tr>
<tr>
<td>71</td>
<td>C</td>
<td>R</td>
<td>S</td>
</tr>
<tr>
<td>72</td>
<td>C</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>73</td>
<td>C</td>
<td>R</td>
<td>P</td>
</tr>
</tbody>
</table>

<sup>a</sup> DsbA variants were classified after determined *in vivo* activity in a β-galactosidase (Bardwell *et al.*, 1991) and a motility assay (Dailey & Berg, 1993): active in both assays (class I), active in the β-galactosidase assay but inactive in the motility assay (class II) and inactive in both assays (class III).

<sup>b</sup> the purified variants for *in vitro* measurements are marked (•).

<sup>c</sup> % buried of the side chain of the amino acid was calculated with the program MOLMOL (Koradi *et al.*, 1996).

<sup>d</sup> expression level of the DsbA variants was classified in: strong overexpression (+++), overexpression (++), expression only visible with immunostaining (+) and no expression detectable even with immunostaining (-).
### Class III: inactive DsbA variants

<table>
<thead>
<tr>
<th>variant number</th>
<th>amino acid number</th>
<th>expres. level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>% buried</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% buried of the side chain of the amino acid was calculated with the program MOLMOL (Koradi et al., 1996).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>74</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>75</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>76</td>
<td>C</td>
<td>K</td>
</tr>
<tr>
<td>77</td>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td>78</td>
<td>T</td>
<td>N</td>
</tr>
<tr>
<td>79</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>80</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>81</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>82</td>
<td>C</td>
<td>K</td>
</tr>
<tr>
<td>83</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>84</td>
<td>C</td>
<td>K</td>
</tr>
<tr>
<td>85</td>
<td>C</td>
<td>Y</td>
</tr>
<tr>
<td>86</td>
<td>C</td>
<td>K</td>
</tr>
<tr>
<td>87</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>88</td>
<td>C</td>
<td>Q</td>
</tr>
<tr>
<td>89</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>90</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>91</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>92</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>93</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>94</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>95</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>96</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>97</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>98</td>
<td>C</td>
<td>A</td>
</tr>
</tbody>
</table>

---

a DsbA variants were classified after determined in vivo activity in a β-galactosidase (Bardwell et al., 1991) and a motility assay (Dailey & Berg, 1993): active in both assays (class I), active in the β-galactosidase assay but inactive in the motility assay (class II) and inactive in both assays (class III).

b the purified variants for in vitro measurements are marked (+•).

c % buried of the side chain of the amino acid was calculated with the program MOLMOL (Koradi et al., 1996).

d expression level of the DsbA variants was classified in: strong overexpression (+++), overexpression (++), expression only visible with immunostaining (+), and no expression detectable even with immunostaining (-).
Table II: Properties of the purified DsbA variants

<table>
<thead>
<tr>
<th>DsbA variants</th>
<th>In vivo activity</th>
<th>Hirudin refolding</th>
<th>$K_{eq}$ with glutathione</th>
<th>$E'_0$ (mV)</th>
<th>$pK_a$ of Cys$^{30}$</th>
<th>$F_{ox}/F_{red}$</th>
<th>Reoxidation with DsbB</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$K_M$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>+</td>
<td>++</td>
<td>0.112 (± 0.003)</td>
<td>-123</td>
<td>3.30 (± 0.05)</td>
<td>3.00</td>
<td>26.9 (± 0.7)</td>
<td>10.9 (± 0.7)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>+</td>
<td>++</td>
<td>93.9 (± 3.7)</td>
<td>-210</td>
<td>4.98 (± 0.04)</td>
<td>2.73</td>
<td>145 (± 12)</td>
<td>24.6 (± 3.4)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>++</td>
<td>2.96 (± 0.07)</td>
<td>-165</td>
<td>4.90 (± 0.05)</td>
<td>3.63</td>
<td>29.8 (± 0.9)</td>
<td>12.1 (± 0.8)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>10.2 (± 0.5)</td>
<td>-181</td>
<td>6.60 (± 0.06)</td>
<td>2.28</td>
<td>29.2 (± 2.9)</td>
<td>20.2 (± 4.0)</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>±</td>
<td>++</td>
<td>3.11 (± 0.26)</td>
<td>-166</td>
<td>5.42 (± 0.04)</td>
<td>2.16</td>
<td>80.0 (± 5.1)</td>
<td>12.4 (± 1.7)</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>±</td>
<td>+</td>
<td>13.1 (± 1.5)</td>
<td>-185</td>
<td>4.92 (± 0.09)</td>
<td>1.98</td>
<td>26.7 (± 0.1)</td>
<td>4.85 (± 0.08)</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>-</td>
<td>+</td>
<td>240 (± 36)</td>
<td>-222</td>
<td>5.72 (± 0.63)</td>
<td>2.13</td>
<td>25.6 (± 1.0)</td>
<td>4.39 (± 0.48)</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>-</td>
<td>-</td>
<td>180 (± 10)</td>
<td>-218</td>
<td>6.45 (± 0.04)</td>
<td>1.80</td>
<td>9.68 (± 0.48)</td>
<td>4.99 (± 0.77)</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>-</td>
<td>-</td>
<td>167 (± 18)</td>
<td>-217</td>
<td>5.75 (± 0.32)</td>
<td>1.86</td>
<td>not determined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>-</td>
<td>-</td>
<td>149 (± 2)</td>
<td>-216</td>
<td>6.31 (± 0.03)</td>
<td>1.86</td>
<td>not determined</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

$a$ The ability of the different variants to restore DsbA deficiency in a *dsbA* deletion strain was probed with oxidative inactivation of periplasmically oriented β-galactosidase and recovery of cell motility (see also Table I). (+) Active DsbA in both complementation assays; (±) complementation in the β-galactosidase assay, but no recovery of motility; (-) no DsbA activity in both complementation assays.

$b$ Stoichiometric oxidation of hirudin with the different DsbA variants was classified into 3 categories: (-): less than 2% of the initial velocity of native hirudin formation compared to DsbA wild type; (+), 10-15% of DsbA wild type activity; (++) ≥ 100% of wild type activity.

$c$ The standard redox potential of DsbA was calculated from the equilibrium constant ($K_{eq}$) with glutathione ($E'_0 = -240$ mV for GSH/GSSG).

$d$ $F_{red}/F_{ox}$: relative increase in fluorescence intensity at 324 nm ($\lambda_{ex} = 280$ nm) upon reduction of the active-site disulfide bridge.
Figure 1
### Table B

<table>
<thead>
<tr>
<th>Protein and species</th>
<th>Number of amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td>DsbA from <em>Escherichia coli</em></td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Shigella flexneri</em></td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Salmonella typhimurium</em></td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Yersinia pestis</em></td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Erwinia carotovora</em></td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Pectobacterium carotovorum</em> subsp. atrosepticum</td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Erwinia chrysanthemi</em></td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Pasteurella multocida</em></td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Haemophilus influenzae</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Buchnera aphidicola</em></td>
<td>P</td>
</tr>
<tr>
<td>DsbA from <em>Vibrio cholerae</em></td>
<td>P</td>
</tr>
<tr>
<td>Dlp from <em>Salmonella enteritidis</em></td>
<td>A</td>
</tr>
<tr>
<td>Dlt protein from <em>Salmonella typhi</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Pseudomonas syringae</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Azobacter vinelandii</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Pseudomonas aeruginosa</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Burkholderia cepacia</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Neisseria meningitidis</em></td>
<td>A</td>
</tr>
<tr>
<td>STM3193 from <em>Salmonella typhimurium</em> LT2</td>
<td>A</td>
</tr>
<tr>
<td>XF1437 from <em>Xylella fastidiosa</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Klebsiella pneumoniae</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Enterobacter amnigenus</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Legionella pneumophila</em></td>
<td>A</td>
</tr>
<tr>
<td>NMA2209 from <em>Neisseria meningitidis</em></td>
<td>A</td>
</tr>
<tr>
<td>XF1436 from <em>Xylella fastidiosa</em></td>
<td>A</td>
</tr>
<tr>
<td>DsbA from <em>Cytobacter freundii</em></td>
<td>A</td>
</tr>
<tr>
<td>CC0375 from <em>Caulobacter crescentus</em></td>
<td>A</td>
</tr>
<tr>
<td>SMc00023 from <em>Rhizobium meliloti</em></td>
<td>A</td>
</tr>
<tr>
<td>AGR_C_1464 from <em>Agrobacterium tumefaciens</em></td>
<td>A</td>
</tr>
<tr>
<td>RP025 from <em>Rickettsia prowazekii</em></td>
<td>A</td>
</tr>
<tr>
<td>DR0753 from <em>Deinococcus radiodurans</em></td>
<td>A</td>
</tr>
</tbody>
</table>
Results

Figure 2

Randomization of the active-site helix of DsbA

Oxidized

Reduced

near-UV

far-UV
Figure 3

A

![Graph showing fraction reduced (%) vs. pK_{a,nuc} for wild type and variants 3, 70, and 83.]

B

![Graph showing Keq (M) vs. pK_{a,nuc} for WT, 82, 84, 85, 71, 3, 2, 70, and WT.]
Figure 4

Results

Randomization of the active-site helix of DsbA

HPLC retention time

**Figure 4**

**wild type**

**variant 1**

**variant 70**

**variant 84**
Randomization of the active-site helix of DsbA Results

Figure 5

Graph showing nmol oxidized DsbA per second against DsbA variant (μM) for different variants including wild type, variant 1 to variant 85.
5.3 Requirement for two active sites for efficient disulfide isomerase activity of disulfide isomerases

P. Frei, B. Philipps & R. Glockshuber
Requirement for two active sites for efficient disulfide isomerase activity of disulfide isomerases

Patrick Frei, Björn Philipps and Rudi Glockshuber

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Hönnggerberg, CH-8093 Zürich, Switzerland

To whom correspondence should be addressed:

Rudi Glockshuber
Institut für Molekularbiologie und Biophysik
Eidgenössische Technische Hochschule Hönnggerberg
CH-8093 Zürich
Switzerland
Phone: +41-1-633 6819
Fax: +41-1-633 1036
E-mail: rudi@mol.biol.ethz.ch
Abstract

Oxidative folding of proteins with multiple disulfide bonds involves both oxidation of cysteine thiols and isomerization of non-native disulfide bonds. Both reactions are catalyzed by enzymes of the thioredoxin family *in vivo*. While protein disulfide isomerase (PDI) from the endoplasmic reticulum of eukaryotes simultaneously possesses dithiol oxidase and disulfide isomerase activity, these activities are catalyzed by separate proteins in the bacterial periplasm. Whereas the monomeric enzyme DsbA represents the main oxidant for folding polypeptides in the periplasm of *Escherichia coli*, the homodimeric enzyme DsbC catalyzes rearrangements of non-native disulfide bonds. PDI, like the DsbC homodimer, possesses two catalytically active thioredoxin domains. As the isolated catalytic domains of PDI and DsbC are only poor isomerases, efficient disulfide isomerase activity may be linked to the presence of two active sites in the catalyst. Here, we have tested this hypothesis by converting the monomeric disulfide oxidase DsbA into a disulfide isomerase by artificial dimerization. For this purpose, DsbA was fused to the C-terminus of the dimerization domain of DsbC. The resulting fusion indeed proved to be a homodimer with about 2-fold higher isomerase activity compared to wild type DsbA.
Introduction

Oxidative protein folding in the bacterial periplasm requires the periplasmic enzymes DsbA, DsbC and possibly DsbG (Andersen et al., 1997; Bardwell et al., 1991; Kamitani et al., 1992; Missiakas et al., 1994; Rietsch et al., 1996; Ritz & Beckwith, 2001). DsbA, DsbC and DsbG belong to the family of thiol-disulfide oxidoreductases which share the thioredoxin fold and a catalytic disulfide bond with the active-site sequence Cys-Xaa-Xaa-Cys (Martin et al., 1993). The monomeric 21 kDa protein DsbA randomly and rapidly introduces disulfide bonds into folding substrate proteins (Wunderlich & Glockshuber, 1993b; Zapun et al., 1993) and is thought to be the generic oxidant of all periplasmic disulfide bonds. DsbA contains a single active-site disulfide bond, and the three-dimensional structures of the enzyme have been solved for both redox states of the protein (Guddat et al., 1998; Guddat et al., 1997; Martin, 1995; Schirra et al., 1998). DsbA is maintained in the oxidized state by disulfide exchange with the inner membrane protein DsbB (Bader et al., 1999; Bardwell et al., 1993; Dailey & Berg, 1993; Guilhot et al., 1995; Kishigami & Ito, 1996; Missiakas et al., 1993). DsbA has only very weak disulfide isomerase activity (Zapun et al., 1995). This function is efficiently fulfilled by DsbC in the periplasm (Missiakas et al., 1994; Zapun et al., 1995).

In contrast to monomeric DsbA, DsbC is a homodimer of 23 kDa subunits (Zapun et al., 1995). The three-dimensional structure of DsbC in the oxidized form has been determined by X-ray crystallography (McCarthy et al., 2000). Each subunit consists of an N-terminal dimerization domain connected to a C-terminal, catalytic domain via a helical linker segment (Fig. 1). Both catalytic domains appear to act independently (Zapun et al., 1995) and do not interact with each other in the X-ray structure of the enzyme (the average distance between both active-site disulfides is 38 Å) (McCarthy et al., 2000). The catalytically active, reduced state of DsbC is maintained by disulfide exchange with the inner membrane protein DsbD (Joly & Swartz, 1997; Missiakas et al., 1995; Rietsch et al., 1996; Rietsch et al., 1997). The DsbA-DsbB and DsbC-DsbD redox systems are kept separate by kinetically restricted disulfide exchange between DsbA and DsbC (Darby et al., 1998b), and by the dimeric state of DsbC which prevents oxidation of DsbC by DsbB (Bader et al., 2001). In *E. coli*, there is another, dimeric disulfide isomerase, DsbG, which is homologous to DsbC (Andersen et al., 1997; Bessette et al., 1999). Like DsbC, DsbG is kept reduced by
DsbD (Bessette et al., 1999) and acts as disulfide isomerase, albeit perhaps with a narrower substrate range (van Straaten et al., 1998).

It is interesting to note that all known, efficient disulfide isomerases involved in catalysis of oxidative protein folding, i.e., bacterial DsbC and DsbG, and eukaryotic protein disulfide isomerase (PDI), possess two catalytic domains. PDI is composed of four thioredoxin-like domains, denoted a, b, b’, and a’, of which the a and a’ domains are catalytically active (Darby et al., 1998a; Kemmink et al., 1997). A particular feature of DsbC and PDI is their ability to make buried disulfide bonds in kinetically trapped folding intermediates accessible (Darby et al., 1998b; Weissman & Kim, 1993; Zapun et al., 1995). For both enzymes, this property has been shown to depend on the presence of both active sites, as the isolated, catalytic domains of these enzymes are only weak disulfide isomerases (Darby et al., 1998a; Sun & Wang, 2000). The ability of PDI, DsbC and DsbG to make disulfide bonds in kinetically trapped folding intermediates accessible is most likely due to a high-affinity binding site for unfolded or partially folded polypeptides. For example, DsbC has a ca. 10⁴-fold higher affinity for peptides than DsbA (Darby et al., 1998b), and PDI accelerates reshuffling of kinetically trapped BPTI folding intermediates up to 6000-fold (Weissman & Kim, 1993).

A plausible explanation for the high-affinity peptide binding sites in disulfide isomerases would be the creation of a continuous peptide binding groove through dimerization or covalent connection of domains with low-affinity binding sites. In the case of PDI, the b and, in particular, the b’ domain are likely to contribute to substrate binding (Darby et al., 1998a; Klappa et al., 1997). A high-affinity binding site for unfolded polypeptide chains is also fully consistent with the intrinsic chaperone activities described for PDI, DsbC and DsbG (Chen et al., 1999; Noiva, 1999; Shao et al., 2000). Binding to an unfolded peptide segment in a folding intermediate would cause extremely high effective concentrations of the active sites relative to the substrate and thus strongly accelerate catalysis of disulfide isomerization. Another mechanism, according to which two active sites would improve disulfide isomerase activity independent of peptide binding would be formation of a mixed disulfide between one of the active sites and the substrate, which again would strongly increase the effective concentration of the other active site relative to the substrate.
In the present study, we have addressed the question of how efficient disulfide isomerase activity is achieved in enzymes of the thioredoxin family by artificial dimerization of the very poor, monomeric disulfide isomerase DsbA. To this end, we have made a construct analogous to DsbC which is composed of the N-terminal dimerization domain of DsbC fused to DsbA. We have compared the disulfide isomerase activity of the resulting DsbA dimer with those of DsbA, DsbC and the isolated, catalytic domain of DsbC. Moreover, we have investigated whether and why these constructs are capable of complementing DsbA deficiency in a dsbA deletion strain.
Results

Construction of an artificial DsbA homodimer analogous to DsbC

The three-dimensional structure of the oxidized DsbC homodimer shows that each subunit consists of an N-terminal dimerization domain (residues 1-61), connected via a helical hinge segment (residues 62-77) to a catalytic, thioredoxin-like domain (residues 78-216) (McCarthy et al., 2000) (Fig. 1). To mimic the architecture of DsbC in an artificial DsbA homodimer, we fused DsbA to the C-terminus of the N-terminal part of DsbC (residues 1-77) containing the dimerization domain of DsbC and the linker helix. In addition, we inserted a glycine residue at the fusion site to account for maximal flexibility of the DsbA moieties in the artificial dimer and to allow free rotation of the DsbA units relative to one another. Moreover, we introduced a (His)$_6$ tag at the C-terminus of the fusion protein for purification via metal chelate affinity chromatography. The fusion protein was termed D-DsbA and consists of 267 amino acids. D-DsbA was produced as a soluble protein in the periplasm of *E. coli* using the DsbA signal sequence, and could also be produced directly in the cytoplasm in soluble form. Because cytoplasmic production resulted in much higher production yields, D-DsbA was purified from the cytoplasm, using chromatography on Ni$^{2+}$-NTA agarose followed by gel filtration (Fig. 2). The total yield of D-DsbA was 10 mg of purified protein per liter of bacterial culture.

Using the DsbA signal sequence, we also expressed the isolated catalytic domain of DsbC (termed DsbC-C) with linker helix (residues 62-216) and C-terminal (His)$_6$ tag in the *E. coli* periplasm as a control construct. The linker segment was retained in the construct due to possible stabilizing interactions between the linker helix with helix $\alpha$6 of the catalytic domain (McCarthy et al., 2000). The yield of DsbC-C after purification to homogeneity from the periplasm (Fig. 2) was 0.5 mg per liter of bacterial culture. Edman sequencing and MALDI-TOF mass spectrometry showed that all constructs were correctly processed and not degraded in the cell or during purification.

Analytical gel filtration experiments

To test the oligomerization state of D-DsbA compared with wild type DsbA, wild type DsbC and DsbC-C, we performed analytical gel filtration experiments. First, we analyzed the apparent molecular mass of DsbC by varying the initial DsbC
concentration over two orders of magnitude (4.6-460 μM). Fig. 3A shows that about 40% of the DsbC molecules dissociated to monomers at the lowest initial DsbC concentration (4.6 μM). As the protein was diluted about 10-20-fold during the gel filtration run, we estimate the dissociation constant of the DsbC homodimer to be in the micromolar region. Moreover, the fact that two distinct peaks were obtained for the monomer and dimer indicates that the association-dissociation equilibrium of DsbC is comparatively slow and not very dynamic. As expected, the isolated catalytic domain of DsbC behaved as a monomer at all concentrations.

Analogous experiments with the artificial DsbA dimer showed very similar results. Complete dimer formation was observed at high protein concentrations. Like DsbC, D-DsbA partially dissociated at the lowest initial concentration of 4.6 μM, although D-DsbA was somewhat more stable against dissociation than wild type DsbC (Fig. 3B). Overall, the gel filtration experiments showed that the D-DsbA homodimer is at least as stable as wild type DsbC and that the N-terminal dimerization domain of DsbC functions essentially independently of the C-terminal fusion partner. The dimerization domain of DsbC may thus represent a generally applicable dimerization module for protein engineering.

Disulfide isomerase and isomerase activity of the artificial DsbA dimer and the isolated DsbC domain

To assess the disulfide isomerase activity of D-DsbA relative to DsbC, DsbA and DsbC-C in both a quantitative and mechanistic manner, we used the bifunctional α-amylase-trypsin inhibitor from Ragi (RBI) as a model system. RBI is a monomeric 122 amino acid protein with a complex disulfide pattern and 5 disulfide bonds in the native state (Strobl et al., 1998; Strobl et al., 1995). Due to the large number of intramolecular disulfide bonds in RBI and the high probability of forming incorrect disulfides during folding, it is not oxidation of cysteine thiols but isomerization of nonnative disulfide bonds that limits the rate of oxidative folding of the inhibitor (Maskos, 1995; Wunderlich & Glockshuber, 1993a). This is particularly true under oxidizing conditions that favor nonnative disulfides. We therefore used strongly oxidizing conditions for oxidative refolding of RBI to detect the disulfide isomerase activities of the individual oxidoreductase constructs specifically acceleration of RBI folding compared to the uncatalyzed reaction was investigated.
RBI can be quantitatively refolded in vitro from the completely reduced state in glutathione redox buffers. At high concentrations of oxidized glutathione (GSSG) and low concentrations of reduced glutathione (GSH), a large number of folding intermediates with nonnative disulfide bonds accumulates (Maskos, 1995). The kinetics of RBI refolding in oxidizing glutathione redox buffers were followed by quenching the reactions with acid after different incubation times and separation of RBI folding intermediates by reversed-phase HPLC. Fig. 4A shows the spontaneous refolding of RBI at pH 7.5 in the presence of 40 mM GSSG and 5 mM GSH. Folding is limited by an enormous number of intermediates with nonnative disulfide bonds, which cannot be resolved by HPLC. Figs. 4B and C show the analogous reactions in the presence of 0.2 molar equivalents (monomer concentrations relative to RBI disulfides) of DsbA and D-DsbA. The reaction conditions were chosen so as to guarantee almost complete dimer formation of DsbC and D-DsbA. In the case of DsbC, the apparent refolding rate was increased by a factor of 6.9 compared to the uncatalyzed reaction, and D-DsbA accelerated RBI refolding 1.7-fold (Fig. 5). Both DsbC and D-DsbA obviously reduced the number of well-populated folding intermediates, as one would expect for disulfide isomerase activity. In contrast, even slightly slower refolding of RBI was observed when the reaction was performed in the presence of DsbA or the isolated catalytic-domain of DsbC, and the number of RBI folding intermediates did not decrease significantly (Figs. 4 and 5). In summary, the disulfide isomerase activity of DsbA could be improved about 2-fold by artificial dimerization compared to wild type DsbA (Fig. 5).

To test the disulfide reductase properties of the dimerized DsbA relative to DsbA, DsbC and the catalytic domain of DsbC we measured the reduction of 130 µM insulin with 1 mM DTT. The observed reduction velocity of insulin in the case of dimeric catalyst DsbC and D-DsbA is faster than by monomeric counterparts DsbC-C and DsbA (Fig. 6).

In vivo complementation of DsbA deficiency

Recent experiments (Bader et al., 2001) have demonstrated that the dimeric state of DsbC is required for its in vivo activity. This was shown by the search for DsbC variants that could complement DsbA deficiency in E. coli, which exclusively yielded monomeric DsbC variants (Bader et al., 2001). The monomeric DsbC variants are, in
contrast to dimeric wild type DsbC, good substrates of DsbB. Dimerization of DsbC thus somehow prevents its oxidation by DsbB, possibly for steric reasons, and guarantees the maintenance of the catalytically active, reduced state of DsbC by DsbD and the coexistence and independence of the DsbA-DsbB and DsbC-DsbD redox systems (Bader et al., 2001). These data predicted that the isolated, monomeric DsbC domain would be a substrate of DsbB when expressed in the periplasm and able to complement DsbA deficiency. This is indeed the case, as shown by the reconstitution of motility of the DsbA deletion strain THZ2. The complementation assay is based on the DsbA-dependent formation of the flagellar P-ring (Dailey & Berg, 1993). In contrast to wild type DsbC, the artificial DsbA dimer D-DsbA also complemented the DsbA deficiency (Fig. 7). The reasons for DsbA complementation by D-DsbA are not clear, however. Previous complementation assays with extremely unstable DsbA variants had shown that minute concentrations of DsbA (below detectability by Western Blots) are already sufficient for reconstitution of the DsbA$^+$ phenotype (Hennecke et al., 1999). As we observed partial proteolytic cleavage of D-DsbA in the linker between the dimerization domain and DsbA on a Western blot (data not shown), it could well be that the reconstitution of motility by D-DsbA is an indirect effect of proteolytic degradation of the fusion by periplasmic proteases that generate DsbA monomers. Edman sequencing revealed preferable cleavage within the linker helix of the Leu$^{67}$-Leu$^{68}$ peptide bond, and alternative cleavage sites at the Asn$^{71}$-Asn$^{72}$ and Asn$^{72}$-Ala$^{73}$ peptide bonds (data not shown). These cleavage sites are very close to the trypsin cleavage site Lys$^{65}$-Met$^{66}$ (Sun & Wang, 2000).
Results

Discussion

The fusion protein consisting of the dimerization domain of DsbC and DsbA acts as a dimer and even shows a somewhat smaller dimerization constant than DsbC, suggesting that fusion of this dimerization domain to any protein may be a useful method for generation of dimeric proteins.

The refolding of fully reduced RBI under strongly oxidizing conditions that favor formation of nonnative disulfide bonds can be used to study the function of disulfide isomerases (Maskos, 1995). Under these conditions, DsbC is able to accelerate the folding rate of RBI by a factor of 6 and D-DsbA by a factor of 2 compared with the uncatalyzed reaction, while DsbC-C and DsbA decrease the folding rate somewhat under these conditions. Similar order was observed by comparison of the reduction velocity of insulin. These results demonstrate that the strongest known disulfide oxidase, DsbA, can be converted into a disulfide isomerase by artificial dimerization. The lower activity of D-DsbA compared to DsbC could be explained by non-optimal distance and orientation of the two active centers compared to DsbC, but an even simpler explanation would be a lower intrinsic peptide binding affinity of DsbA compared to the catalytic domain of DsbC. The fact that DsbC-C and DsbA lack isomerase activity demonstrates the need for two active sites for a good disulfide isomerase activity, as in the oligomeric isomerases DsbC and DsbG (Bader et al., 2001; Bessette et al., 1999; Darby et al., 1998a; Sun & Wang, 2000).

Complementation of the DsbA deficiency in the motility assay, based on the DsbA-dependent folding of the P-ring subunits from the flagellar motor (Dailey & Berg, 1993), was achieved by plasmid-encoded DsbA, dimerized DsbA (D-DsbA) and the catalytic domain of DsbC (DsbC-C), but not DsbC. The ability of D-DsbA to complement the DsbA-deficiency could be explained by partial degradation within the linker region, because it is known that minimal amounts of DsbA in the periplasm are sufficient the restore the DsbA* phenotype (Hennecke et al., 1999). The ability of DsbC-C to complement the DsbA deficiency is in full agreement with results obtained for monomeric variants of DsbC with mutations in the dimerization domain (Bader et al., 2001). These DsbC variants where oxidized by DsbB in vivo and in vitro (Bader et al., 2001). As DsbB has been shown to accept many other disulfide oxidoreductases as substrates, such as thioredoxin and PDI (Jonda et al., 1999), the most likely
reason for accepting DsbC-C, but not dimeric wild type DsbC as a substrate, appears to be a steric clash preventing disulfide exchange.
Material and methods

Materials
1,4-Dithio-DL-threitol (DTT) and ampicillin were purchased from AppliChem (Heidelberg, Germany) and reduced glutathione (GSH), oxidized glutathione (GSSG), 5,5’-dithio-bis-(2-nitrobenzoic acid) (DTNB), 5-bromo-4chloro-3-β-D-galacto-pyranoside (X-Gal) and polymyxin B sulfate from Sigma (Buchs, Switzerland). CM52- and DE52-cellulose were purchased from Whatman (Maidstone, United Kingdom), Ni-NTA column from Quiagen (Hilden, Germany) and HiLoad 26/60 Superdex 75, HR 10/30 Superdex 75, phenyl sepharose and PD 10 gel filtration columns were obtained from Pharmacia (Uppsala, Sweden). All other chemicals were from Merck (Darmstadt, Germany) and of the highest purity available. DNA oligonucleotides were from Microsynth (Balgach, Switzerland) and DNA modifying enzymes from MBI Fermentas (Vilnius, Lithuania), or Axon Lab AG (Baden, Switzerland).

Cloning of the expression plasmids
For construction of the periplasmic expression vector of pD-DsbA, the gene of the dimerization domain of DsbC was amplified via PCR with the oligonucleotides 5’-GGG AAT TCC ATA TGA AGA AAG GTC TTA TGT TGT TAC T-3’ (Ndel restriction site underlined) and 5’-CTG TTA AAG CAG TTG AAT GCG CTC GAG CGG TA-3’ (XhoI restriction site underlined) and the plasmid pDsbC (Maskos, 1995) and the gene of DsbA was amplified with the oligonucleotides 5’- CAC TAC TCG AGA AAG AAG GCG CGC AGT ATG GTA AA -3’ (XhoI restriction site underlined) and 5’- AAA TAT CTG TCC GAG AAA AAA GGC CAT CAC CAT CAC TAA GGA TCC GCG -3’ (BamHI restriction site underlined) and the plasmid pDsbA3 (Hennecke et al., 1999). The PCR products were cleaved with Ndel and XhoI and XhoI and BamHI, respectively and the vector pDsbA3 was cleaved with Ndel and BamHI. The PCR products were cloned in a three-fragment ligation into pDsbA3, yielding the pD-DsbA for periplasmic expression of the fusion between the dimerization domain of DsbC and DsbA with C-terminal (His)_6-tag und control of the trc promoter - lac operator sequence.

For cytoplasmatic expression of D-DsbA, the gene encoding D-DsbA-(His)_6 was amplified from pD-DsbA via PCR with the oligonucleotides 5’-GGG AAT TCC ATA
TGGA TGG ATG ACG CGGCAA TTC AAC AAC G-3′ (Ndel restriction site underlined) and 
5′-AAA TAT CTG TCC GAG AAA AAA GGC CAT CAC CAT CAC CAT CAC TAC TAA 
GGA TCC GCG-3′ (BamHI restriction site underlined). The PCR product was cloned 
into pDsbAcyto (Hennecke et al., 1999) via Ndel and BamHI, yielding the cytoplasmic 
expression vector pD-DsbAcyto for expression of D-DsbA-(His)$_6$ under trc 
promoter/lac operator control.

For periplasmic expression of the isolated catalytic domain of DsbC (residues 78-
216) with C-terminal (His)$_6$-tag, the gene encoding the domain was fused to the 
genetic sequence of the DsbA signal sequence in pDsbA3. The gene for the catalytic 
domain was amplified from pDsbC by PCR with the oligonucleotides 5′-ATT GAC 
TAG CTA GCG TCG TCA CCA ATA AGA TGC TGT TAA AGC AG-3′ (Nhel 
restriction site underlined) and 5′-CAA AAA ATG ACA AGC GGT AAA GGG CAT 
CAT CAT CAT CAT CAT CAT TAA GGA TCC GCG-3′ (BamHI restriction site underlined). 
The PCR fragment was cloned via Nhel and BamHI into pDsbA3, yielding the vector 
pDsbC-C (expression under trp promoter/lac operator control).

Recombinant protein expression and purification

Cytoplasmic production of DsbA using the vector pDsbAcyto was performed 
essentially as described previously (Hennecke et al., 1999). E. coli BL21(DE3) 
harboring the vector pDsbAcyto was grown at 37 °C in rich medium (20 g/l tryptone, 
10 g/l yeast extract, 5 g/l NaCl, 20 ml/l glycerol, 50 mM K$_2$HPO$_4$, 10 mM MgCl$_2$, 10 g/l 
glucose, 100 µg/l ampicillin) to an OD$_{600}$ of 1.0 and induced with 1 mM IPTG. After 
进一步 growth for 5 h at 37 °C the cells were harvested by centrifugation. The cells 
were suspended in 12 ml buffer (50 mM Tris/HCl, pH 8.0, 1 mM MgCl$_2$, 1 mg/l ampicillin) per liter of bacterial culture at 0 °C and disrupted by sonification. After 
centrifugation, the supernatant was dialyzed against 10 mM MOPS/NaOH, pH 7.3, 
and applied to an DE52 cellulose anion exchange column and eluted by a linear NaCl 
gradient from zero to 400 mM. Fractions containing DsbA were pooled, mixed with 
4 M ammonium sulfate to a final concentration of 0.8 M and applied to a phenyl 
sepharose column equilibrated with the same buffer. The protein was eluted by a 
linear gradient from 0.8 to zero M ammonium sulfate. Fractions with pure DsbA were 
pooled, dialyzed against distilled water and lyophilized.
Results

For production of (His)$_6$-tagged D-DsbA, *E. coli* BL21(DE3) cells harboring pD-DsbAcyto were grown in 2xYT medium containing 100 µg/l ampicillin at 30 °C to an OD$_{600}$ of 1.5 and induced with 1 mM IPTG. After further growth at 30 °C for 16 h the cells were harvested by centrifugation, suspended in 30 ml of 50 mM sodium phosphate, pH 8.0, 250 mM NaCl, 1 mg/l lysozyme per liter of bacterial culture at 0 °C and lysed by sonification. The soluble fraction was directly loaded on an Ni-NTA column equilibrated with the same buffer. The protein was eluted with a linear gradient from zero to 300 mM imidazole in 50 mM sodium phosphate, pH 8.0, 300 mM NaCl. The fractions containing D-DsbA were pooled, mixed with EDTA to a final concentration of 1 mM, and applied to an CM52 cation exchange column equilibrated with 10 mM acetic acid/NaOH, pH 5.0. After elution with a linear gradient from zero to 400 mM NaCl, the fractions with D-DsbA were pooled, concentrated after addition of EDTA (final concentration of 0.1 mM) and applied to a HiLoad 26/60 Superdex 75 gel filtration column equilibrated with 20 mM sodium phosphate, pH 7.5, 115 mM NaCl, 0.1 mM EDTA. Fractions with pure, dimeric D-DsbA were pooled, concentrated and dialyzed against distilled water.

The isolated catalytic domain of DsbC (DsbC-C) was produced in *E. coli* HM125 cells harboring pDsbC-C. The cells were grown at 25 °C in 2xYT medium containing 100 µg/l ampicillin and induced with 1 mM IPTG at an OD$_{600}$ of 1.5. After further growth for 16 hours at 25 °C the cells were harvested by centrifugation. The cells were suspended in 10 ml of 10 mM MOPS/NaOH, pH 7.0, 150 mM NaCl, 5 mM EDTA, 1 mg/ml polymyxin B sulfate per liter of bacterial culture, and gently shaken for 1.5 h at 0 °C. After centrifugation, the supernatant was dialyzed against 50 mM sodium phosphate, pH 8.0, 300 mM NaCl and loaded onto an Ni-NTA column equilibrated with the same buffer. The protein was eluted with a linear gradient from zero to 300 mM imidazole in 50 mM sodium phosphate, pH 8.0, 300 mM NaCl. The fractions containing DsbC-C were pooled, concentrated and dialyzed against distilled water.

For periplasmic production of RBI and DsbC, cell growth and harvest were performed as follows: *E. coli* JM83 cells harboring the plasmid pDsbC or pRBI-DsbC were grown in 2xYT medium at 30 °C to an OD$_{600}$ 1.0 and induced with 1 mM IPTG. In the case of RBI production, N-acetyl-cysteine was added to the medium at the time of induction (final concentration: 5 mM). The cells were further grown for 16 h at 30 °C, and harvested by centrifugation. The cells were suspended in 10 ml of 10 mM
MOPS/NaOH, pH 7.0, 150 mM NaCl, 5 mM EDTA, 1 mg/ml polymyxin B sulfate per liter of bacterial culture and gently shaken for 1.5 h at 4 °C. After centrifugation, DsbC or RBI were isolated from the supernatants (periplasmic extracts).

For purification of RBI, periplasmic extracts were dialyzed against 10 mM Tris/HCl, pH 8.0 and applied to an DE52 anion exchange column equilibrated with the same buffer. The elute was directly applied to a CM52 cation exchange column equilibrated with 10 mM Tris/HCl, pH 8.0. After elution of RBI with a linear gradient of zero to 400 mM NaCl in 10 mM Tris/HCl, pH 8.0, fractions with RBI were pooled, dialyzed against 50 mM Tris/HCl, pH 8.0, 100 mM NaCl and ammonium sulfate was added to a final concentration of 1 M. The solution was applied to a Phenyl Superose column equilibrated with the same buffer and RBI was eluted with a linear gradient from 1 to zero M ammonium sulfate in 50 mM Tris/HCl, pH 8.0, 100 mM NaCl. The fractions containing RBI were pooled, concentrated and dialyzed against 40 mM formic acid/NaOH, pH 4.0, 150 mM NaCl.

For purification of DsbC, periplasmic extracts were dialyzed against 10 mM Bis-Tris/HCl, pH 7.0 and applied to a DE52 anion exchange column equilibrated with the same buffer. The protein was eluted with a linear gradient from zero to 400 mM NaCl in 10 mM Bis-Tris/HCl, pH 7.0, DsbC containing fractions were pooled and dialyzed against 40 mM Tris/HCl, pH 7.5, 100 mM NaCl. After addition of ammonium sulfate to a final concentration of 1.5 M, the solution was applied to a Phenyl Sepharose column and the protein was eluted with a linear gradient from 1.5 to zero M ammonium sulfate in 40 mM Tris/HCl, pH 7.5, 100 mM NaCl. The fractions containing DsbC were pooled, dialyzed against distilled water, concentrated to a protein concentration of 500 µM. After complete oxidation of DsbC for 1 h with 100 mM GSSG in 100 mM Tris/HCl, pH 8.0, GSSG was removed on a Superdex 75 gel filtration column equilibrated with 20 mM sodium phosphate, pH 7.4, 115 mM NaCl. Fractions with DsbC were pooled, dialyzed against distilled water and concentrated.

Oxidative refolding of RBI

Completely reduced RBI was prepared by incubation of native RBI (0.5 mM) in 50 mM Tris/HCl, pH 8.7, 6 M guanidinium chloride (GdmCl), 100 mM DTT for 16 h, followed by removal of low-molecular weight species with gel filtration on a PD 10
column, equilibrated with 10 mM HCl. Reduced RBI in 10 mM HCl proved to be stable for several days at 4°C.

RBI refolding was performed under strongly oxidizing conditions favouring formation of nonnative disulfide bonds. Reduced RBI (4.6 µM) was incubated in 2 mM GSH, 40 mM GSSG, 100 mM Tris/HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA at 15°C. Catalysts were added at a concentration of 4.6 µM (calculated for the monomer), i.e., 0.2 molar equivalents of catalytic disulfides relative to RBI disulfides. After different incubation times, samples were removed and mixed with 0.4 volumes of 30 % (w/v) formic acid. Separation of RBI folding intermediates was performed at 55 °C on an analytical C18 reversed phase column (Vydac, Basel, Switzerland) using a linear gradient from 28 to 45 % (v/v) acetonitrile with 0.1 % trifluoroacetic acid. The elution profile was recorded at 275 nm, and the peaks were analysed by integration with the program Data System MT2 (Kontron Instruments, Bio-Tek, Basel, Switzerland).

Reduction of insulin

Turbidimetric reduction of insulin was performed in 100 mM potassium phosphate, pH 7.0, 2 mM EDTA, 1 mM DTT, 1 µM catalyst and started by addition of insulin to a final concentration of 130 µM. The reaction was followed by light scattering measurements at 650 nm (Holmgren, 1979).

Analytical gel filtration

For determination of the oligomeric state of D-DsbA and DsbC, analytical gel filtration was performed at 25°C and same buffer conditions used for RBI refolding (100 mM Tris/HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA). Protein samples (60 µl) with different concentrations were applied to a Superdex 75 HR 10/30 column (10 x 30 cm) and separated at a flow rate of 20 ml min^{-1}. The elution profiles were recorded at 280 and 226 nm.

In vivo DsbA complementation analysis

Phenotypic screening for DsbA⁺ or DsbA⁻ clones by motility assays (Dailey & Berg, 1993) was performed on soft agar plates in the presence of 1 mM IPTG as described (Hennecke et al., 1997), using the E.coli dsbA deletion strain THZ2 (dsbA::kan,
recA::cam, λmalF-lacZ102). The plasmids pDsbA3, pD-DsbA, pDsbC and pDsbC-C were used for the DsbA complementation studies.
Figure legends

**Figure 1.** Ribbon diagram (Koradi *et al.*, 1996) of the X-ray structure of the oxidized DsbC homodimer from *E. coli* (McCarthy *et al.*, 2000). The catalytic domains (residues 78 to 216) are shown in blue and green, and the dimerization domains (residues 1 to 61) are coloured in red and yellow, respectively. The linker helix (residues 62-77) of subunit 1 is drawn in black, and the arrow indicates the C-terminal end of the linker helix of subunit 1 (peptide bond between residues 77 and 78). The side chains of the active-site cysteines (yellow) are indicated in ball-and-stick representation.

**Figure 2.** Coomassie-stained SDS-PAGE of purified proteins used in this study: DsbC (lane 1), isolated catalytic domain of DsbC (DsbC-C) (lane 2), fusion between the dimerization domain of DsbC and DsbA (D-DsbA) (lane 3) and DsbA (lane 4).

**Figure 3.** Concentration-dependent oligomerization state of DsbC and the artificial dimer of DsbA (D-DsbA), analyzed by analytical gel filtration at 25°C and pH 7.5. Proteins were equilibrated with running buffer for 120 min and applied to a Superose 75 column (10 x 30 cm) at the initial concentrations of 460 µM (red), 46 µM (blue) and 4.6 µM (green). For molecular mass determination a mixture of the following standard proteins ( ■ ) was used: Serum albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa) and ribonuclease A (13.7 kDa) was performed and indicated with arrows. **(A)** Concentration-dependent dissociation of DsbC and comparison with the monomeric, catalytic domain of DsbC ( ● ). **(B)** Concentration-dependent dissociation of the artificial DsbA dimer D-DsbA and comparison with monomeric DsbA ( ● ). The applied sample volume was 50 µl in all experiments. Proteins were diluted about 20-fold during the gel filtration runs.

**Figure 4.** HPLC analysis of RBI refolding at pH 7.5 and 15 °C under strongly oxidizing conditions (40 mM GSSG, 2 mM GSH) favoring formation of nonnative disulfide bonds, in the absence **(A)** and presence of 0.2 molar equivalents of DsbA **(B)**, and dimerized DsbA **(C)**. Reactions were quenched with acid after different reaction times and disulfide folding intermediates of RBI were separated by reversed-
phase HPLC using a water-acetonitrile gradient. R and N denote the HPLC retention times of completely reduced RBI (10 SH groups) and oxidized, native RBI (5 disulfide bonds), respectively. The oxidizing conditions of the reaction leads to a complete loss of peak for the reduced RBI within the manual mixing time.

**Figure 5.** Comparison of the kinetics of RBI refolding under oxidative conditions in the presence of catalytic amounts of DsbC (■), the isolated, catalytic DsbC domain (◀), DsbA (◆), and the artificial DsbA dimer D-DsbA (●) with the uncatalyzed reaction (▲).

**Figure 6.** Turbidimetric assay of insulin reduction (130 µM) by DTT (1 mM), in the absence (▲) and presence of DsbA (◆), dimerized DsbA (●), the isolated, catalytic domain of DsbC (◀) and DsbC (■). The assay was performed at 25 °C and pH 7.0 with a 1 µM concentration of the catalyst.

**Figure 7.** Recovery of motility on soft agar plates of the dsbA deletion strain THZ2 by plasmid-encoded, periplasmic expression of DsbA (A), D-DsbA (B), the catalytic domain of DsbC (C) and DsbC (D). Assays were performed in the presence of IPTG at 37 °C and analyzed after growth for 24 hours after inoculation in the center of the plates. Cells with DsbA+ phenotype (A-C) spread over the culture plate and form a bacterial lawn, and cells lacking DsbA only grow only at the point of inoculation (D).
References


Guddat, L. W., Bardwell, J. C., Zander, T. & Martin, J. L. (1997). The uncharged surface features surrounding the active site of Escherichia coli DsbA are conserved and are implicated in peptide binding. *Protein Sci* 6(6), 1148-56.


Results

Requirement for two active-sites for disulfide isomerase activity


Figure 1
Figure 2

Results

Requirement for two active-sites for disulfide isomerase activity
Figure 3

(A) Absorption vs. Retention volume (ml)

(B) Absorption vs. Retention volume (ml)
Figure 4

A

B

C
Figure 5

Percentage of native RBI vs. Time (min)
Results

Requirement for two active-sites for disulfide isomerase activity

Figure 6

Absorbance at 650 nm

Time (min)
Figure 7

Requirement for two active-sites for disulfide isomerase activity

Results
6. Curriculum vitae

Name: Björn PHILIPPS
Date of birth: March 12, 1973
Place of birth: Hamburg, Germany
Marital status: Married
Nationality: German

Education:

1979 - 1983: Grundschule Windmühlenweg, Hamburg, Germany
1983 - 1992: Humanistic Gymnasium Christianeum, Hamburg, Germany
1992: Final Examination: Abitur

1989 - 1993: Gymnasium Altona and Berufsfachschule für Chemisch-
Technische Assistenz, Hamburg, Germany
1993: Final Examination: Chemisch-technischer Assistant

1993 - 1995: Studies in Chemistry at the University of Tübingen, Germany
1994: Vordiplom in Chemistry
1994 - 1995: Studies in Chemistry at the University of Manchester Institute of
Science and Technology (UMIST), United Kingdom
1995 - 1998: Studies in Chemistry at the University of Freiburg im Breisgau,
Germany
1997 - 1998: Diploma thesis director: Prof. Dr. Georg E. Schulz
1998: Diplom in Chemistry

1995: Practical training in the preclinical research of Sandoz Pharma
AG, Basel, Switzerland
1998: Practical training in the preclinical research of Novartis Pharma
AG, Basel, Switzerland

1998 - 2002: PhD student at the Institut für Molekularbiologie und Biophysik,
Eidgenössische Technische Hochschule Zürich, Switzerland
7. Danksagung


Herrn Prof. Dr. Rudi Glockshuber danke ich für die Betreuung und die Unterstützung meiner Arbeit, sowie die Diskussionsbereitschaft und den mir überlassenen Freiraum.

Herrn Prof. Dr. Donald Hilvert danke ich für die Übernahme des Korreferats.

Mein ganz besonderer Dank geht an Simon Topell für die vielen Anregungen und Diskussionen, sowie die große Hilfe bei der Systemadministration und vieles mehr.

Bei Dr. René Brunisholz bedanke ich mich für die N-terminalen Sequenzierungen, bei Dr. Thomas Denzinger für die Einweisung und Hilfe bei der Massenspektrometrie und bei Eva Niederer für die Zellsortierungen. Bei Verena Eggli bedanke ich mich für die häufige technische Hilfe und die Messungen während meiner gehandikapten Zeit und bei Vreni Graf für die Übernahme des administrativen Ballastes.

Bei Patrick Frei bedanke ich mich für die Mitarbeit im Rahmen seines Vollpraktikums und seiner Diplomarbeit.

Allen meinen jetzigen und ehemaligen Laborkollegen möchte ich für das angenehme Arbeitsklima, die zahlreichen Hilfestellungen und Unterstützungen, sowie für die vielen anregenden Diskussionen danken, ganz besonders aber: Heike Eberl, Dr. Ulla Grauschopf, Dr. Uta Hermanns, Dr. Ekkehard Mössner, Dr. Peter Sebbel, Michael Vetsch, Dr. Gudrun Wildegger und Dr. Eva Zobeley.

Schließlich möchte ich mich bei meiner Familie, meinen Freunden und allen, die direkt oder indirekt zum Erfolg dieser Arbeit beigetragen haben bedanken.

Danke Maresa.