THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 46, pp. 33257–33264, November 16, 2007

Functional Tat Transport of Unstructured, Small, Hydrophilic Proteins*[S]

Received for publication, April 19, 2007, and in revised form, September 10, 2007 Published, JBC Papers in Press, September 11, 2007, DOI 10.1074/jbc.M703303200

Silke Richter[‡], Ute Lindenstrauss[‡], Christian Lücke[§], Richard Bayliss[¶1], and Thomas Brüser[‡2]

From the [‡]Institute of Biology/Microbiology, University of Halle-Wittenberg, Kurt-Mothes-Strasse 3, D-06120 Halle, Germany, the [§]Max Planck Research Unit for Enzymology of Protein Folding, Weinbergweg 22, D-06120 Halle, Germany, and the [¶]Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, United Kingdom

The twin-arginine translocation (Tat) system is a protein translocation system that is adapted to the translocation of folded proteins across biological membranes. An understanding of the folding requirements for Tat substrates is of fundamental importance for the elucidation of the transport mechanism. We now demonstrate for the first time Tat transport for fully unstructured proteins, using signal sequence fusions to naturally unfolded FG repeats from the yeast Nsp1p nuclear pore protein. The transport of unfolded proteins becomes less efficient with increasing size, consistent with only a single interaction between the system and the substrate. Strikingly, the introduction of six residues from the hydrophobic core of a globular protein completely blocked translocation. Physiological data suggest that hydrophobic surface patches abort transport at a late stage, most likely by membrane interactions during transport. This study thus explains the observed restriction of the Tat system to folded globular proteins on a molecular level.

In many organisms, a subset of proteins has to be translocated across biological membranes in a folded state to allow essential folding events, such as cofactor assemblies or oligomerizations, prior to translocation. The Tat[3] system is the general translocation machinery that is adapted to such folded protein substrates (1). It is of prokaryotic origin and widely occurring in archaea and bacteria, and it also plays a pivotal role in the translocation across the thylakoid membranes in plant plastids (2). In proteobacteria and plants, it consists of three membrane proteins, namely TatA, TatB, and TatC (3). Proteins that are translocated by the Tat system are recognized by a usually cleavable N-terminal signal sequence, which contains the eponymous twin-arginine motif, a conserved sequence pattern with two consecutive and almost invariant arginine residues. So far, all evidence supports the view that such a signal sequence at an N terminus of a folded protein is sufficient to target proteins to the Tat translocon (4–7). The Tat system therefore can be regarded as a general transport system that is not specialized to certain individual proteins but rather accepts a very broad range of proteins as substrates. The smallest known Tat substrates are high potential iron-sulfur proteins (HiPIP) of less than 10 kDa (8), and the largest are heterodimeric formate dehydrogenases of nearly 150 kDa (9). This variability shows that the shapes of the transported proteins cannot be very important. Hence, the Tat system extends the spectrum of transportable proteins to those that are folded in the cytoplasm and thus not compatible with the more commonly used Sec transport system.

Several Tat signal sequence-fused Sec substrates, such as *c*-type cytochromes or alkaline phosphatase (PhoA), have been clearly shown to become Tat-compatible only after cytoplasmic folding (6, 7). This requirement for folding supported the hypothesis of a folding quality control at the Tat translocon (7). Others suggested quality control steps prior to translocon binding (10, 11). Nevertheless, Tat-incompatible unfolded twin-arginine signal-fused PhoA (RR-PhoA) has been shown to interact with the Tat system, thereby causing inhibition of growth (12). Translocon binding and growth inhibition suggested that the selectivity for folded proteins at the Tat translocon is not a physiological folding quality control.

We suspected that the bacterial Tat system may not directly detect whether substrate is folded or not and that translocation of unfolded proteins should in principle be possible. To test this hypothesis, we analyzed the Tat targeting of proteins that are not able to fold and thus carry out their physiological function in an unfolded state. We demonstrate here that the bacterial Tat system can indeed translocate such naturally unfolded polypeptides and that folding therefore is not a general prerequisite for Tat transport. Instead, we show that the absence of an exposed hydrophobic core motif is the critical factor that determines whether or not a protein can be translocated by the Tat system.

MATERIALS AND METHODS

Strains and Growth Conditions—Escherichia coli MC4100 (13) and its *tatABCDE*-deficient derivative DADE (14) were used for expressions and physiological studies, and *E. coli* XL1-Blue Mrf′ Kan (Stratagene) was employed for cloning. The bacteria were grown aerobically at 37 °C on LB medium (1% tryptone, 1% NaCl, 0.5% yeast extract) in the presence of the appropriate antibiotics (100 μg/ml ampicillin, 20 μg/ml chloramphenicol).

* This work was supported by Deutsche Forschungsgemeinschaft Grant BR2285/1-3 and grants from the state Sachsen-Anhalt (Exzellenz-Cluster Biowissenschaften) and the Fonds der Chemischen Industrie (to T. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1 and S2.

[1] Supported by the Royal Society and the Institute of Cancer Research.

[2] To whom correspondence should be addressed. Tel.: 49-345-5526360; E-mail: t.brueser@mikrobiologie.uni-halle.de.

[3] The abbreviations used are: Tat, twin-arginine translocation; HiPIP, high potential iron-sulfur protein.

Plasmids and Genetic Methods—For construction of twin-arginine signal sequence-FG repeat fusions, the *hip*-containing fragment of pEXH5-BamHI^{SP} (12) was excised with NdeI and ScaI and cloned into the corresponding sites of pBW22 (15), resulting in pBWH-BamHI^{SP}. The FG repeat-encoding regions of *nsp1p* were amplified with the plasmid pFF18 (16) as template, using the forward primer *nsp1p*-BamHI-F (5′-CAT ATG GAT CCC ATG GCA AAA CCA GCC CTC TC-3′) in combination with either *nsp1p*-XhoI-R3 (5′-CCA AAC TCG AGA GCT GGT TTT GAG GAA TTA TC-3′), *nsp1p*-XhoI-R5 (5′-CCA AAC TCG AGT GCC GGC TTG GAG GTT TCA TTG-3′), *nsp1p*-XhoI-R7 (5′-CCG AAC TCG AGG GCA GGC TTA GAA GTA GCA C-3′), *nsp1p*-XhoI-R10 (5′-CCA AAC TCG AGT GCC GGC TTG GAA GTT TCA TTG-3′), or *nsp1p*-XhoI-R17 (5′-CCA AAC TCG AGC GCT GGC TTG CTC GTA TCA C-3′). The fragments were restricted with BamHI/XhoI and cloned into the corresponding sites of pBWH-BamHI^{SP}, resulting in plasmids encoding C-terminally His_6-tagged fusions of the HiPIP twin-arginine signal sequence with FG repeats of varied lengths, named pBW-R3, pBW-R5, pBW-R7, pBW-R10, and pBW-R17, respectively. For construction of the twin-lysine variants, the twin-arginine codons of the plasmid pBWH-BamHI^{SP} were changed to two lysine codons by QuikChange^{TM} mutagenesis (Stratagene), using the primer pair *hip*-KK-F (5′-CGA TAA GCC AAT CAG CAA GAG CAA GAA AGA CGC TGT CAA AGT GAT G-3′)/*hip*-KK-R (5′-CAT CAC TTT GAC AGC GTC TTT CTT GCT CTT GCT GAT TGG CTT ATC G-3′), resulting in pBW-R3-KK, pBW-R5-KK, pBW-R7-KK, pBW-R10-KK, and pBW-R17-KK.

The codons encoding the hydrophobic core residues NIILLI of PhoA were introduced into the third FG repeat by two subsequent QuikChange^{TM} mutagenesis steps using at first the primer pair *nsp1p*-core-1F (5′-GCT GGT GCT ACT TCT AAG AAT ATC ATC TCA TTC GGT GCT AAG C-3′)/*nsp1p*-core-1R (5′-GCT TAG CAC CGA ATG AGA TGA TAT TCT TAG AAG TAG CAC CAG C-3′) and then the primer pair *nsp1p*-core-2F (5′-CTT CTA AGA ATA TCA TCT TAT TGA TTG CTA AGC CAG AAG AAA AG-3′)/*nsp1p*-core-2R (5′-CTT TTC TTC TGG CTT AGC AAT CAA TAA GAT GAT ATT CTT AGA AG-3′), resulting in the plasmids pBW-R3-Φ, pBW-R5-Φ, pBW-R7-Φ, pBW-R10-Φ, and pBW-R17-Φ. For introduction of the same hydrophobic core into the fifth FG repeat, the primer pair *nsp1p*-core2-1F (5′-CAA GAT GGA ACG GCT AAG AAT ATC ATC TCC TTT GGT GCC AAG CCC-3′)/*nsp1p*-core2-1R (5′-GGG CTT GGC ACC AAA GGA GAT GAT ATT CTT AGC CGT TCC ATC TTG-3′) and then the primer pair *nsp1p*-core2-2F (5′-CGG CTA AGA ATA TCA TCT TAT TGA TTG CCA AGC CCG CGG AG-3′)/*nsp1p*-core2-2R (5′-CTC CGC GGG CTT GGC AAT CAA TAA GAT GAT ATT CTT AGC CG-3′) was used with pBW-R5 as template, resulting in the plasmid pBW-R5-Φ_c. For the alteration of the hydrophobic core to NIIL<u>WL</u>, the primer pair R5-Φ-WL-F (5′-CTT CTA AGA ATA TCA TCT TAT GGC TTG CTA AGC CAG AAG AAA AG-3′)/R5-Φ-WL-R (5′-CTT TTC TTC TGG CTT AGC AAG CCA TAA GAT GAT ATT CTT AGA AG-3′) was used, and for the alteration of the hydrophobic core to NI<u>FV</u>LI, the primer pair R5-Φ-FV-F (5′-GCT ACT TCT AAG AAT ATC TTC GTG TTG ATT GCT AAG CCA GAA G-3′)/R5-Φ-FV-R (5′-CTT CTG GCT TAG CAA TCA ACA CGA AGA TAT TCT TAG AAG TAG C-3′) was used, with pBW-R5-Φ as template, resulting in pBW-R5-Φ-WL and pBW-R5-Φ-FV, respectively. For shortening the hydrophobic core in FG5-Φ, we exchanged codons for LI by codons for ST, using the primer pair FG5-ST-F (5′-CTA AGA ATA TCA TCT TAT CGA CCG CTA AGC CAG AAG AAA AG-3′)/(5′-CTT TTC TTC TGG CTT AGC GGT CGA TAA GAT GAT ATT CTT AG-3′) with pBW-FG5-Φ as template, resulting in pBW-R5-Φ-ST.

For production of HiPIP, we used the plasmid pEXH5-tac (12) and derivatives with one, two, three, or four Cys → Asp mutations. Primers used were as follows: hip-C43D-F (3′-CCC GAA GAG CAG CAC GAC GCC AAC TGC CAG-5′)/hip-C43D-R (3′-CTG GCA GTT GGC GTC GTG CTG CTC TTC GGG-5′), hip-C43,46D-F (3′-GCA GCA CGA CGC CAA CGA CCA GTT CAT GCA GG-5′)/hip-C43,46D-R (3′-CCT GCA TGA ACT GGT CGT TGG CGT CGT GCT GC-5′), hip-C63D-F (3′-GAG TGG AAG GGC GAC CAG CTC TTC CCC GG-5′)/hip-C63D-R (3′-CCG GGG AAG AGC TGG TCG CCC TTC CAC TC-5′), hip-C77D-F (3′-CGT CAA CGG CTG GGA CGC TTC CTG GAC C-5′)/hip C77D-R (3′-GGT CCA GGA AGC GTC CCA GCC GTT GAC G-5′). Arg-Arg → Lys-Lys mutations were generated using the primer pair *hip*-KK-F/*hip*-KK-R (see above), to allow the assessment of Tat dependence of translocation.

For purification and structural analyses, the C-terminally His_6-tagged FG5 and FG5-Φ domains were produced from the arabinose-regulated plasmids pBAD-*nsp*R5-H6 and pBAD-*nsp*R5Φ-H6, which were constructed by PCR amplification with pFF18 as template, using the primer pair *nsp1p*-NcoI-F (5′-GGG ATC CAT GGC AAA ACC AGC CCT C-3′)/*nsp1p*-XhoI-R5, followed by restriction of the fragment with NcoI and XhoI and ligation together with the His_6 tag-containing XhoI-ScaI fragment from pEXH5 (17) into the corresponding sites of pBAD22 (18), which resulted in pBAD-*nsp*R5-H6. For the generation of pBAD-*nsp*R5Φ-H6, the hydrophobic core codons were introduced by QuikChange^{TM} as described above, using the primer pairs *nsp1p*-core-1F/*nsp1p*-core-1R and *nsp1p*-core-2F/*nsp1p*-core-2R. All constructs were confirmed by restriction analyses and sequencing. The vector pABS-*tatABC* was used for co-expression of the *tatABC* genes (19).

Biochemical Methods—SDS-PAGE and protein estimations were carried out by standard methods (20, 21). Subcellular fractionations as well as Western blotting and blot developments were carried out as previously described (8). Antibodies against the HiPIP signal sequence and the hexahistidine tag (17) were used.

For NMR data collection, samples of FG5 or FG5-Φ (0.2 mM each) in 20 mM potassium phosphate buffer (pH 7.0) were measured in Shigemi tubes at 10 °C. All NMR spectra were acquired using a Bruker DMX 500 spectrometer operating at 500.13 MHz proton resonance frequency and equipped with a 5-mm triple resonance $^1H\{^{13}C/^{15}N\}$ probe that had XYZ gradient capability. Homonuclear two-dimensional $^1H/^1H$ TOCSY experiments (22) were collected with the carrier placed in the center of the spectrum on the water resonance, which was suppressed by presaturation. Quadrature detection in the indi-

rectly detected dimension was obtained by the states-time-proportional phase incrementation method. All NMR spectra were acquired and processed on Silicon Graphics computers using the program XWINNMR 3.5 (Bruker Bio-Spin, Rheinstetten, Germany). A 90° phase-shifted squared sine-bell function was used for apodization in all dimensions. Polynomial baseline correction was applied to the processed spectra in the directly detected ^{1}H dimension. The chemical shifts were referenced to external 2,2-dimethyl-2-silapentane-5-sulfonate.

RESULTS

Tat Transport of Natively Unfolded Proteins—In bacteria, proteins that are translocated by the Tat system are believed to require folding prior to translocation (23). A strict folding requirement has not been perceived in the Tat system of plant thylakoids, since the translocation of malfolded proteins has been demonstrated in that system (24). It is therefore an important issue in protein translocation research to address the reasons for the incompatibility of certain unfolded Tat substrates with the bacterial Tat system.

Based on our previous observation of an artificially unfolded protein that is targeted to the Tat translocon (12), we were encouraged to assess the translocation of a natively unfolded protein. Such natively unfolded soluble proteins are very hydrophilic and generally do not fold into any globular conformation. We found that the so-called FG repeat domain of the yeast nuclear pore protein Nsp1p was an ideal tool for our studies. It has been shown in the past by circular dichroism that this polar domain is 80–95% in a naturally unfolded conformation (25). The multiple repeats of 19 amino acids function as an unstructured flexible peptide in the transport through nuclear pores (16, 26–28). The repetitive nature of the FG repeat domain allowed us to construct a series of Tat signal-fused unfolded proteins that differ in length. We analyzed the translocation of 3, 5, 7, 10, and 17 repeats, which were N-terminally fused to the signal sequence of our model Tat substrate HiPIP. The variation of the repeat number served the purpose to examine possible size restrictions in the transport of unfolded proteins by the Tat system. The constructs are shown schematically in Fig. 1*A*. For clarity, the constructs are termed RR-(FG3) to RR-(FG17), where "RR" represents the twin-arginine signal sequence and where the number of fused FG repeats is given in parentheses. All constructs were hexahistidine-tagged to allowed their detection in Western blot analyses as well as the rapid purification and biophysical analyses.

When the translocation of the naturally unfolded Tat substrates was examined, we found a clear Tat-dependent translocation with the RR-(FG3), RR-(FG5), RR-(FG7), and RR-(FG10) constructs (Fig. 1*B*). Significant translocation required co-expression of *tatABC*, indicating Tat dependence of transport. Without co-expression of the *tat* genes, transport was hardly detectable. The Tat dependence of transport was further confirmed by the analysis of transport with a mutated signal sequence, in which the two arginines were substituted by lysines. Such a substitution efficiently and selectively blocks Tat-dependent translocation (29, 30). As expected, the Arg-Arg → Lys-Lys mutation resulted in a block of translocation, as shown for FG5 in Fig. 1*C*, confirming that the translocation of the RR constructs was indeed Tat-dependent. The data thus demonstrate that unfolded proteins can be translocated by the Tat system up to a size of 20–30 kDa. The translocation efficiency decreased with increasing length until, in the RR-(FG17) construct, the translocation was undetectable, indicating some inverse correlation of size and translocation efficiency. As typically observed with recombinant Tat substrates (31), the translocation of the Tat-compatible constructs was not complete, and precursor as well as signal sequence-degraded mature proteins accumulated

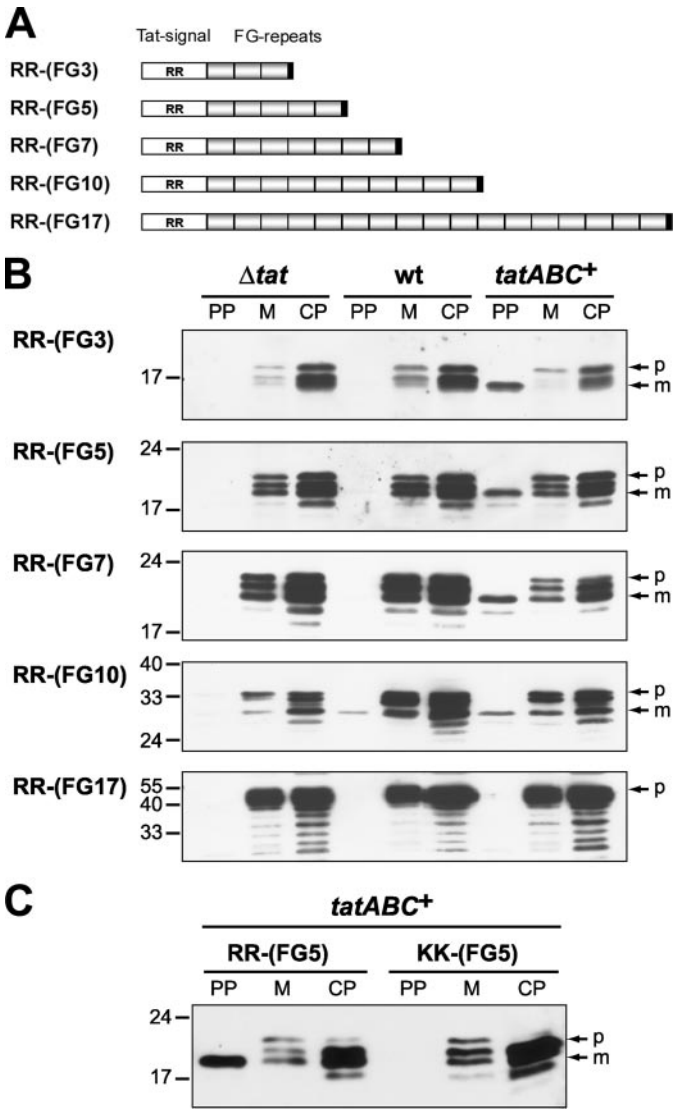

FIGURE 1. **Transport analysis of Tat signal sequence-fused natively unfolded FG-domains.** *A*, schematic representation of the FG domain constructs used in this study. The indicated numbers of FG repeats (*gray*) have been fused to the N-terminal twin-arginine signal sequence of HiPIP (*white*) and C-terminally tagged by a hexahistidine-tag (*black*). *B*, detection of the natively unfolded protein constructs in subcellular fractions of the TatABCDE-deficient strain DADE (Δ*tat*), the parental wild type strain MC4100 (*wt*), and the strain MC4100/pABS-*tatABC* (*tatABC*+), all containing the expression vectors for the natively unfolded Tat substrates indicated to the *left* of the respective blots. *C*, analysis of the twin-arginine dependence of transport. Shown is detection of the twin-arginine signal sequence-fused natively unfolded protein RR-(FG5) and its twin-lysine derivative KK-(FG5) in subcellular fractions after SDS-PAGE and Western blotting using antiserum recognizing the His_6 tag. *PP*, periplasm; *M*, membranes; *CP*, cytoplasm. The positions of precursor proteins (*p*) and the signal peptide-processed mature proteins (*m*) are indicated on the *right*. Molecular mass markers (in kDa) are indicated on the *left*.

in the cytoplasm. The FG repeats were only slightly sensitive to degradation, since predominantly full size and mature size protein was detected. The results thus demonstrate that proteins have to be neither globular nor tightly folded to be compatible with the bacterial Tat system.

Hydrophobic Surface Patches Block Tat Transport—The transport of unfolded FG repeats prompted the question of why the other proteins studied thus far required folding to become compatible with the Tat system. Certainly, some tested proteins, such as RR-PhoA, were too large to be translocated in an unfolded conformation. However, also smaller proteins, such as cytochromes, are reported to be incompatible with the Tat system when they cannot fold properly prior to transport (6). We thus suspected that there are additional characteristics in proteins that are not properly folded that prevent translocation. These characteristics had to be absent in natively unfolded proteins, such as the above described FG repeats, because these polypeptides can be transported by the Tat system.

Globular proteins have hydrophobic core regions that contribute to the stability of the protein fold (32). When globular proteins are prevented from proper folding (*e.g.* by not allowing cofactor assembly or disulfide formations), these proteins can expose hydrophobic core regions to the surface. Such hydrophobic patches are absent on the surface of natively unfolded proteins, such as the FG repeats used in this study, which are very polar and highly charged. It thus seemed feasible to suppose that Tat substrates must be polar on their surface.

We tested this hypothesis by introducing the residues NIILLI from the hydrophobic core of PhoA into the third repeat of the unfolded RR-(FGn) constructs. A comparison of the hydrophobicity of the constructs with and without the artificial hydrophobic core is shown in Fig. 2*A* for RR-(FG5). As expected, the highest hydrophobicity is found in the N-terminal signal sequence. In contrast, the mature domain of the translocated RR-(FG5) construct is hydrophilic overall and shows some periodic fluctuations that are due to the repetitive nature of the construct, with slightly hydrophobic regions alternating with entirely hydrophilic regions. Importantly, the introduction of the PhoA hydrophobic core motif results in a punctually increased hydrophobicity at the insertion site in the third repeat, thereby mimicking an exposed hydrophobic core of an unfolded globular protein (Fig. 2*A*). In the following, we term such hydrophobic core-containing constructs RR-(FGn)-Φ.

When we examined the transport of these constructs, we found that the introduced hydrophobic core region completely blocked Tat-dependent translocation in all cases (Fig. 2*B*). This block of translocation was independent of the total length of the constructs. We then tested whether the position of the hydrophobic core residues within the unfolded FG repeat constructs had any influence on the block of Tat transport. For that purpose, we introduced the same hydrophobic core residues into the fifth repeat of RR-(FG5), which resulted in the construct RR-(FG5)-Φ_c. As in the case of RR-(FG5)-Φ, we could not detect any transport with RR-(FG5)-Φ_c (Fig. 2*C*). In this experiment, the position of the hydrophobic core residues within the unfolded protein was therefore not relevant. The detailed RR-(FG5) sequence and the positions of the introduced hydrophobic core regions in RR-(FG5)-Φ and RR-(FG5)-Φ_c are shown in supplemental Fig. S1. The data indicate that already a small number of surface-exposed hydrophobic residues can result in an abortion of Tat-dependent translocation.

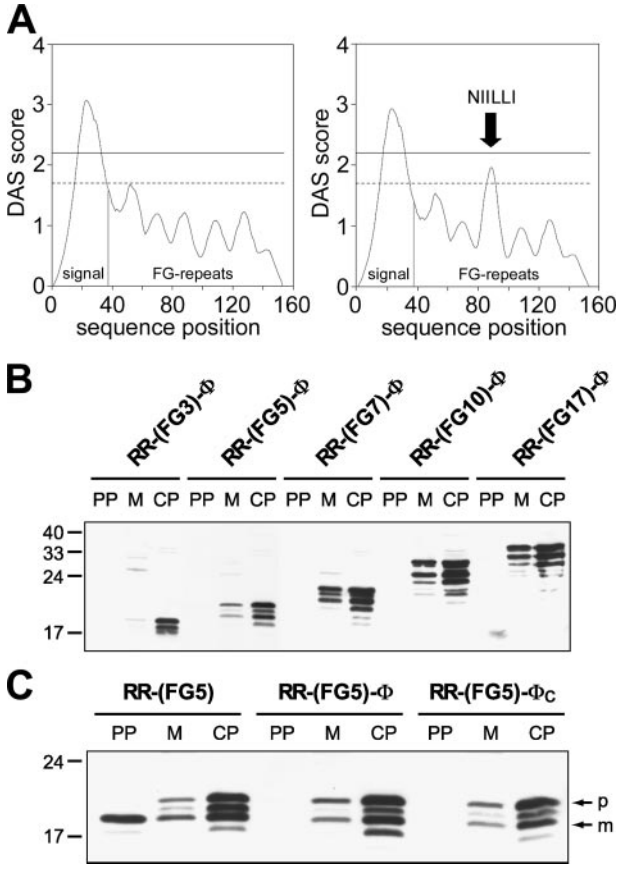

FIGURE 2. **Transport analysis of natively unfolded FG domains with introduced hydrophobic cores.** *A*, comparison of the hydropathy of the RR-(FG5) and RR-(FG5)-Φ constructs by the dense alignment surface (*DAS*) method (43). *B*, transport analyses of the RR-(FG3)-Φ, RR-(FG5)-Φ, RR-(FG7)-Φ, RR-(FG10)-Φ, and RR-(FG17)-Φ constructs in the strain MC4100/pABS-*tatABC* (*tatABC*+). *C*, comparative transport analysis of the RR-(FG5), RR-(FG5)-Φ, and RR-(FG5)-Φ_c constructs. Detection of the constructs in subcellular fractions was as described in the legend to Fig. 1.

Hydrophobic Surface Patches Do Not Affect the Unstructured Conformation of the FG Repeats—We analyzed the structure of the five FG repeats (FG5), which was representative for the transported unfolded proteins. The protein was purified by affinity chromatography and cation exchange chromatography and analyzed by two-dimensional NMR spectroscopy. Strikingly, all amide proton resonances were found within a narrow region of the $^1H/^1H$ TOCSY spectrum between 8.1 and 8.8 ppm, similar to the natively unfolded α-synuclein (33, 34), thus indicating the complete absence of secondary structure in the construct (Fig. 3*A*). This was additionally confirmed by analysis of the side-chain resonances, which for example showed that the resonances of all 15 lysine residues in the construct are essentially superimposed, indicating that there are no structural features present that would cause a variation of these resonances (see supplemental Fig. S2*A*). These results are in full agreement with the published CD spectroscopy data, which showed that the FG repeats are natively unfolded (25, 35).

Next, it was important to compare these data with a construct that was blocked in transport due to the introduction of

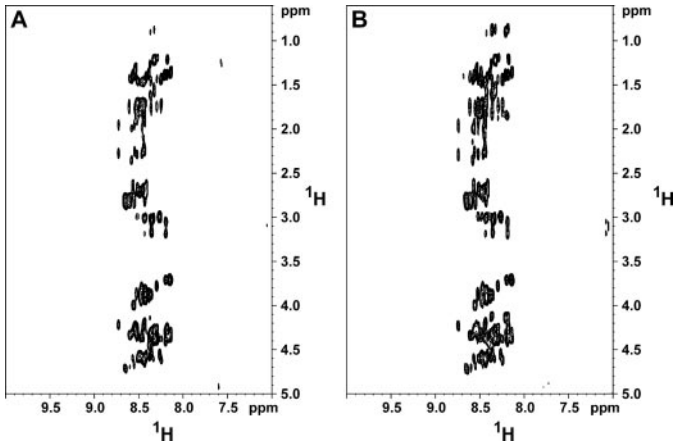

FIGURE 3. **Conformation analysis of FG5 and FG5-Φ by NMR.** Sections from the two-dimensional ^{1}H/^{1}H TOCSY spectra of purified FG5 (*A*) and FG5-Φ (*B*) show the respective amide regions. The rather low spectral dispersion between 8.1 and 8.8 ppm is typically observed in unstructured polypeptides. The most significant differences between FG5 and FG5-Φ are observed near 0.9 ppm, where additional signals occur in the FG5-Φ spectrum due to the methyl resonances of the introduced hydrophobic core.

hydrophobic core residues. We therefore purified and analyzed the FG5-Φ construct in the same way. Strikingly, there was hardly any change in the NMR spectra. Again, all amide signals concentrated between 8.1 and 8.8 ppm (Fig. 3*B*). Few resonance changes occurred due to the six amino acid substitutions, but this had no effect on the unstructured conformation of the construct. Again, this was confirmed by the side-chain resonances of the lysine residues, all of which were found at identical positions (see supplemental Fig. S2*B*). From these experiments, it can be concluded that (*i*) the herein tested translocated proteins indeed adopted a fully unfolded conformation and (*ii*) the introduction of the hydrophobic core residues did not affect the structural properties of the unfolded proteins.

Hydrophobic Surface Patches Affect Growth—It is known that recombinant production of Tat signal sequence-fused PhoA (RR-PhoA) causes an inhibition of the growth of cell cultures (12). This inhibition depends on the twin-arginine motif in the signal sequence, is strongly enhanced by co-expression of the *tatABC* genes, and is most likely due to effects on membrane integrity (12). The reported effect on membrane integrity could be caused by exposed hydrophobic core residues in unfolded RR-PhoA.

When we carried out such growth experiments with a strain that produced the nontransported RR-(FG17) construct and with a strain that produced the twin-lysine derivative of that construct, KK-(FG17), there was hardly any effect on growth upon gene induction, and the minor effects observed clearly did not depend on the twin-arginines (Fig. 4*A*). The results obtained with constructs of the other lengths were the same (data not shown).

However, when we analyzed the effects of RR-(FGn)-Φ derivatives on growth, we found a clear growth inhibition (Fig. 4*B*). After induction, all strains producing RR-(FGn)-Φ derivatives showed a lowered growth rate, independent of the size of the constructs, whereas the twin-lysine derivatives KK-(FGn)-Φ did not show this effect. Obviously, the introduced hydrophobic core residues were harmful for the cells, but only when they

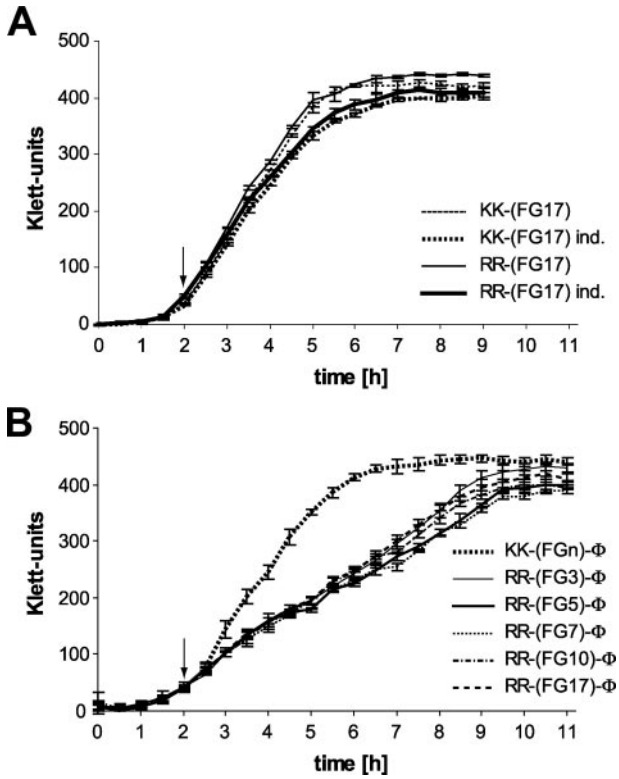

FIGURE 4. **Growth of strains producing Tat signal sequence-fused natively unfolded proteins.** *A*, growth of the strain MC4100/pABS-*tatABC* (*tatABC*+) containing either pBW-R17 or its twin-lysine derivative pBW-R17-KK. Production of the protein constructs was either not induced or was induced by the addition of 0.5% rhamnose at the indicated time point. Cultures were grown at 37 °C in LB medium under aerobic conditions, and culture densities were determined at the given time points in Klett units. *B*, effect of the hydrophobic core residues in the natively unfolded proteins on cellular growth. MC4100/pABS-*tatABC* (*tatABC*+) producing RR-(FG3)-Φ, RR-(FG5)-Φ, RR-(FG7)-Φ, RR-(FG10)-Φ, and RR-(FG17)-Φ or the KK-derivatives of these constructs were grown as described above. All cultures were induced at the indicated time point. KK-(FGn) represents the averaged curve for all strains producing the twin-lysine derivatives of the different sizes, all of which gave the same results. The data points are medians from three cultures. S.D. values are indicated.

were fused to an intact twin-arginine signal sequence. Hence, the tested specific hydrophobic core sequence was not toxic to the cells *per se*. This demonstrates the Tat dependence of the growth inhibition and its dependence on the presence of exposed hydrophobic core residues. Since the hydrophobic residues in RR-(FGn)-Φ constructs were deduced from the PhoA hydrophobic core, and since RR-PhoA caused qualitatively the same but only stronger phenotype, the data indicate that the hydrophobic core residues that cause a block of RR-(FGn)-Φ translocation are harmful for the cells in a Tat-dependent manner. This strongly suggests that the Tat transport of RR-(FGn)-Φ constructs is initiated and aborted at a late stage of translocation, most likely due to stable hydrophobic interactions of the exposed nonpolar residues within the membrane during transport.

The Size but Not the Specific Sequence of the Hydrophobic Patches Is Relevant—Since it was possible that the specific sequence and not the general hydrophobicity of the introduced core residues from PhoA inhibited Tat transport, we mutated the NIILLI pattern in RR-(FG5)-Φ to either NIILWL or NIFVLI and tested the translocation of these constructs. We found that

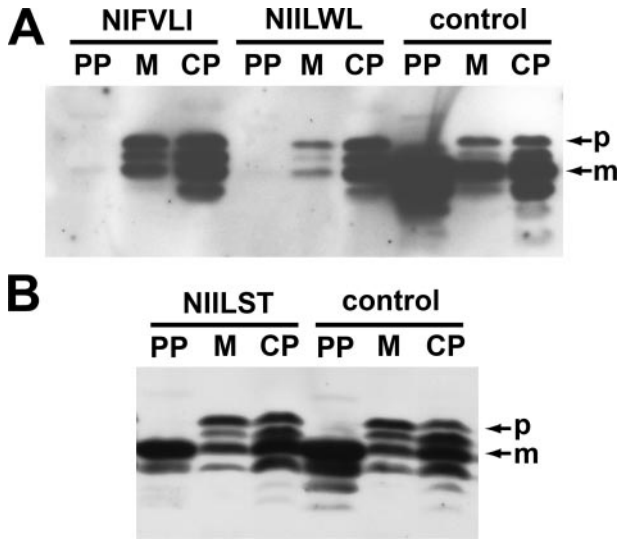

FIGURE 5. **Transport analyses of FG5 constructs with varied hydrophobic cores.** *A*, transport analyses of RR-(FG5)-Φ variants with altered hydrophobic cores NIFVLI or NIILWL in comparison with the transported hydrophilic RR-(FG5) construct. *B*, transport analyses of a RR-(FG5)-Φ variant with only three hydrophobic residues remaining in its hydrophobic core region, NIILST. All analyses in *A* and *B* were carried out in the strain MC4100/pABS-*tatABC* (*tatABC*+) with the respective expression vectors for the natively unfolded substrates. Detection of the constructs in subcellular fractions was as described in the legend to Fig. 1.

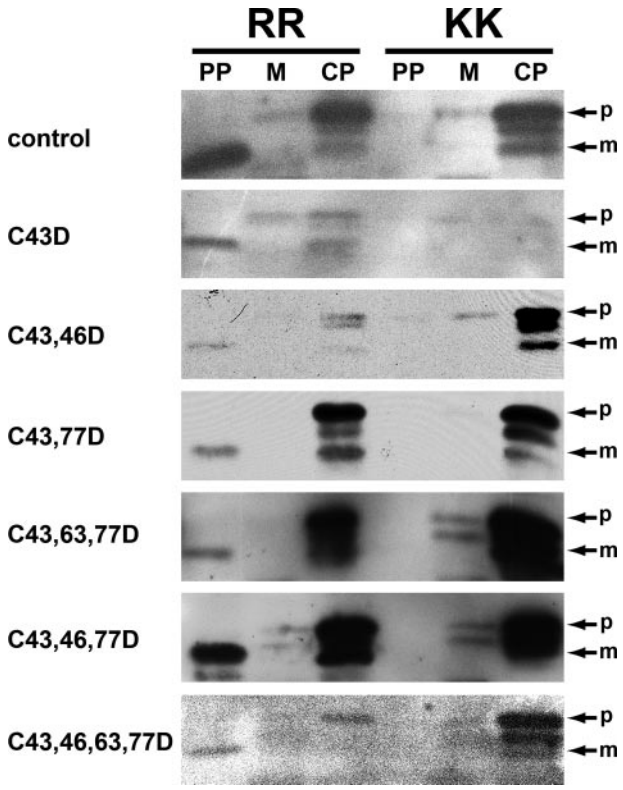

FIGURE 6. **Transport analyses of the Tat substrate HiPIP with folding-preventing mutations of the cofactor ligands.** Transport of native HiPIP (control) and HiPIP with the indicated cysteine mutations was monitored by immunoblots after SDS-PAGE of subcellular fractions as described in the legend to Fig. 1, using antiserum against HiPIP. *RR* and *KK* designate fractionations with HiPIP derivatives containing twin-arginine or twin-lysine signal peptides, respectively.

both altered hydrophobic cores blocked translocation, whereas the hydrophilic RR-(FG5) construct was transported efficiently (Fig. 5*A*). Therefore, the hydrophobicity and not the specific sequence of the introduced hydrophobic core caused the block of translocation.

To gain an idea about the size of hydrophobic patches that could be transported by the Tat system, we then mutated the NIILLI pattern of RR-(FG5)-Φ to NIILST, thereby reducing the number of consecutive hydrophobic residues to three (IIL). Notably, translocation was only weakly affected, indicating that three consecutive hydrophobic residues could not block translocation (Fig. 5*B*). The system therefore is rather tolerant of short hydrophobic segments, but five hydrophobic residues in series are above the tolerated threshold.

Transport of an Unfolded Small Natural Tat Substrate—Having established that artificial Tat substrates can be translocated by the Tat system in a completely unfolded conformation if only they are small and sufficiently hydrophilic, we tested the transport of an unfolded small natural Tat substrate. We chose HiPIP, which is the smallest established Tat substrate so far. HiPIP requires the assembly of a $[4Fe\text{-}4S]^{2+/3+}$ cofactor for folding, since the cluster assembly induces folding, and the loss of the cluster results in unstructured protein (36, 37). In addition, HiPIP has no large hydrophobic core region. We mutated the cysteines that are essential for iron binding to aspartates in order to exclude a cofactor assembly and to introduce repelling charges into the center of HiPIP and tested the translocation of the respective constructs. Tat dependence of transport was confirmed by the analysis of derivatives with Arg-Arg → Lys-Lys mutations in their signal sequence (Fig. 6). Remarkably, all tested folding-incompetent HiPIP derivatives were Tat-dependently translocated, since periplasmic mature protein was detectable only in fractionations of constructs with twin-arginine (RR) signal sequences. Even with very protease-sensitive derivatives, weak bands were detected that indicate a Tat dependent translocation.

DISCUSSION

Within the last few years, studies on bacterial Tat systems resulted in the postulation of a quality control at the Tat translocon, which could prevent the translocation of unfolded proteins (7). The "quality control" hypothesis was mainly based on observations that proteins that were not allowed to fold properly inside the cytoplasm were not translocated by the Tat system (6, 7). However, in the thylakoidal Tat system, a restriction to folded proteins is not presumed, because it has been shown that malfolded Tat substrates can be translocated by this Tat system (24). The data thus implied that there are differences between the plant and bacterial Tat systems with respect to the restriction to folded proteins (38).

We started to assess the restriction of the bacterial Tat system to fully folded protein substrates by a more detailed analysis of the targeting of a nontranslocated malfolded protein. In contrast to what would be expected if a quality control would function at the Tat translocon, we found that this protein bound the Tat machinery, and the machinery most likely attempted unsuccessfully its translocation (12). This resulted in a significant growth inhibition, which was traced back to effects on the membrane integrity. Based on these results, we sus-

pected that no physiologically meaningful quality control takes place at the Tat translocon.

Our present study now reveals the basis for the selectivity at the Tat translocon. We show that, in principle, unfolded proteins *can* be translocated by the Tat system. However, the Tat system is restricted to substrates that are sufficiently hydrophilic on their surface. For most natural Tat substrates, this means that any hydrophobic core regions have to be hidden inside the folded protein. However, the system can tolerate a certain degree of surface hydrophobicity, since the transported unfolded FG repeats contain two separate Phe residues per repeat. Similarly, three consecutive hydrophobic residues hardly affect Tat transport. Preliminary results indicate that the Tat system tolerates hydrophobic patches at the N terminus of mature domains better than at other positions of unfolded proteins.[4]

Taken together, our results now suggest that most natural Tat substrates may not be successfully transported by the Tat system in an unfolded conformation if they expose their hydrophobic core residues to the surface. Most likely, transport of such substrates will be blocked by interactions of these hydrophobic surface patches with the membrane lipids during transport. However, small natural Tat substrates, such as HiPIP, are sufficiently hydrophilic in their unfolded state to be compatible with the Tat system. For these substrates, only efficient cytoplasmic folding can guarantee the transport in a folded state. Slowly folding Tat substrates bind specific factors that most likely ensure complete folding prior to the translocon interaction (11).

We further demonstrate that transport of unfolded Tat substrates becomes inefficient when the polypeptide is too long. This is a very interesting aspect, as it supports the idea that Tat transport does not involve consecutive binding events along a peptide chain. It is well known that tight binding of Tat signal sequences occurs to the TatBC complexes (39), and Tat substrates have only one conserved Tat system binding site: the signal sequence. Further binding events most likely do not occur; therefore, unfolded Tat substrates that are too long cannot be translocated.

Another mechanistic aspect of our results is that a Tat substrate does not need to be globular to become translocated. This finding argues against models of the Tat system that explain transport by pushing mechanisms or by gated pore mechanisms. It is very unlikely that a gated pore mechanism can function with an unfolded protein of over 20 kDa. It is also very unlikely that pushing triggers transport, since the idea of pushing assumes globular Tat substrates. Among the current models for Tat transport, the membrane weakening and pulling model fits best to our observations (10). In this model, the signal sequence is predicted to bind to the TatC subunit of the TatBC complex, TatA then associates with the ternary complex, and this induces a conformational change of TatC, most likely triggered by the proton motive force. This conformational change pulls the protein across the membrane, which is weakened by the assembly of multiple TatA subunits. With globular proteins, one pulling event is likely to transfer the complete substrate across the membrane, which makes a back-diffusion impossible and the transport unidirectional. This model recently gained additional support from the finding that Tat substrates can be translocated even after covalent attachment to TatC (40). It is further supported by reports that demonstrate that TatA assembles with the ternary TatBC-Tat substrate complex (41). In our working model, we explain the observations with unfolded Tat substrates as follows. Transport of the natively unfolded FG repeats by the Tat translocon is unlikely to be completed in one pulling step, at least in the case of longer constructs. The halfway transported protein is most likely released and can escape from the membrane on either side. As we have demonstrated, translocation can be therefore quite efficient with short unfolded Tat substrates. However, when most of the unfolded protein remains on the cytoplasmic side (*i.e.* in the case of long proteins), the probability for a completion of the transport becomes low. When a hydrophobic core is exposed, this blocks transport by interaction with membrane lipids or hydrophobic domains of membrane-integral proteins. The membrane-trapped unfolded proteins probably cause a proton leakage, which results in the lower growth rate (12).

The requirement for hydrophilic surfaces in successfully translocated Tat substrates can be useful for biotechnological applications, such as selections for more soluble proteins (42). However, this mechanistic restriction of the Tat system does not appear to represent a physiological folding quality control, because the unsuccessful attempt to translocate an unfolded protein with hydrophobic surfaces is likely to be energetically more costly than a rare translocation would be. It is further possible that malfolded proteins can be translocated, if only their hydrophobic cores are hidden or small. The translocation of a malfolded protein by the thylakoidal Tat system can thus be explained without postulating major differences between bacterial and thylakoidal Tat systems in this respect (24). Under physiological conditions, factors such as the cytoplasmic folding chaperones are most likely sufficient to ensure that Tat substrates are targeted to the translocon in a folded conformation. Further studies on the aspects of substrate selectivity are vitally important, since they will probably reveal additional features of the translocation mechanism.

Acknowledgment—We are grateful to Jan R. Andreesen for discussions and support.

[4] U. Lindenstrauss and T. Brüser, unpublished results.

REFERENCES

1. Berks, B. C., Palmer, T., and Sargent, F. (2005) *Curr. Opin. Microbiol.* **8,** 174–181
2. Yen, M. R., Tseng, Y. H., Nguyen, E. H., Wu, L. F., and Saier, M. H., Jr. (2002) *Arch. Microbiol.* **177,** 441–450
3. Müller, M., and Klösgen, R. B. (2005) *Mol. Membr. Biol.* **22,** 113–121
4. Thomas, J. D., Daniel, R. A., Errington, J., and Robinson, C. (2001) *Mol. Microbiol.* **39,** 47–53
5. Santini, C. L., Bernadac, A., Zhang, M., Chanal, A., Ize, B., Blanco, C., and Wu, L. F. (2001) *J. Biol. Chem.* **276,** 8159–8164
6. Sanders, C., Wethkamp, N., and Lill, H. (2001) *Mol. Microbiol.* **41,** 241–246
7. DeLisa, M. P., Tullman, D., and Georgiou, G. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6115–6120

8. Brüser, T., Yano, T., Brune, D. C., and Daldal, F. (2003) *Eur. J. Biochem.* **270,** 1211–1221
9. Sargent, F., Berks, B. C., and Palmer, T. (2002) *Arch. Microbiol.* **178,** 77–84
10. Brüser, T., and Sanders, C. (2003) *Microbiol. Res.* **158,** 7–17
11. Jack, R. L., Buchanan, G., Dubini, A., Hatzixanthis, K., Palmer, T., and Sargent, F. (2004) *EMBO J.* **23,** 3962–3972
12. Richter, S., and Brüser, T. (2005) *J. Biol. Chem.* **280,** 42723–42730
13. Casadaban, M. J. (1976) *J. Mol. Biol.* **104,** 541–555
14. Wexler, M., Sargent, F., Jack, R. L., Stanley, N. R., Bogsch, E. G., Robinson, C., Berks, B. C., and Palmer, T. (2000) *J. Biol. Chem.* **275,** 16717–16722
15. Wilms, B., Hauck, A., Reuss, M., Syldatk, C., Mattes, R., Siemann, M., and Altenbuchner, J. (2001) *Biotechnol. Bioeng.* **73,** 95–103
16. Clarkson, W. D., Kent, H. M., and Stewart, M. (1996) *J. Mol. Biol.* **263,** 517–524
17. Brüser, T., Deutzmann, R., and Dahl, C. (1998) *FEMS Microbiol. Lett.* **164,** 329–336
18. Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) *J. Bacteriol.* **177,** 4121–4130
19. Berthelmann, F., and Brüser, T. (2004) *FEBS Lett.* **569,** 82–88
20. Laemmli, U. K. (1970) *Nature* **227,** 680–685
21. Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951) *J. Biol. Chem.* **193,** 265–275
22. Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R. R. (1988) *J. Am. Chem. Soc.* **110,** 7870–7872
23. Lee, P. A., Tullman-Ercek, D., and Georgiou, G. (2006) *Annu. Rev. Microbiol.* **60,** 373–395
24. Hynds, P. J., Robinson, D., and Robinson, C. (1998) *J. Biol. Chem.* **273,** 34868–34874
25. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L., and Rexach, M. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 2450–2455
26. Hurt, E. C. (1989) *J. Cell Sci. Suppl.* **12,** 243–252
27. Bayliss, R., Littlewood, T., and Stewart, M. (2000) *Cell* **102,** 99–108
28. Frey, S., Richter, R. P., and Gorlich, D. (2006) *Science* **314,** 815–817
29. Halbig, D., Wiegert, T., Blaudeck, N., Freudl, R., and Sprenger, G. A. (1999) *Eur J Biochem.* **263,** 543–551
30. Buchanan, G., Sargent, F., Berks, B. C., and Palmer, T. (2001) *Arch. Microbiol.* **177,** 107–112
31. Sturm, A., Schierhorn, A., Lindenstrauss, U., Lilie, H., and Brüser, T. (2006) *J. Biol. Chem.* **281,** 13972–13978
32. Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M., and Regan, L. (1996) *Protein Sci.* **5,** 1584–1593
33. Eliezer, D., Kutluay, E., Bussell, R., Jr., and Browne, G. (2001) *J. Mol. Biol.* **307,** 1061–1073
34. Lücke, C., Gantz, D. L., Klimtchuk, E., and Hamilton, J. A. (2006) *J. Lipid Res.* **47,** 1714–1724
35. Bayliss, R., Kent, H. M., Corbett, A. H., and Stewart, M. (2000) *J. Struct. Biol.* **131,** 240–247
36. Natarajan, K., and Cowan, J. A. (1997) *J. Am. Chem. Soc.* **119,** 4082–4083
37. Bentrop, D., Bertini, I., Iacoviello, R., Luchinat, C., Niikura, Y., Piccioli, M., Presenti, C., and Rosato, A. (1999) *Biochemistry* **38,** 4669–4680
38. Robinson, C., and Bolhuis, A. (2004) *Biochim. Biophys. Acta* **1694,** 135–147
39. Gerard, F., and Cline, K. (2007) *J. Biol. Chem.* **282,** 5263–5272
40. Gerard, F., and Cline, K. (2006) *J. Biol. Chem.* **281,** 6130–6135
41. Dabney-Smith, C., Mori, H., and Cline, K. (2006) *J. Biol. Chem.* **281,** 5476–5483
42. Fisher, A. C., Kim, W., and DeLisa, M. P. (2006) *Protein. Sci.* **15,** 449–458
43. Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997) *Protein Eng.* **10,** 673–676