Size Dependence of Protein Diffusion in the Cytoplasm of Escherichia coli

Nenninger, A; Mastroianni, G; Mullineaux, CW

Size-dependence of protein diffusion in the cytoplasm of *Escherichia coli*

Anja Nenninger, Giulia Mastroianni and Conrad W. Mullineaux*

School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, U.K.

Running title

Protein diffusion in the cytoplasm of *E. coli*

*Correspondence footnote

Correspondence to: C.W. Mullineaux, School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, U.K. tel: +44 20 7882 8440; fax: +44 20 8983 0973; e-mail: c.mullineaux@qmul.ac.uk

Abstract

Diffusion in the bacterial cytoplasm is regarded as the primary method of intracellular protein movement and must play a major role in controlling the rates of cell processes. A number of recent studies have used Green Fluorescent Protein (GFP)-tagging and fluorescence microscopy to probe the movement and distribution of proteins in the bacterial cytoplasm. However, the dynamic behaviour of indigenous proteins must be controlled by a complex mixture of specific interactions, combined with the basic physical constraints imposed by the viscosity and macromolecular crowding of the cytoplasm. These factors are difficult to unravel in studies with indigenous proteins. It has also remained unknown to what extent the addition of a GFP tag might affect the movement of a protein through the cytoplasm. To resolve these problems, we have carried out a systematic study of the size-dependence of protein diffusion coefficients in the *Escherichia coli* cytoplasm, using engineered GFP multimers (from 2 to 6 covalently-linked GFP molecules). Diffusion coefficients were measured using confocal Fluorescence Recovery after Photobleaching (FRAP). At least up to 110 kDa (4 linked GFP molecules), diffusion coefficient varies with size roughly as would be predicted from the Einstein-Stokes equation for a classical (Newtonian) fluid. Thus protein diffusion coefficients are predictable over this range. GFP-tagging of proteins has little impact on the diffusion coefficient over this size-range and therefore need not significantly perturb protein movement. Two indigenous *E. coli* proteins were used to show that their specific interactions within the cell are the main controllers of diffusion rate.

Introduction

The use of fluorescence microscopic techniques to monitor macromolecular diffusion in eukaryotic (HeLa) cells showed that the diffusion of DNA is strongly size dependent, but also that two fluorescently-labelled dextrans (70 kDa and 580 kDa) can diffuse freely in the cytoplasm and nucleus (17). Within bacterial cells such as *E. coli* similar measurements are challenging because of the small dimensions of the cell. Nevertheless, studies of the mobility of fluorescently-tagged proteins are starting to give powerful insights into the dynamics of processes occurring in living bacterial cells. Examples include studies of the mobility of signal transduction proteins in the *E. coli* cytoplasm (23), mobility and distribution of transporters and respiratory complexes in the plasma membrane (15,16), and the dynamic assembly/disassembly of the flagellar motor (14). All these studies depend on the use of cells engineered to express fusion proteins, in which the protein of interest is fused to a fluorescent protein tag, usually a variant of Green Fluorescent Protein (GFP). In many cases, the fluorescent tag is comparable in size, or even larger, than the protein of interest. For example, the chemotaxis signal transducer CheY (14 kDa) was tagged with Yellow Fluorescent Protein (YFP), producing a fusion protein of about 41 kDa (4,23) It remains an open question how much the addition of a substantial fluorescent tag might perturb the mobility of the protein of interest.

The bacterial cytoplasm is a complex, crowded environment (6). The movement of proteins within the cytoplasm must be constrained by a combination of viscosity, macromolecular crowding, and specific interactions of the protein with other cell components (e.g. other proteins, nucleic acids and the cytoplasmic membrane). Any indigenous protein is likely to have specific interactions with other cell components. Therefore it is difficult to dissect out the specific aspects of its behaviour from the more general physical constraints in the cytoplasm. The effects of crowding in the cytoplasm could be complex: for example it is conceivable that macromolecules could form a molecular sieve imposing a distinct size limit on protein mobility. The diffusion of fluorescent proteins in the *E. coli* cytoplasm can

conveniently be measured using Fluorescence Recovery after Photobleaching (FRAP) (7, 12, 19). To resolve the question of the size-dependence of protein diffusion in the *E. coli* cytoplasm, FRAP was used to measure diffusion coefficients for a series of engineered GFP oligomers, ranging in size from 30 kDa (GFP monomers) to 165 kDa (6 linked GFP molecules). The compact barrel-like structure of GFP minimises its interactions with other proteins. Diffusion in the cytoplasm is independent of the type and amount of co-expressed protein and overcrowding of the cytoplasm does not seem to lead to self-interaction of GFP (25). Since GFP is not indigenous to *E. coli* and is unlikely to have specific interactions with other cell components, it can be assumed that the behaviour of GFP oligomers reflects only the simple physical constraints controlling protein movement in the cytoplasm.

Materials and Methods

GFP, vector, and bacterial strain. In all experiments the GFP mut3* (5) was used and the constructs were expressed from the arabinose-inducible pBAD24 vector (8). All constructs were cloned into the *E. coli* strain DH5α (*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*). For a control experiment the MC4100 (*F- araD139 delta(argF-lac)U169 rpsL150 relA1 deoC1 rbsR fthD5301 fruA25 lambda-*) Δ*tatABCDE* (29) strain was used.

GFP multimers. The torA-GFP (GFP tagged at the N-terminus with the signal sequence of TMAO reductase from *E. coli*) construct pJTD1 (28) was used as template for PCR. For torA-GFP2 *torA-gfp* was amplified by PCR to delete the GFP stop codon and to create an EcoRI and KpnI side. The resulting *torA-gfp*$_{(\text{STOP})}$ was cloned into pBad24. The second *gfp* was then amplified with N-terminal asparagine (x5) linker with and without stop codon and inserted via ligation behind *torA-gfp*$_{(\text{STOP})}$ using KpnI and XpaI. The resulting pBad24_torA-GFP2$_{(\text{STOP})}$ vector was used to create *torA-gfp3*. Additional *gfp* (up to GFP5) genes were cloned in frame as described using XbaI-PstI for GFP3, PstI-SphI for

GFP4, and SphI-HindIII for GFP5.A sixth *gfp* was cloned at the end via the HindIII side using pBad24_torA-GFP5$_{(\text{~~STOP~~})}$ as vector and resulting colonies were screened for the right orientation of *gfp6*.

For the pBad24_GFP2 construct the first *gfp* was amplified from pJDT1 without the TMAO signal sequence and stop codon and cloned into pBad24 via EcoRI and KpnI. The second *gfp* was also amplified from pJDT1 with a N-terminal asparagine (x5) linker and cloned behind the first *gfp* using the KpnI and HindIII sides.

<u>AmiA and NlpA constructs.</u> For the two additional constructs *amiA* and *nlpA* were amplified via PCR from genomic DNA from *E. coli* and *gfp* was amplified from pJDT1. To fuse the two PCR products with the *gfp* overlap extension PCR (modified from (24)) was used. In a first PCR, chimeric primers produced overlapping regions at the 5' ends. In a second PCR, external primers were used to generate *amiA-gfp* and *nlpA-gfp* and to create restriction sides. The extended PCR products were cloned into pBad24 using EcoRI and HindIII.

For the modified proteins *amiA* was amplified without the Tat signal sequence (AmiA$_{(\text{noSP})}$). To determine the signal peptide (SP) and the cleavage side the free available prediction software 'SignalP' was used. The first 34 residues containing the twin arginine were deleted and the 'start' codon was moved. For NlpA$_{(\text{noLB})}$ the prediction software 'LipoP' was used to find the lipoprotein signal peptide. As lipoprotein of the plasma membrane lipoprotein 28 requires an aspartame residue in the +2 position after the fatty-acylated cysteine for retention in the plasma membrane. For the shortened construct the lipobox (LB) was deleted and the +2 aspartate was replaced by a methionine.

92 All restriction enzymes (FastDigest®) were purchased from Fermentas. For all PCR steps the PfuUltra
93 II Fusion HS DNA Polymerase from Stratagene was used. Ligation was performed using the Quick
94 Ligase Kit from NEB. For a list of primer sequences see Supplementary Material.
95
96 Growth of cells and sample preparation. Bacterial cultures were cultivated aerobically over night in
97 Luria-Bertani (LB) medium supplemented with ampicillin (50 μg/ml) at 37°C under constant shaking
98 (180 rpm). For measurements the culture was diluted approximately 1:100 into the same media and
99 grown at 37°C under constant shaking. Long non-septated cells were produced by adding the antibiotic
100 cephalexin (10) to a final concentration of 30 μg/ml to a growing culture. Cells were never treated with
101 cephalexin for longer than 120 min. In the case of slower growth, the dilution from the start culture was
102 decreased. GFP expression was induced by adding arabinose in concentrations from 200 μM to 133mM
103 (2%) depending on the construct used, and cultures were grown to mid-exponential phase. A droplet of
104 the culture was spotted onto LB agar plates and cells were allowed to settle down by drying of excess
105 liquid. Small blocks of the agar with the cells adsorbed onto the surface were placed in a laboratory-
106 built sample holder connected to a temperature-controlled circulating water-bath (19). The cells were
107 covered with a glass cover-slip and placed under the microscope objective and samples was maintained
108 at 37°C during FRAP measurements.
109
110 FRAP measurements and data analysis. FRAP measurements were carried out as described and
111 illustrated in (19) using a Nikon PCM2000 laser-scanning confocal microscope equipped with an
112 Argon laser run at 120 mW. For imaging the power was reduced by a factor of 32 using neutral density
113 filters and only the bleach was performed at high laser power. The 488 nm laser line was selected for
114 GFP excitation. Pre- and post-bleach *xy* scans were recorded at 1.64s intervals over an area of 512 x
115 512 pixels corresponding to physical dimensions of either 29 x 29 μm or 58 x 58 μm, depending on the
116 zoom. GFP fluorescence was detected between 500 and 527 nm, selected by an interference band-pass

filter. The FRAP bleach was carried out by switching to *x*-scanning mode and scanning a line across the short axis of the elongated cell, close to the centre of the cell. The laser power was increased by manually raising the neutral density filters for about 1-2s, and after switching back to *xy* mode a series of post-bleach images was recorded. Data analysis was done using the Image-Pro Plus 6.2 software (Media Cybernetics). Pre- and post-bleach images were merged into a sequence file and one-dimensional fluorescence profiles were extracted as a line profile summing data widthways across the cell for the entire experiment. Curve fitting and statistical analyses were performed with SigmaPlot 10.0. The diffusion coefficient D in $\mu m^2\ s^{-1}$ was obtained by fitting to a one-dimensional diffusion equation as described and illustrated in (19).

Cell Fractionation and SDS-PAGE with Western blotting. Cultures were induced with arabinose but not treated with cephalexin. Cells were fractionated according to Randall and Hardy (21) and proteins were separated by SDS-PAGE (10% or 15% polyacrylamide). After electrophoresis the proteins were semi-dry electroblotted onto Hybond-P PVDF membrane (GE Healthcare) and probed with antibody against GFP (Invitrogen). For visualisation of proteins a horseradish peroxidase-conjugated anti-mouse IgG secondary antibody and ECL detection kit (GE Healthcare) were used.

Results

Construction and expression of GFP multimers. Previous measurements from several groups have shown that in the *E. coli* cytoplasm the diffusion coefficient for unmodified GFP is about 10 times lower than in water (see Table 1 for references). The slower diffusion in the cytoplasm may be due to a combination of classical viscosity and macromolecular crowding. To explore constraints on protein diffusion in the *E. coli* cytoplasm, a series of gene constructs expressing linked multimers of GFP were made, based on a torA-GFP construct, which includes an N-terminal signal sequence for the twin-

arginine translocation (Tat) export system (19,28). We used this construct with the intention to explore mobility of the multimers in the periplasm as well as the cytoplasm. However, under our growth and expression conditions in DH5α cells the GFP multimers appeared only in the cytoplasm. Monomeric torA-GFP in the cytoplasm showed a size range from 27 to 30 kDa (Fig. 1). This size range is the result of proteolytic clipping at the N-terminus, where the 30 kDa protein is the chimeric precursor protein, whilst the 27 kDa product is assumed to be the mature GFP (19, 28). The linked GFP multimers all showed a predominant band at the expected molecular weight in the cytoplasmic fraction, but also show some degradation of the multimers into smaller products (Fig. 1).We did not try to quantify these degradation products from the SDS-PAGE because it is difficult to achieve quantitative transfer of proteins over a wide size range (3). Probably for this reason it was also impossible to get a band for the largest multimer (torA-GFP6) (Fig. 1). Note that the smaller degradation products blot more readily and therefore their prevalence is exaggerated in the blot (3).

Effects of GFP overexpression. The pBad24 expression vector system uses the P_{BAD} promoter and shows moderately high expression levels in the presence of the inducer arabinose. However, the level of expression can vary considerably in individual cells. For FRAP measurements, cells were induced with a high level of arabinose (up to 133 mM). For each construct the highest usable concentration was determined microscopically. If the concentration of the inducer is too high, GFP aggregates and/or inclusion bodies become visible (Supplementary Fig.1). When induced with 500 µM arabinose the elongated *E. coli* cell shows a bright and uniform distribution of torA-GFP2 fluorescence in the cytoplasm; overexpression with 1 mM arabinose results in aggregation of torA-GFP2. This problem became even more obvious with the torA-GFP6 constructs, where induction with a range of different arabinose concentrations (from 200 µM – 1.5 mM) led only to either very weak fluorescence or GFP aggregates. The bright, uniform fluorescence needed for FRAP measurements could not be obtained for this construct.

Diffusion of GFP multimers in the cytoplasm. FRAP measurements as previously described and illustrated (19) were used to determine the diffusion coefficients for GFP multimers (from torA-GFP2-torA-GFP5) in the *E. coli* cytoplasm. In all cases, expression levels were chosen to give a bright and uniform loading of GFP fluorescence in the cytoplasm, without detectable aggregates or inclusion bodies. *E. coli* cells were elongated by treatment with cephalexin: this gives highly elongated cells with a continuous cytoplasm containing no diffusion barriers (7, 19). The use of elongated cells allows much more accurate estimation of D because diffusion in the cytoplasm is so rapid that in normal-sized cells fluorescence almost completely re-equilibrates during the bleach (19).

The results from our FRAP measurements are shown and compared with other studies in Table 1. Our results for the GFP multimers show a trend of decreasing mean D with increasing molecular size (illustrated in Fig. 2). An ANOVA test showed that the probability of the null hypothesis (all measured diffusion coefficients are the same) is 0.001 (F-value 6.749), indicating a relationship between size and mobility. Although there is a clear trend of decreasing D with increasing size, the effects of adding one additional GFP molecule up to GFP4 were not significant. Unpaired T-tests gave the following P values: torA-GFP2 compared to torA-GFP3: P = 0.221; torA-GFP3 compared to torA-GFP4: P = 0.441. Only the transition from torA-GFP4 to torA-GFP5 led to a significant decrease in D with P = 0.0025.

The TorA signal peptide does not perturb GFP mobility

Our GFP multimers were assembled with an N-terminal TorA signal peptide. Although this did not result in any detectable translocation to the periplasm under our conditions, it is conceivable that an interaction with the twin arginine translocon of the plasma membrane could influence diffusion. To check this possibility we carried out two controls. Firstly, torA-GFP2 was expressed in the *E. coli* strain MC4100Δ*tatABCDE* (29) lacking the twin arginine translocon. The diffusion coefficient

measured (see Table 1) did not show any significant difference (P = 0.684) to the value obtained for torA-GFP2 in the strain DH5α. We also constructed a GFP dimer (GFP2) lacking the TorA signal peptide and expressed it in DH5α. Again, the diffusion coefficient did not differ significantly from torA-GFP2 in DH5α (Table 1). The Western blots (Fig. 1) suggest efficient cleavage of the TorA signal sequence in the cytoplasm of DH5α, which may explain its lack of influence on diffusion.

Comparison of the size-dependence of D with the predictions of the Einstein-Stokes equation. The Einstein-Stokes equation for diffusion of spherical particles in a classical fluid predicts that D should be inversely proportional to radius:

$$D = \frac{k_B T}{6 \pi \eta a}$$

where 'k_B' is Boltzmann's constant, 'T' is absolute temperature, 'η' is viscosity in PaS and 'a' is the molecular radius. The relation can be used to predict diffusion coefficients for GFP multimers, taking the approximation that mean radius is proportional to $(\text{volume})^{1/3}$, which is in turn proportional to $N^{1/3}$, where N is the number of GFPs in the multimer. It should be noted that the approximation is quite crude as tandem repeats of GFP may behave more like a polymer chain than a globular protein. Values for the diffusion coefficient of monomeric GFP (see Table 1) are very similar to the diffusion coefficient of 9.0 ± 2.1 $\mu m^2 s^{-1}$ measured for torA-GFP (19). The grey shaded area in Fig. 2 shows an extrapolation from the measurement for torA-GFP to give a prediction (± S.D) of D for the GFP multimers. Comparison with the experimental values shows that from GFP1 to GFP4 there is a good match with the predictions of the Einstein-Stokes equation. However, D for GFP5 falls significantly below the value extrapolated from torA-GFP.

Diffusion of GFP-tagged AmiA. For comparison with the diffusion coefficients of GFP multimers, we produced *E. coli* cells expressing GFP-tagged variants of two indigenous periplasmic proteins, AmiA

and Lipoprotein-28 (NlpA). Both proteins have molecular weights similar to GFP, therefore the GFP-tagged variants are comparable in size to GFP2. AmiA is a 31 kDa amidase, which exhibits a predicted signal peptide with the consensus twin-arginine motif (1) at the N-terminus and has been shown to be a Tat substrate for export to the periplasm (11). Western blotting confirmed the predicted size of AmiA-GFP (Fig. 3). GFP can be exported in fluorescent form to the periplasm via the TAT pathway (though not via the Sec pathway) (28). However, when overexpressed from the P_{BAD} promoter, AmiA-GFP is not translocated into the periplasm. There was significant association with the membrane, with the construct partitioned between the membrane and cytoplasmic fractions (Fig. 3). Overexpression of AmiA-GFP with 133 mM arabinose led to significant cell elongation, allowing FRAP measurements to be carried out without cephalexin treatment. A diffusion coefficient of $1.8 \pm 0.8\ \mu m^2 s^{-1}$ (± S.D, n=10) for AmiA-GFP was measured. As a control, we also measured AmiA-GFP diffusion in cephalexin-treated cells, obtaining a value of $1.8 \pm 1.2\ \mu m^2\ s^{-1}$ (± S.D, n=10). The diffusion coefficient for AmiA-GFP is significantly lower than that for the similarly-sized GFP2 construct, which could be due to a tendency of the chimeric protein to associate with the membrane (Fig. 3) via the TAT signal peptide. To test this idea, *amiA* was amplified without the signal sequence and tagged with GFP. For $AmiA_{(noSP)}$-GFP a diffusion coefficient of $7.1 \pm 3.6\ \mu m^2\ s^{-1}$ (± S.D, n=10) was measured with no significant difference (P = 0.488) to the GFP multimer of similar size (GFP2). Overexpression of the modified protein did not lead to cell elongation, so cells had to be treated with cephalexin for FRAP measurements. Western blots revealed no attachment to the plasma membrane for $AmiA_{(noSP)}$-GFP (Fig. 3).

Diffusion of GFP-tagged Lipoprotein-28. The *nlpA* gene encodes a 28 kDa lipoprotein associated with the periplasmic side of the plasma membrane (Lipoprotein 28) (18,31). The estimated size for the chimeric protein of 55 kDa was verified by Western blotting (Fig 3). Overexpressed (133mM arabinose) NlpA-GFP showed an even distribution in the cytoplasm, without translocation to the

periplasm (Fig. 3). The measured diffusion coefficients for this construct showed great variance (Table 1). The overall mean diffusion coefficient was 2.1 ± 1.4 $\mu m^2 s^{-1}$ (± S.D, n=34). Individual cells showed no obvious differences in expression level, size or GFP distribution in the cytoplasm. The images give no clear indication of attachment of NlpA-GFP to the membrane. However, Western blots (Fig. 3) show a significant proportion of NlpA-GFP in the membrane fraction as well as in the cytoplasm.

NlpA has the typical lipobox (LB) [L-[A/S/T]-[G/A]-C (9) and an aspartate residue in the +2 position after the fatty acylated cysteine which is required for plasma membrane lipoproteins (22). In an attempt to modify NlpA-GFP association with the membrane, a modified version ($NlpA_{(noLB)}$-GFP) was produced by amplifying *nlpA* without the lipobox. The +2 aspartate was replaced by a methionine and the modified protein tagged with GFP. Western blotting for the modified protein showed no attachment to the membrane (Fig. 3). When induced with 133mM arabinose the diffusion coefficient for the modified construct showed an even wider range of values in different cells, with a mean of 2.7 ± 3.2 $\mu m^2 s^{-1}$ (± S.D, n=26).

Discussion

The results presented are a systematic study of the effects of protein size on diffusion coefficient in the cytoplasm of *E. coli*, using engineered multimers of GFP expressed from the arabinose-induced pBAD24 vector. Western blots (Fig .1) confirm that the proteins have the expected size, with only limited cleavage in the cytoplasm. For constructs with sizes up to 138 kDa (GFP5) we were able to find levels of induction which gave bright, evenly distributed GFP fluorescence in the cytoplasm, without visible aggregates or inclusion bodies; conditions that allowed the use of confocal FRAP as previously described (18) to measure the diffusion coefficient of the construct. As in some other studies (2, 7, 19) *E. coli* cells were elongated by treatment with cephalexin. The elongation was necessary to obtain accurate diffusion coefficients for constructs diffusing rapidly in the cytoplasm. Cephalexin-treated

cells have a continuous cytoplasm without diffusion barriers (7) and studies on diffusion of monomeric GFP give comparable results with and without cephalexin treatment (Table 1). For GFP6 (165 kDa) we could not obtain good levels of expression without also getting a very inhomogeneous distribution of GFP fluorescence, with much of the protein concentrated into aggregates or inclusion bodies; therefore no diffusion coefficient was obtained for this construct.

Results for diffusion coefficients of the GFP multimers are summarised in Fig. 2, showing a clear dependence of GFP diffusion coefficient on the size of the construct. The diffusion coefficient decreases gradually with increasing size. Although the trend is not severe, it is significant: an ANOVA test gives a probability of only 0.001 that there is no correlation of size and diffusion coefficient. Fig. 2 also indicates the extent to which the size-dependence of the diffusion coefficient conforms to the Einstein-Stokes equation for diffusion in a classical fluid, showing the expected diffusion coefficients for larger proteins extrapolated from the measured diffusion coefficient for GFP1. Up to GFP4 (111 kDa) the mean diffusion coefficient falls close to the Einstein-Stokes prediction for a viscous fluid. This suggests that proteins up to this size do not encounter significant diffusion barriers due to macromolecular crowding or a meshwork of macromolecular structures in the cytoplasm. Note, however, that it is likely that proteins in this size range will encounter size-barriers if the cytoplasm is shrunk and concentrated by osmotic stress (2). The mean diffusion coefficient for the GFP5 construct falls significantly below the expectation from the Einstein-Stokes equation (Fig. 2), and this may provide the first indication of a size-limit for protein diffusion in cells that are not subject to osmotic stress.

Another recent study in *E. coli* indicates a very much steeper reduction in cytoplasmic protein mobility with protein size than we observed (13). The discrepancy could be explained by the nature of the proteins used, since all the larger protein constructs used by Kumar et al. (13) contain elements of

native cytoplasmic *E. coli* proteins. We suggest that it is the specific interactions of these proteins, rather than simply their size, that leads to the drastically slower diffusion of the larger constructs.

Our results indicate that in the absence of specific interactions with other cell components, protein diffusion rates in the *E. coli* cytoplasm are quite predictable, at least within the range from 27-111 kDa. They also suggest that any protein within this size-range that diffuses significantly slower than the expectation shown in Fig. 2 must be slowed by specific interactions with other cell components. The deviation from the expectation in Fig. 2 could be used to estimate the drag due to these interactions. An example from the literature is the CheY chemotaxis signal transduction protein, which, when conjugated with YFP, has a size of about 41 kDa (4 ,23). On the basis of Fig. 5 we would predict a diffusion coefficient of about 8 $\mu m^2 s^{-1}$, thus the measured diffusion coefficient of 4.6 ± 0.8 $\mu m^2 s^{-1}$ (4) suggests drag due to binding to interaction partners in the cytoplasm. To further illustrate this point, the diffusion coefficients for two indigenous *E. coli* proteins tagged with GFP, AmiA and Lipoprotein-28 were determined. Both have mean diffusion coefficients well below the expectation for GFP multimers, and NlpA-GFP additionally shows a much greater variation in diffusion coefficient from cell to cell, indicative of complex and variable interactions in the cell. (Fig. 2). Cell fractionation and Western blotting indicates some interaction with the cell membrane in both cases (Fig. 3). In the case of AmiA we were able to prevent membrane interaction by truncating the protein to remove the TAT signal peptide, and the truncated protein ($AmiA_{(noSP)}$-GFP) showed a diffusion coefficient close to the expectation from GFP multimers (Fig. 2). With Lipoprotein-28, removal of the lipobox led to loss of membrane interaction as judged from cell fractionation and Western blotting (Fig. 3). However, the diffusion coefficient remained low on average and very variable from cell to cell (Fig. 2). AmiA is indigenous to the periplasm and therefore may lack interaction partners in the cytoplasm, leading to rapid diffusion in the cytoplasm once the membrane association is lost. The slow and variable diffusion

316 of Lipoprotein-28 suggests strong interactions in the cytoplasm even though it is indigenous to the
317 periplasmic side of the membrane: interaction with cytoplasmic chaperones is one possibility.
318
319 A useful conclusion from this study is that, for target proteins from 27 - 84 kDa GFP tagging has rather
320 little effect on protein mobility, provided that specific interactions of the protein are not perturbed. The
321 approximation from the Einstein-Stokes equation would suggest that adding a GFP tag should decrease
322 the diffusion coefficient by a maximum of about 13%, in the case of a freely-diffusing target protein of
323 27 kDa. In all other cases (a larger target protein or one with diffusion impeded by specific
324 interactions) the effect of adding the GFP tag will be even smaller. Experimentally, such small changes
325 are usually well within the standard deviation of the measurement. For example, in the measurements
326 reported here, the addition of a single extra GFP molecule never resulted in a statistically-significant
327 decrease in diffusion coefficient over the range from GFP1-GFP4. Thus, observation of GFP-tagged
328 proteins in this size-range can give an accurate picture of the behaviour of the native protein in the *E.*
329 *coli* cytoplasm. However, the results do give an indication of a size-limit between 111 and 138 kDa
330 (Fig. 2). Beyond this size-limit, diffusion of proteins may be more strongly impeded by crowding or
331 meshwork in the cytoplasm, and GFP tags that take a target protein over the size limit are likely to
332 significantly perturb the behaviour of the protein.
333
334 **Acknowledgements**
335 This work was supported by Biotechnology and Biological Sciences Research Council Grant to
336 C.W.M., and also used equipment purchased with a Wellcome Trust grant to C.W.M. We thank Dr
337 Rosemary Bailey and Dr Steven Le Comber for advice on statistics.
338
339 **References**

1. **Berks, B.** 1996. A common export pathway for proteins binding complex redox cofactors? Mol. Microbiol. **22**:393-404.

2. **Bogaart, van den, G., N. Hermans, V. Krasnikov, and B. Poolman.** 2007. Protein mobility and diffusive barriers in *Escherichia coli*: consequences of osmotic stress. Mol. Microbiol. **64**:858-871.

3. **Bolt, M. W., P. A. Mahoney**. 1997. High-efficiency blotting of proteins of diverse sizes following sodium dodecyl sulphate-polyacrylamide gel electrophoresis. Anal. Biochem. **247**:185-192.

4. **Cluzel, P., M. Surette, and S. Leibler.** 2000. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. Science **287**:1652-1655.

5. **Cormack, B. P., R. H. Valdivia, and S. Falkow.** 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene **173**:33-38.

6. **Ellis R. J.** 2001. Macromolecular crowding: an important but neglected aspect of intracellular environment. Curr. Opin. Struct. Biol. **11**:114-119.

7. **Elowitz, M. B., M. G. Surette, P. E. Wolf, J. B. Stock, and S. Leibler.** 1999. Protein mobility in the cytoplasm of *Escherichia coli*. J. Bacteriol. **181**:197-203.

8. **Guzman, L.-M., D. Belin, M. J. Carson, and J. Beckwith.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P-BAD promoter. J. Bacteriol. **177**:4121-4130.

9. **Hutchings, I. M., T. Palmer, D. J. Harrington, and I. C. Sutcliffe.** 2009. Lipoprotein biogenesis in Gram-positive bacteria; knowing when to hold ‘em, knowing when to fold ‘em. Trends Microbiol. **17**:13-21.

10. **Ishihara, A., J. E. Segall, S. M. Block, and H. C. Berg.** 1983. Coordination of flagella on filamentous cells of *Escherichia coli*. J. Bacteriol. 155:228-237.

11. **Ize, B., N. R. Stanley, G. Buchanan, and T. Palmer.** 2003. Role of the *Escherichia coli* Tat pathway in outer membrane integrity. Mol. Microbiol. **48**:183-1193.

12. **Konopka, M.C., K.A. Sochaki, B.P. Bratton, I.A. Shkel, M.T. Record, and J.C. Weisshaar**. 2009. Cytoplasmic protein mobility in osmotically stressed *Escherichia coli.* J. Bacteriol. **191:** 231-237.

13. **Kumar M., M. S. Mimmer, and V. Sourjik.** 2010. Mobility of cytoplasmic, membrane and DNA-binding proteins in *Escherichia coli*. Biophys. J. **98**:552-559.

14. **Leake, M.C., J.H. Chandler, G.H. Wadhams, F. Bai, R.M. Berry, and J.P. Armitage**. 2006. Stoichiometry and turnover in single, functioning membrane protein complexes. Nature **443:** 355-358.

15. **Leake, M.C., N.P. Green, R.M. Godun, T. Granjon, G. Buchanan, S. Chen, R.M. Berry, T. Palmer and B.C. Berks.** 2008. Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by *in vivo* single-molecule imaging. Proc. Nat. Acad. Sci. USA **105:** 15376-15381.

16. **Lenn, T., M.C. Leake, M.C., and C.W. Mullineaux.** 2008. Clustering and dynamics of cytochrome *bd*-I complexes in the *Escherichia coli* plasma membrane *in vivo*. Mol. Microbiol. **70:** 1397-1407.

17. **Lukacs, G. L., P. Haggie, O. Seksek, D. Lechardeur, N. Freedman, and A. S. Verkman.** 2000. Size-dependent DNA mobility in cytoplasm and nucleus. J. Biol. Chem. **275**:1625-1629.

18. **McBroom, A. J., A. P. Johnson, S. Vemulapalli, and M. J. Kuehn.** 2006. Outer membrane vesicle production by *Escherichia coli* is independent of membrane instability. J. Bacteriol. **188**:5385-5392.

19. **Mullineaux, C. W., A. Nenninger, N. Ray, and C. Robinson.** 2006. Diffusion of green fluorescent protein in three cell environments in *Escherichia coli*. J. Bacteriol. **188**:3442-3448.

20. **Potma, E. O., W. P. de Boeij, L. Bosgraaf, J. Roelofs, P. J. M. van Haastert, and D. A. Wiersma.** 2001. Reduced protein diffusion rate by cytoskeleton in vegetative and polarized *Dictyostelium* cells. Biophys. J. **81:**2010–2019.

21. **Randall L. L., and S. L. S. Hardy.** 1986. Correlation of competence for export with lack of tertiary structure of the mature species: a study *in vivo* of maltose-binding in *E. coli*. Cell **46**:921-928.

22. **Robichon, C., D. Vidal-Ingigliardi, and A. P. Pugsley.** 2005. Depletion of apolipoprotein *N*-Acyltransferase cases mislocalization of outer membrane lipoproteins in *Escherichia coli*. J. Biol. Chem. **280**2: 974-983.

23. **Schulmeister, S., M. Ruttorf, S. Thiem, D. Kentner, D. Lebiedz, and V. Sourjik.** 2008. Protein exchange dynamics at chemoreceptor clusters in *Escherichia coli.* Proc. Nat. Acad. Sci. USA **105:** 6403-6408.

24. **Shevchuk, N. A., A. V. Bryksin, Y. A. Nusinovich, F. C. Cabello, M. Sutherland, and S. Ladisch.** 2004. Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. Nucleic Acids Res. **32**: e19.

25. **Slade K. M., R. Baker, M. Chua, N. L. Thompson, and G. J. Pielak.** 2009. Effects of recombinant protein expression on green fluorescent protein diffusion in *Escherichia coli*. Biochem. **48**:5083-5089.

26. **Slade K. M., B. L. Steel, G. J. Pielak, and N. L. Thompson.** 2009. Quantifying green fluorescent protein diffusion in *Escherichia coli* by using continuous photobleaching with evanescent illumination. J. Phys. Chem. B **113**:4837-4845.

27. **Swaminathan R., C. P. Hoang, and A. S. Verkman.** 1997. Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: cytoplasmic viscosity probed by green fluorescent protein translation and rotational diffusion. Biophys. J. **72**:1900-1907.

28. **Thomas, J. D., R. A. Daniel, J. Errington, and C. Robinson.** 2001. Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*. Mol. Microbiol. **39**:47-53.

29. **Wexler, M, F. Sargent, R. L. Jack, N. R. Stanley, E. G. Bogsch, C. Robinson, B. C. Berks, and T. Palmer.** 2000. TatD is a cytoplasmic protein with DNase activity. J. Biol. Chem. **275**:16717-16722.

30. **Yang, F., L.G. Moss, and G.N. Phillips**. 1996. The molecular structure of green fluorescent protein. Nature Biotech. **14**: 1246-1251.

31. **Yu, F., S. Inouye, and M. Inouye.** 1986. Lipoprotein-28, a cytoplasmic membrane lipoprotein from *Escherichia coli*. J. Biol. Chem**. 261:**2284-2288.

Table 1

Diffusion coefficients determined for GFP constructs in the cytoplasm of *E. coli* cells (unless indicated otherwise). Techniques used were FRAP and photoactivation of a red-emitting fluorescence state of GFP (7), Fluorescence Correlation Spectroscopy (4), confocal FRAP (13,19), pulsed-FRAP (2) and continuous photobleaching with evanescent illumination (26).

Protein	**Mw (kDa)**	**D ($\mu m^2 s^{-1}$)**	**treatment**	**Ref.**
GFP in water	27	87		27
GFP	27	7.7 ± 2.5	induced with 100 µM IPTG	7
GFP	27	3.6 ± 0.7	induced with 1 mM IPTG	7
eYFP	26.5	7.08 ± 0.3		13
GFP-his_6	27+	4.0 ± 2.0		7
cMBP-GFP	72	2.5 ± 0.6		7
CheY-GFP	41	4.6 ± 0.8		4
CFP-CheW-YFP	71	1.5 ± 0.05		13
CFP-CheR-YFP	86.2	1.7 ± 0.05		13
torA-GFP	30	9.0 ± 2.1	cephalexin	19
GFP	27	9.8 ± 3.6	cephalexin	2
GFP	27	0.4 ± 0.3	after osmotic upshock with NaCl	2
GFP	27	6.3 ± 1.1		26
GFP	27	3.1 ± 1.0	after osmotic shock	26
torA-GFP2 in DADE	57	7.5 ± 3.9	cephalexin, 2 % arabinose	This study
GFP2	27	9.1 ± 5.1	cephalexin	This study
torA-GFP2	57	8.3 ± 4.2	cephalexin, 500 µM arabinose	This study
torA-GFP3	84	6.3 ± 2.6	cephalexin, 200 µM arabinose	This study
torA-GFP4	111	5.5 ± 1.9	cephalexin, 1 mM arabinose	This study
torA-GFP5	138	2.8 ± 1.5	cephalexin, 800 µM arabinose	This study
AmiA-GFP	58	1.8 ± 0.8	2 % arabinose	This study
AmiA-GFP	58	1.8 ± 1.2	cephalexin, 2 % arabinose	This study
$AmiA_{(noSP)}$-GFP	58	7.1 ± 3.6	cephalexin, 2 % arabinose	This study
NlpA-GFP	55	2.1 ± 1.4	cephalexin, 2 % arabinose	This study
$NlpA_{(noLB)}$-GFP	55	2.7 ± 3.2	cephalexin, 2 % arabinose	This study

LEGENDS TO FIGURES

Figure 1

Size of GFP-tagged constructs expressed in *E . coli* DH5α. Proteins of the cytoplasmic fraction were separated on a 10% denaturing SDS-PAGE and immunoblotted using antibodies to GFP. Lanes are denoted (A) torA-GFP, (B) torA-GFP2, (C) GFP2, (D) torA-GFP3, (E) torA-GFP4, (F) torA-GFP5, (G) torA-GFP6, and (H) empty vector.

Figure 2

Diffusion coefficients for GFP-tagged proteins in the *E. coli* DH5α cytoplasm. Mean diffusion coefficient ± S.D. is shown. "GFPn" are multimers of torA-GFP from GFP1 - GFP5 (GFP1 on the left, through to GFP5 on the right), as described in the text. The line shows predicted D (± S.D. shown by the grey shaded area) estimated by using the Einstein-Stokes equation to extrapolate from data for GFP1 (19) to larger proteins. Note that GFP multimers up to GFP4 show diffusion coefficients in line with this prediction.

Figure 3

Cellular location of AmiA-GFP, NlpA-GFP and modified forms of these proteins lacking the TatA signal peptide ($AmiA_{(noSP)}$-GFP) and the lipobox ($NlpA_{(noLB)}$-GFP). *E. coli* DH5α cells expressing the constructs were fractionated into periplasmic (P), membrane (M) and cytoplasmic (C) fractions and Western blots performed with anti-GFP antibody.

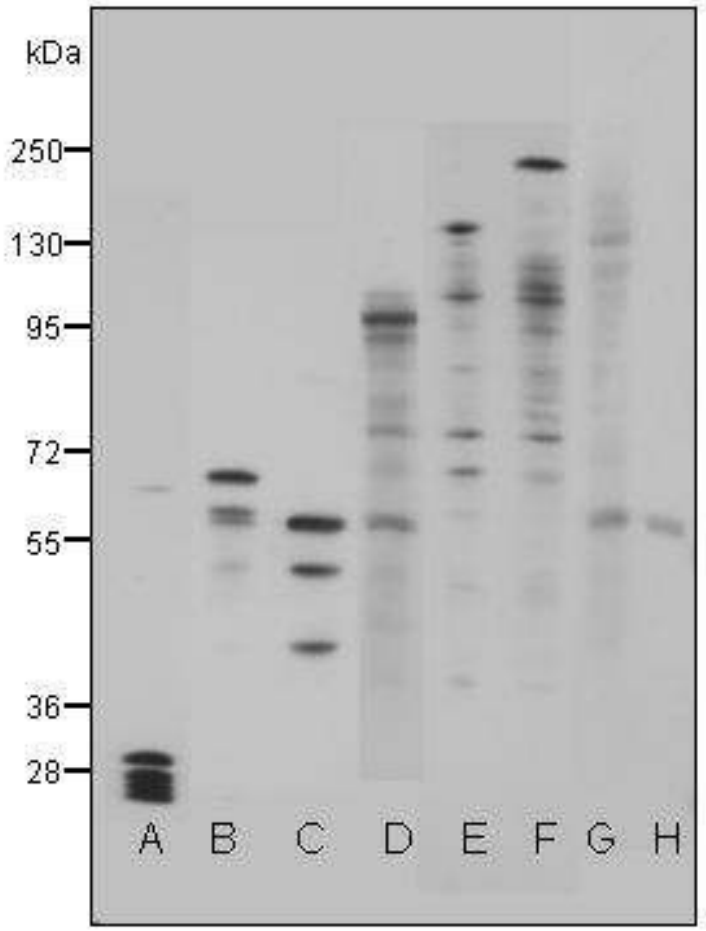

Figure 1

Size of GFP-tagged constructs expressed in *E . coli* DH5α. Proteins of the cytoplasmic fraction were separated on a 10% denaturing SDS-PAGE and immunoblotted using antibodies to GFP. Lanes are denoted (A) torA-GFP, (B) torA-GFP2, (C) GFP2, (D) torA-GFP3, (E) torA-GFP4, (F) torA-GFP5, (G) torA-GFP6, and (H) empty vector.

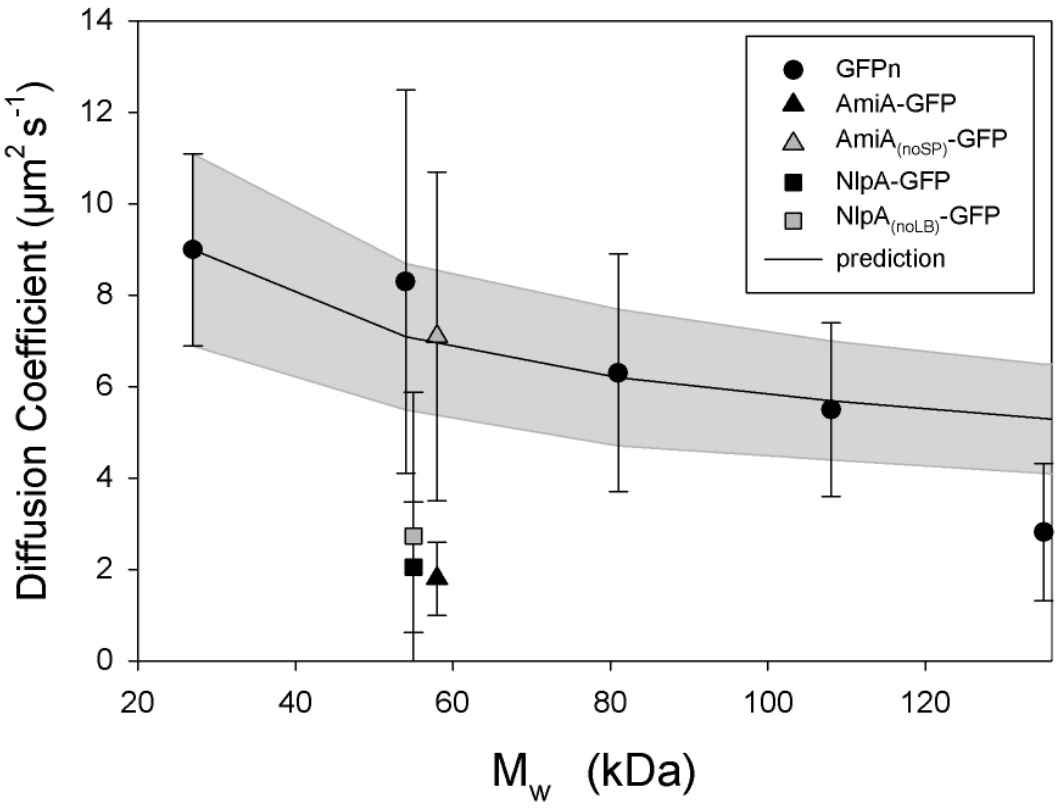

Figure 2

Diffusion coefficients for GFP-tagged proteins in the *E. coli* DH5α cytoplasm. Mean diffusion coefficient ± S.D. is shown. "GFPn" are multimers of torA-GFP from GFP1 - GFP5 (GFP1 on the left, through to GFP5 on the right), as described in the text. The line shows predicted D (± S.D. shown by the grey shaded area) estimated by using the Einstein-Stokes equation to extrapolate from data for GFP1 (19) to larger proteins. Note that GFP multimers up to GFP4 show diffusion coefficients in line with this prediction.

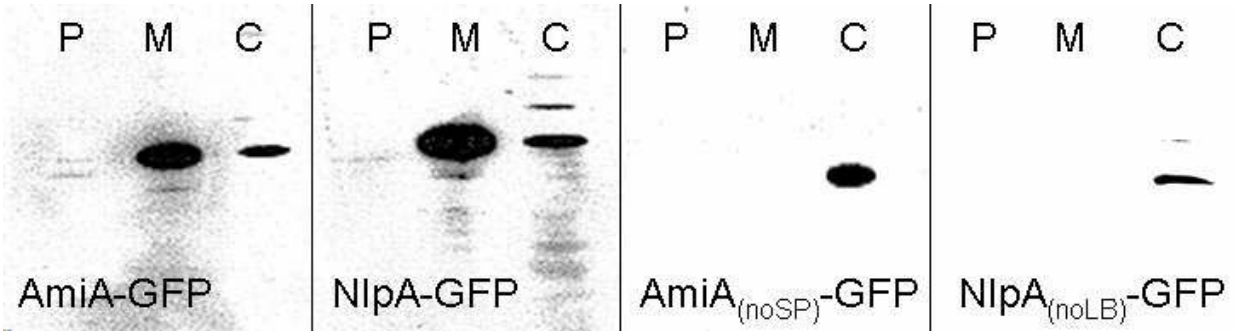

Figure 3

Cellular location of AmiA-GFP, NlpA-GFP and modified forms of these proteins lacking the TatA signal peptide ($AmiA_{(noSP)}$-GFP) and the lipobox ($NlpA_{(noLB)}$-GFP). *E. coli* DH5α cells expressing the constructs were fractionated into periplasmic (P), membrane (M) and cytoplasmic (C) fractions and Western blots performed with anti-GFP antibody.