Peptide Signals Encode Protein Localization[▽]

Jay H. Russell and Kenneth C. Keiler*

Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802

Received 20 June 2007/Accepted 23 August 2007

Many bacterial proteins are localized to precise intracellular locations, but in most cases the mechanism for encoding localization information is not known. Screening libraries of peptides fused to green fluorescent protein identified sequences that directed the protein to helical structures or to midcell. These peptides indicate that protein localization can be encoded in 20-amino-acid peptides instead of complex protein-protein interactions and raise the possibility that the location of a protein within the cell could be predicted from bioinformatic data.

Many proteins have precise intracellular addresses in bacteria, i.e., they are restricted to an area measured in tens of nanometers and not merely to domains such as the cytoplasm or membrane (19, 27, 30). Bacterial protein localization can change in response to external stimuli or internal genetic programs, and essential processes such as cell division and differentiation depend on correct protein localization (13, 23). It is clear that structural proteins must be localized to form functional assemblies, such as flagella, pili, and the cytokinetic Z ring. Yet, many regulatory proteins have a specific intracellular address as well. Studies of several bacterial species, including *Caulobacter crescentus*, have revealed the localization of histidine kinases and transcription factors, and in many cases localization is required for proper function of these proteins in vivo (19). In addition, the identification of bacterial analogs of tubulin (12, 26), actin (20, 36), and intermediate filaments (2) indicates that bacteria have a cytoskeleton and suggests that there may be a high degree of intracellular organization (32).

A priori, there are several possible ways that protein localization could be specified. Each protein could be localized through a unique set of interactions with other localized proteins. This mechanism has been observed for some chemoreceptors (22), proteins associated with the FtsZ ring at the cell division plane (13), and a histidine kinase (28). Alternatively, protein localization could be mediated by global sorting pathways in which localized proteins contain sequence or structural information that targets them to a particular site through interactions with general sorting machinery or cytoskeletal filaments. This type of mechanism is used for nuclear localization in eukaryotes (18, 25) and for protein secretion in bacteria (5). Any global mechanism would require homologous sequence or structural features, but protein localization signals have not been reported.

One simple mechanism to direct protein fate is through peptide signal sequences. The N-end rule and tmRNA peptide signals target proteins for degradation (21, 35), and N-terminal signal sequences target proteins to the Sec or Tat (twin-arginine translocation) pathway for secretion (6, 10). In each of

these examples, the pathways were uncovered by genetic studies, first to identify the *cis* information that targets the substrate to the pathway and then to identify the *trans*-acting components of the cellular export machinery. In principle, peptide-encoded information could also target proteins to a specific location. Here, we test whether short peptides contain sufficient information to localize proteins and find that multiple complex protein localization patterns can be produced by 20-amino-acid peptides.

MATERIALS AND METHODS

Strains and plasmids. The *C. crescentus* strain used in this work was CB15N (14). A plasmid library was constructed by amplifying the *egfp* gene (Clontech) by PCR using the primers GGAATTCATATG(N)₆₀GTGAGCAAGGGCGAGG AGCTGTTTC and CTAGTTCTAGATTACTTGTACAGC and cloning the product into pJS14 (33) under control of a xylose-inducible promoter (24) to generate *N₆₀-egfp*. *Escherichia coli* cells were transformed with the plasmid library and grown as described previously (9), resulting in $\sim 2 \times 10^5$ colonies. *E. coli* colonies were pooled, and plasmid DNA was prepared and mobilized into *C. crescentus* by electroporation. To generate LP2-mCherry, the mCherry coding sequence was cloned into pJS14, and a DNA fragment containing the xylose-inducible promoter and the LP2 sequence was amplified by PCR and cloned into a BamHI site at the 5' end of mCherry. Due to the BamHI site, the LP2-mCherry protein contains a glycine-serine linker between LP2 and mCherry.

Screen for localization signals. Cells from the *C. crescentus* library were grown to an optical density at 660 nm of 0.30 to 0.40 in PYE medium (11) with 0.3% xylose to induce expression of *N₆₀-egfp* and visualized using epifluorescence microscopy. Localization patterns were confirmed by growing independent cultures in PYE with xylose, or in M2X (24), and examining the cells with epifluorescence microscopy. In all cases, the localization patterns were unchanged in different media. Plasmid DNA was prepared from cells containing localized fluorescence signal and sequenced to identify the N-terminal peptide fused to green fluorescent protein (GFP). The effect of A22 was assayed by adding the inhibitor to a final concentration of 10 μ g/ml for 1 hour prior to observation.

Microscopy. Cells were immobilized on 1% agar and imaged using an Eclipse E600 microscope (Nikon) with a 100 $\times$ Plan Fluor oil numerical aperture 1.30 objective in conjunction with a CoolSNAP *fx* charge-coupled device camera (Photometrics) controlled by Image-Pro Discovery software (Media Cybernetics). Images were deconvolved using the 2D Blind Deconvolution algorithm in AutoQuant software (Media Cybernetics).

For optical sectioning, an IX70 laser scanning confocal microscope with a 100 $\times$ UPlanFL oil numerical aperture 1.3 objective and Fluoview software (Olympus) was used to obtain a Z-series of confocal images at increments of 0.15 μ M. Each series of optical sections was deconvolved through 30 iterations using AutoQuant software.

* Corresponding author. Mailing address: 401 Althouse, University Park, PA 16827. Phone: (814) 863-0787. Fax: (814) 863-7024. E-mail: kkeiler@psu.edu.

[▽] Published ahead of print on 31 August 2007.

FIG. 1. Peptides can specify discrete locations within the bacterial cell. *C. crescentus* cells producing GFP (A) or GFP with a 20-amino-acid peptide at the N terminus (B to D) were imaged by epifluorescence (left panels) or differential interference contrast (right panels) microscopy. The position of the focus of fluorescence in cells producing LP3-GFP, shown by schematic diagram (D, bottom left), was measured as a fraction of the distance from the stalked pole (fractional distance = 0) to the flagellar pole (fractional distance = 1.0), and a histogram is shown (D, bottom right).

RESULTS AND DISCUSSION

Identification of peptide localization signals. To screen for peptides that contain a localization signal, a library of genes containing 20 redundant codons fused to the 5' end of a gene encoding GFP was expressed in *C. crescentus*, and the cells

FIG. 2. LP1-GFP localizes to two helices in *C. crescentus*. Confocal microscopy was used to acquire a Z-series of fluorescence images of *C. crescentus* producing LP1-GFP. Schematic diagrams indicate the relative position of the optical section. A differential interference contrast image (top) shows the cell boundaries.

were screened for localized GFP fluorescence. Cells producing GFP without an N-terminal peptide extension had fluorescence distributed evenly throughout the cell (Fig. 1A), as did the vast majority of clones from the library, demonstrating that GFP has no inherent localization signals. Five of the ~3,000 clones had specific patterns of fluorescence within the cells, and three of these were investigated in more detail.

Epifluorescence images of cells producing LP1-GFP showed a pattern of three to four bands extending across the short axis of the cell (Fig. 1B). Similar patterns have been observed in fluorescence images by proteins localized in a three-dimensional helical structure (1, 8, 16, 31, 32, 34). To determine if LP1-GFP was localized in a three-dimensional structure, optical sections were obtained by taking confocal images at 0.15- μ m increments along the z axis (Fig. 2). These optical sections revealed two helices of fluorescence spanning the length of the cell.

The epifluorescence images of cells producing LP2-GFP showed an array of spots (Fig. 1C), a pattern which is also consistent with a three-dimensional structure (1, 8, 16, 31, 32, 34). Optical sections of these cells revealed a helix of GFP fluorescence along the long axis of the cell (Fig. 3). The pattern produced by LP2-GFP is clearly distinct from the pattern produced by LP1-GFP, demonstrating that 20-amino-acid peptides target GFP to at least two different helical structures in the cell.

A third fusion protein, LP3-GFP, produced a single focus of fluorescence signal near the middle of each cell (Fig. 1D). To more precisely determine the position of the LP3-GFP focus, the distance from the stalked pole was measured in 54 cells. In 87% of the cells the focus was one-fourth to one-half of the distance from the stalked pole to the flagellar pole (Fig. 1D). Similar patterns have been observed for the chromosomal replication machinery and proteins required for cytokinesis and chromosome segregation (27, 32), although it is not clear that LP3-GFP colocalizes with any of these structures.

Helices do not require MreB filaments. Helical patterns have been observed for the actin-like protein MreB in several species of bacteria, including *C. crescentus* (16, 32). When LP1-GFP or LP2-GFP was produced in cells that contained MreB fused to the red fluorescent protein mCherry, the GFP and mCherry fluorescence signals were not colocalized (data not shown), indicating that the LP1 and LP2 peptides were not directing GFP to the MreB helix. The MreB filaments can be disrupted *in vivo* by the addition of the small molecule A22 (17). When A22 was added to cells producing LP1-GFP or LP2-GFP, the GFP helices remained intact for more than 1 hour (Fig. 4). Under the same conditions, the mCherry-MreB signal was completely dispersed. These results confirm that localization mediated by LP1 and LP2 does not require MreB filaments. Electron tomography studies of *C. crescentus* showed several MreB-independent filaments, although the composition of these filaments has not been determined (7). LP1 and LP2 may be targeting GFP to one of these filaments or to a previously unobserved structure.

Localization is independent of GFP. It is likely that GFP does not contribute to the localization signal in the LP fusion proteins, because GFP with no N-terminal peptide has no localization and the vast majority of 20-mer peptides do not induce localization, whereas the three LP peptides each pro-

FIG. 3. LP2-GFP localizes to a helix in *C. crescentus*. Confocal microscopy was used to acquire a Z-series of fluorescence images of *C. crescentus* producing LP2-GFP. Schematic diagrams indicate the relative position of the optical section. A differential interference contrast image (top) shows the cell boundaries.

FIG. 4. LP peptide-directed protein localization is not affected by the MreB inhibitor A22. *C. crescentus* cells producing mCherry-MreB (A) or GFP with a localizing peptide at the N terminus (B and C) were imaged by epifluorescence (left panels) and differential interference contrast (right panels) microscopy before (–A22) and after (+A22) treatment with A22 for 1 h.

duce a different pattern. To determine if other proteins can also be localized by LP signals, LP2 was fused to mCherry, which has little sequence similarity to GFP (29). Localization of LP2-mCherry was indistinguishable from that of LP2-GFP (Fig. 5), demonstrating that the information contained in the LP2 sequence is sufficient for localization of two distinct proteins.

Protein localization produced by locons. The precise localization patterns produced by the LP peptides demonstrate that protein localization can be encoded in a short amino acid sequence. We suggest naming these signals “locons,” by analogy with the “degron” degradation signals (37). Although the possibility that the LP1 and LP2 peptides polymerize into filaments has not been excluded, the small size of these peptides limits the surface area available for protein-protein interactions and makes polymerization improbable. The sequences of the LP peptides suggest that they do not interact with known protein secretion systems. The signal sequence prediction programs PSORT (15), SignalP (3), and TatP (4) all predict that each of the LP-GFP fusion proteins is cytoplasmic. It is more likely that the localizing peptides bind to unknown structures

FIG. 5. LP2 localizes mCherry to the same pattern as GFP. LP2-GFP and LP2-mCherry were imaged by epifluorescence (left panels) and differential interference contrast (right panels) microscopy. The localization patterns are indistinguishable, indicating that LP2 is sufficient to localize two different proteins to the same pattern in *C. crescentus*.

in the cell or interact with an uncharacterized protein sorting system. No *C. crescentus* protein contains an exact match to one of the complete LP peptides, but many proteins have seven to nine residues in common with one of the peptides, raising the possibility that localization signals for endogenous proteins are contained in a shorter peptide sequence or are partially redundant. The discovery and further characterization of the LP localization signals will permit the placement of biophysical probes and other molecules at particular sites within the cell and allow protein localization to be predicted from bioinformatic data.

ACKNOWLEDGMENTS

Thanks to Elisabeth Mahen for help with strain construction; to Elaine Kunze and Nicole Bem at the Center for Quantitative Cell Analysis at the Huck Institutes of the Life Sciences, Penn State University, for assistance with confocal microscopy and visualization software; to Zemer Gitai for providing the mCherry-GFP strain; and to Patrick Viollier for providing A22.

This work was supported by NIH grant GM68720.

REFERENCES

1. Addinall, S. G., and J. Lutkenhaus. 1996. FtsZ-spirals and -arcs determine the shape of the invaginating septa in some mutants of *Escherichia coli*. *Mol. Microbiol.* **22**:231–237.
2. Ausmees, N., J. R. Kuhn, and C. Jacobs-Wagner. 2003. The bacterial cytoskeleton: an intermediate filament-like function in cell shape. *Cell* **115**:705–713.
3. Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**:783–795.
4. Bendtsen, J. D., H. Nielsen, D. Widdick, T. Palmer, and S. Brunak. 2005. Prediction of twin-arginine signal peptides. *BMC Bioinformatics* **6**:167.
5. Benson, S. A., M. N. Hall, and T. J. Silhavy. 1985. Genetic analysis of protein export in *Escherichia coli* K12. *Annu. Rev. Biochem.* **54**:101–134.
6. Berks, B. C., T. Palmer, and F. Sargent. 2005. Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Curr. Opin. Microbiol.* **8**:174–181.
7. Briegel, A., D. P. Dias, Z. Li, R. B. Jensen, A. S. Frangakis, and G. J. Jensen. 2006. Multiple large filament bundles observed in *Caulobacter crescentus* by electron cryotomography. *Mol. Microbiol.* **62**:5–14.
8. Campo, N., H. Tjalsma, G. Buist, D. Stepniak, M. Meijer, M. Veenhuis,

- M. Westermann, J. P. Muller, S. Bron, J. Kok, O. P. Kuipers, and J. D. Jongbloed. 2004. Subcellular sites for bacterial protein export. *Mol. Microbiol.* **53**:1583–1599.
9. Cheng, L., T. Naumann, A. Horswill, S.-J. Hong, B. Venters, J. Tomsho, S. Benkovic, and K. Keiler. 2007. Discovery of antibacterial cyclic peptides that inhibit the ClpXP protease. *Protein Sci.* **16**:1535–1542.
10. Danese, P. N., and T. J. Silhavy. 1998. Targeting and assembly of periplasmic and outer-membrane proteins in *Escherichia coli*. *Annu. Rev. Genet.* **32**:59–94.
11. Ely, B. 1991. Genetics of *Caulobacter crescentus*. *Methods Enzymol.* **204**:372–384.
12. Erickson, H. P., D. W. Taylor, K. A. Taylor, and D. Bramhill. 1996. Bacterial cell division protein FtsZ assembles into protofilament sheets and minirings, structural homologs of tubulin polymers. *Proc. Natl. Acad. Sci. USA* **93**:519–523.
13. Errington, J., R. A. Daniel, and D. J. Scheffers. 2003. Cytokinesis in bacteria. *Microbiol. Mol. Biol. Rev.* **67**:52–65.
14. Evinger, M., and N. Agabian. 1977. Envelope-associated nucleoid from *Caulobacter crescentus* stalked and swarmer cells. *J. Bacteriol.* **132**:294–301.
15. Gardy, J. L., M. R. Laird, F. Chen, S. Rey, C. J. Walsh, M. Ester, and F. S. Brinkman. 2005. PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21**:617–623.
16. Gitai, Z., N. Dye, and L. Shapiro. 2004. An actin-like gene can determine cell polarity in bacteria. *Proc. Natl. Acad. Sci. USA* **101**:8643–8648.
17. Gitai, Z., N. A. Dye, A. Reisenauer, M. Wachi, and L. Shapiro. 2005. MreB actin-mediated segregation of a specific region of a bacterial chromosome. *Cell* **120**:329–341.
18. Goldfarb, D. S., J. Gariepy, G. Schoolnik, and R. D. Kornberg. 1986. Synthetic peptides as nuclear localization signals. *Nature* **322**:641–644.
19. Jacobs-Wagner, C. 2004. Regulatory proteins with a sense of direction: cell cycle signalling network in *Caulobacter*. *Mol. Microbiol.* **51**:7–13.
20. Jones, L. J., R. Carballido-Lopez, and J. Errington. 2001. Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. *Cell* **104**:913–922.
21. Keiler, K. C., P. R. Waller, and R. T. Sauer. 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**:990–993.
22. Maddock, J. R., and L. Shapiro. 1993. Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259**:1717–1723.
23. McAdams, H. H., and L. Shapiro. 2003. A bacterial cell-cycle regulatory network operating in time and space. *Science* **301**:1874–1877.
24. Meisenzahl, A. C., L. Shapiro, and U. Jenal. 1997. Isolation and characterization of a xylose-dependent promoter from *Caulobacter crescentus*. *J. Bacteriol.* **179**:592–600.
25. Moreland, R. B., H. G. Nam, L. M. Hereford, and H. M. Fried. 1985. Identification of a nuclear localization signal of a yeast ribosomal protein. *Proc. Natl. Acad. Sci. USA* **82**:6561–6565.
26. Mukherjee, A., K. Dai, and J. Lutkenhaus. 1993. *Escherichia coli* cell division protein FtsZ is a guanine nucleotide binding protein. *Proc. Natl. Acad. Sci. USA* **90**:1053–1057.
27. Ryan, K. R., and L. Shapiro. 2003. Temporal and spatial regulation in prokaryotic cell cycle progression and development. *Annu. Rev. Biochem.* **72**:367–394.
28. Sciochetti, S. A., T. Lane, N. Ohta, and A. Newton. 2002. Protein sequences and cellular factors required for polar localization of a histidine kinase in *Caulobacter crescentus*. *J. Bacteriol.* **184**:6037–6049.
29. Shaner, N. C., R. E. Campbell, P. A. Steinbach, B. N. Giepmans, A. E. Palmer, and R. Y. Tsien. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**:1567–1572.
30. Shapiro, L., H. H. McAdams, and R. Losick. 2002. Generating and exploiting polarity in bacteria. *Science* **298**:1942–1946.
31. Shih, Y. L., T. Le, and L. Rothfield. 2003. Division site selection in *Escherichia coli* involves dynamic redistribution of Min proteins within coiled structures that extend between the two cell poles. *Proc. Natl. Acad. Sci. USA* **100**:7865–7870.
32. Shih, Y. L., and L. Rothfield. 2006. The bacterial cytoskeleton. *Microbiol. Mol. Biol. Rev.* **70**:729–754.
33. Skerker, J. M., and L. Shapiro. 2000. Identification and cell cycle control of a novel pilus system in *Caulobacter crescentus*. *EMBO J.* **19**:3223–3234.
34. Taghbalout, A., and L. Rothfield. 2007. RNaseE and the other constituents of the RNA degradosome are components of the bacterial cytoskeleton. *Proc. Natl. Acad. Sci. USA* **104**:1667–1672.
35. Tobias, J. W., T. E. Shrader, G. Rocap, and A. Varshavsky. 1991. The N-end rule in bacteria. *Science* **254**:1374–1377.
36. van den Ent, F., L. A. Amos, and J. Lowe. 2001. Prokaryotic origin of the actin cytoskeleton. *Nature* **413**:39–44.
37. Varshavsky, A. 1991. Naming a targeting signal. *Cell* **64**:13–15.