Structure, Vol. 12, 1355–1360, August, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.str.2004.06.012

Optimization of Protein Production in Mammalian Cells with a Coexpressed Fluorescent Marker

Filippo Mancia,[1,4] Saurabh D. Patel,[1] Michael W. Rajala,[3] Philipp E. Scherer,[2,3] Adriana Nemes,[4] Ira Schieren,[4] Wayne A. Hendrickson,[1,4,*] and Lawrence Shapiro[1,5,6,*]

[1]Department of Biochemistry and Molecular Biophysics
Columbia University
New York, New York 10032
[2]Department of Cell Biology
Albert Einstein College of Medicine
1300 Morris Park Avenue
Bronx, New York 10461
[3]Diabetes Research and Training Center and Department of Medicine
Division of Endocrinology
Albert Einstein College of Medicine
1300 Morris Park Avenue
Bronx, New York 10461
[4]Howard Hughes Medical Institute
Columbia University
New York, New York 10032
[5]Naiomi Berrie Diabetes Center
Columbia University
New York, New York 10032
[6]Edward S. Harkness Eye Institute
Columbia University
New York, New York 10032

Summary

The expression of mammalian proteins in sufficient abundance and quality for structural studies often presents formidable challenges. Many express poorly in bacterial systems, whereas it can be time consuming and expensive to produce them from cells of higher organisms. Here we describe a procedure for the direct selection of stable mammalian cell lines that express proteins of interest in high yield. Coexpression of a marker protein, such as green fluorescent protein, is linked to that of the desired protein through an internal ribosome entry site in the vector that is transfected into cells in culture. The coexpressed marker is used to select for highly expressing clonal cell lines. Applications are described to a membrane protein, the 5HT2c serotonin receptor, and to a secreted cysteine-rich protein, resistin. Besides providing an expeditious means for producing mammalian proteins for structural work, the resulting cell lines also readily support tests of functional properties and structure-inspired hypotheses.

Introduction

High-resolution structural studies of proteins generally require large amounts of pure, properly folded material. Indeed, the advent of gene manipulation techniques for producing recombinant protein in heterologous systems is arguably the most important breakthrough of the last 30 years for structural biology, exceeding even the wonderous developments in synchrotron crystallography (Hendrickson, 2000) and NMR spectroscopy (Wuthrich, 2003). Bacterial expression systems, primarily based on the gram-negative bacterium *Escherichia coli*, have been by far the most successful for the production of recombinant proteins for structural studies. Of the 23,257 protein structures deposited in the Protein Data Bank, 14,011 have "Expression_System" records (http://www.rcsb.org). Of these, over 90% were produced using *E. coli*; ~3.5% were produced in yeast; ~2.5% with insect cells; and ~1.5% using mammalian cells. The remaining ~3% of these structures were determined using proteins expressed in other systems, including bacterial hosts such as *B. subtilis* and cell-free expression systems. The success of bacteria-based expression systems arises from several factors, including the ease with which such organisms can be genetically manipulated; the thorough understanding of their transcription and translation machinery, which has led to the ability to achieve high levels of protein expression; the rapidity of their growth; and the relatively low cost of their use. Despite these advantages, bacterial systems often fail in their application to the expression of eukaryotic proteins (Baneyx, 1999; Makrides, 1999; Geisse et al., 1996). Failure to achieve acceptable expression often arises from toxicity of the foreign protein or its inability to fold or be targeted properly in the bacterial cell. Such problems inevitably result in low levels of expression or protein misfolding (Baneyx, 1999; Makrides, 1999; Geisse et al., 1996). Thus, despite drawbacks in efficiency, alternative expression systems based on eukaryotic hosts have been developed for large-scale protein production. These include expression in yeast, insect cells, and mammalian cells, all of which have been used successfully in producing proteins for structure determination.

Expression of mammalian proteins has proven to be particularly challenging in bacterial systems. The success rate has been better in heterologous eukaryotic cell systems, such as yeast, baculovirus-infected insect cells, and stably transformed insect cells, but optimal expression of some mammalian proteins may require mammalian host cells. These proteins evolved in a mammalian cellular milieu, and it is understandable that both proper folding and stability may depend on this environment. Unique properties of the mammalian cell environment that may facilitate homologous expression include specific lipid compositions of the various membrane compartments, cell-type specific chaperones, and unique posttranslational modifications, including defined glycosylation, sulfonation, and other covalent modifications. Although it is now routine to use mammalian protein expression systems in functional studies, their application to large-scale protein production has often been deterred by difficulty in obtaining large

*Correspondence: wayne@convex.hhmi.columbia.edu (W.A.H.), shapiro@convex.hhmi.columbia.edu (L.S.)

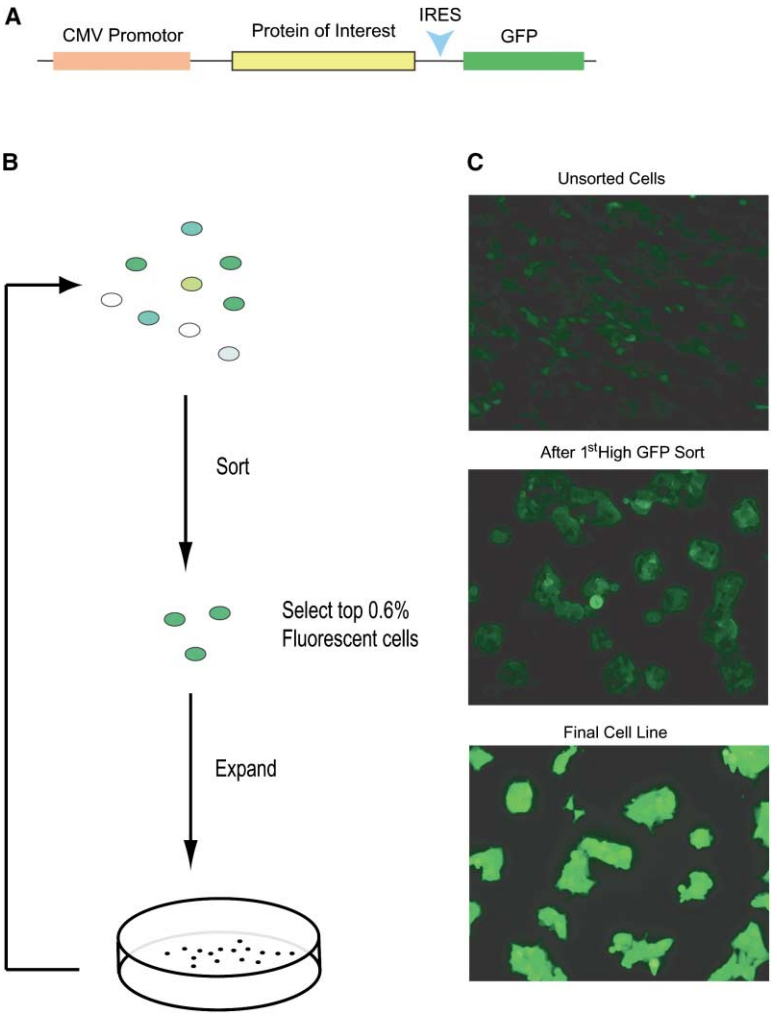

Figure 1. Schematic Representation of the GFP-Selection Mammalian Expression System

(A) The expression vector pFM-1.1. The protein of interest is placed downstream from the strong constitutive CMV promoter. Following the termination codon of the protein of interest is an internal ribosome entry site (IRES) that enables translation of GFP to be initiated from an internal site of the bicistronic mRNA transcript. This enables production of two separate proteins: GFP and the protein of interest. Similar vectors are now commercially available (for example, pIRES-GFP from Clontech, Inc.). pFM-1.2 differs from pFM-1.1 in that it contains an antibiotic resistance gene for puromycin under control of a separate promoter. The pFM vectors are based on a pBluescript parent vector, which was modified by the insertion of the CMV promoter region, followed by a multiple cloning site, and an IRES-GFP segment which included a appropriate poly-A tail.

(B) Enrichment procedure. A highly expressing cell line is developed by repeated rounds of cell sorting, selecting for the highest levels of GFP-derived fluorescence. Since both GFP and the protein of interest are expressed from the same mRNA, GFP fluorescence provides a useful surrogate correlated to levels of the protein of interest.

(C) Progression of the enrichment process monitored by fluorescence microscopy. The cells depicted here express the serotonin receptor $5HT2_C$. Three stages are shown: after antibiotic selection, after the first FACS cycle, and after the final one.

amounts of material rapidly at a reasonable cost. When this is achieved, however, advantages ensue both for the structural work and also for functional tests of structure-inspired hypotheses.

Production of recombinant protein in mammalian cells can be accomplished either through viral infection, transient transfection, or integration of expression constructs into the host genome. Each of these methods has advantages and disadvantages. For example, viral infection typically requires the propagation of recombinant virus in a "packaging" cell line, isolation of the virus, and determination of viral titer prior to infection of the host cells. In transiently transfected mammalian cells, protein expression levels peak around 48–72 hr after transfection and inevitably decline thereafter. Although the production of stable transfectants is time consuming, stably transfected cells are desirable for their ability to provide a constant source of recombinant protein. Generation of stably producing cell lines requires integration of the expression construct into the genome of the host cell. This leads to heterogeneity in expression levels, arising from differences in the number of integrants, and their sites of integration. Thus, a key step in harnessing such cells for protein production is the selection of those single cells that achieve the highest expression levels.

Here we describe a system for the selection of highly expressing stable mammalian cells, based on detection of the fluorescence intensity of a coexpressed marker, the green fluorescent protein (GFP) (Tsien, 1998; Chalfie, 1995). In this system, the coding sequence for the gene of interest is placed under the control of a strong constitutive promoter (such as the promoter element derived from cytomegalovirus [CMV]) (Thomsen et al., 1984). Downstream, after the termination codon for the gene of interest, an internal ribosome entry site (IRES) (Vagner et al., 2001) is followed by the coding sequence for GFP. Transcription from this construct produces a single bicistronic messenger RNA encoding both genes. The IRES element enables binding of the ribosome at the initiation site of GFP. Thus, two separate proteins—the gene of interest and GFP—are translated from the same message, and expression levels of both proteins are thereby coupled. This system enables efficient selection of high expressors by monitoring the fluorescence intensity of cells expressing variable amounts of GFP. Use of fluorescence-activated cell-sorting (FACS) (Ibrahim and van den Engh, 2003; Galbraith et al., 1999) technology allows

for rapid selection of either clonal or non-clonal populations of highly expressing cells.

The use of IRES-GFP elements as binary markers for successfully transfected cells is a well-established technique. However, the use of fluorescence intensity as an indicator of target-protein expression levels has seen far fewer applications (for example, see Meng et al., 2000; Liu et al., 2000). To our knowledge, only one study has employed this technique for large scale protein production for structural studies (Patel et al., 2004), and this is one of the two examples presented in this manuscript.

Methods and Results

As a test for the applicability of this system to high-level expression of functional proteins, we chose two targets, each presenting different challenges. The first target is the rat serotonin receptor subtype 2c (5HT2c) (Julius et al., 1988), a G protein-coupled receptor (GPCR). GPCRs are a large family of integral membrane proteins characterized by seven-transmembrane spanning helices. GPCRs are notoriously resistant to crystallization, in part due to the difficulty of attaining high-level expression of functional protein (Grisshammer et al., 1995; Sarramegna et al., 2003). To date only one GPCR, bovine rhodopsin, has yielded a high-resolution structure (Okada et al., 2002; Palczewski et al., 2000). Rhodopsin, unlike other GPCRs, is present at high levels in rod cell outer segments where it is naturally expressed. The crystal structure of rhodopsin was determined using material purified from natural sources, rather than with a recombinant expression system.

The second target, mouse resistin, is a highly disulfide-linked hormone that is naturally secreted from adipocytes (Steppan et al., 2001; Holcomb et al., 2000). Attempts at expression of resistin in *E. coli*, either as soluble protein or refolded from inclusion bodies, does not yield properly folded functional protein (Juan et al., 2003; Aruna et al., 2003). Resistin adopts a complex multimeric structure (Patel et al., 2004), which inevitably represents a challenge to reproduce with fidelity in heterologous expression hosts.

Expression of 5HT2c

The cDNA for the rat 5HT2c, which encodes a protein of 460 amino acids with three potential glycosylation sites and one palmitoylation site, was inserted into the multiple cloning site of the pCMV-IRES-GFP vector (pFM1.2; Figure 1A). pFM1.2 carries an antibiotic resistance gene for puromycin under the control of a separate promoter. This construct was transfected into T antigen-transformed human embryonic kidney 293 (HEK-293T) cells using lipofectamine (Invitrogen, Inc.). Stable integrants were selected by growth in puromycin-containing media for a period of approximately 3 weeks. This resulted in the growth of individual colonies, displaying varying levels of GFP-generated fluorescence. These colonies were pooled, and the resulting cell suspension was then sorted by GFP fluorescence on a Coulter Epics 753 Flow Cytometer. The top 0.1% of the most highly fluorescent cells was separated from the rest, re-plated, and allowed to propagate. This cycle of cell sorting followed by regrowth was repeated five times until a homogeneous level of fluorescence was exhibited by all cells. Figure 2A shows fluorescence profiles for these cells at different stages of the procedure. Figure 2B shows quantitative western blots corresponding to whole cell lysates at each of these stages. These data show correlation between GFP-derived fluorescence and the expression level of 5HT2c. Functionality of the expressed protein was assessed by ligand binding analysis of the recombinant protein (Figures 2C) and reveal saturable binding equivalent to that observed for the naturally produced protein. Furthermore, these data provide a means to quantitate the levels of functional protein, which reach 140–160 pmol/(mg membrane protein), corresponding to approximately 3×10^6 5HT2c molecules per cell, or 2.5 mg per 10^{10} cells (about 1–5 liters of suspension culture or 2–3 10-layer cell farms [6320cm^2]).

Expression of Resistin

For expression of mouse resistin, we inserted its coding sequence into pFM1.1 (Figure 1A), which is identical to pFM1.2, but lacks an antibiotic selection cassette. Thus, we co-transfected this expression vector with a separate plasmid encoding puromycin resistance, pRSV-puro. Stable HEK-293T cells were obtained by growth in puromycin-containing media, and a FACS-based enrichment protocol similar to that described above was employed. Since resistin is a secreted protein, cell supernatants were used to monitor protein production levels. Similar to 5HT2c, these levels appear to correlate well with the fluorescence of the cells (Figure 3). Yields on the order of 5 mg/liter were routinely obtained using serum-free media that facilitated purification. Like natural resistin produced by adipocytes and detected in mouse serum, recombinant resistin is hexameric. Physiological insulin clamp studies in mice showed the recombinant protein to function as a potent antagonist of insulin action in the liver (Rajala et al., 2003). Furthermore, this protein produced crystals that were suitable for diffraction analysis (Patel et al., 2004).

Discussion

The method described here enables the rapid generation of high-expressing stable mammalian cell lines. The entire procedure, from transfection to obtaining the final cell line, can be accomplished in less than 2 months time. While this is slow in comparison to bacterial expression methods, it is comparable to the timescale of other widely used methods, such as infection of insect cells with recombinant baculovirus, generation of yeast stable integrants, or the production of mammalian cell lines using traditional techniques.

The GFP selection method provides significant advantages in comparison to conventional methods of cell line generation. The isolation of stable integrants in mammalian cells is generally accomplished with the use of antibiotic markers. Expression levels among these antibiotic-resistant colonies are highly variable. Traditionally, to screen for highly expressing cells, individual colonies are handpicked and assayed for their levels of protein production by biochemical methods, usually involving

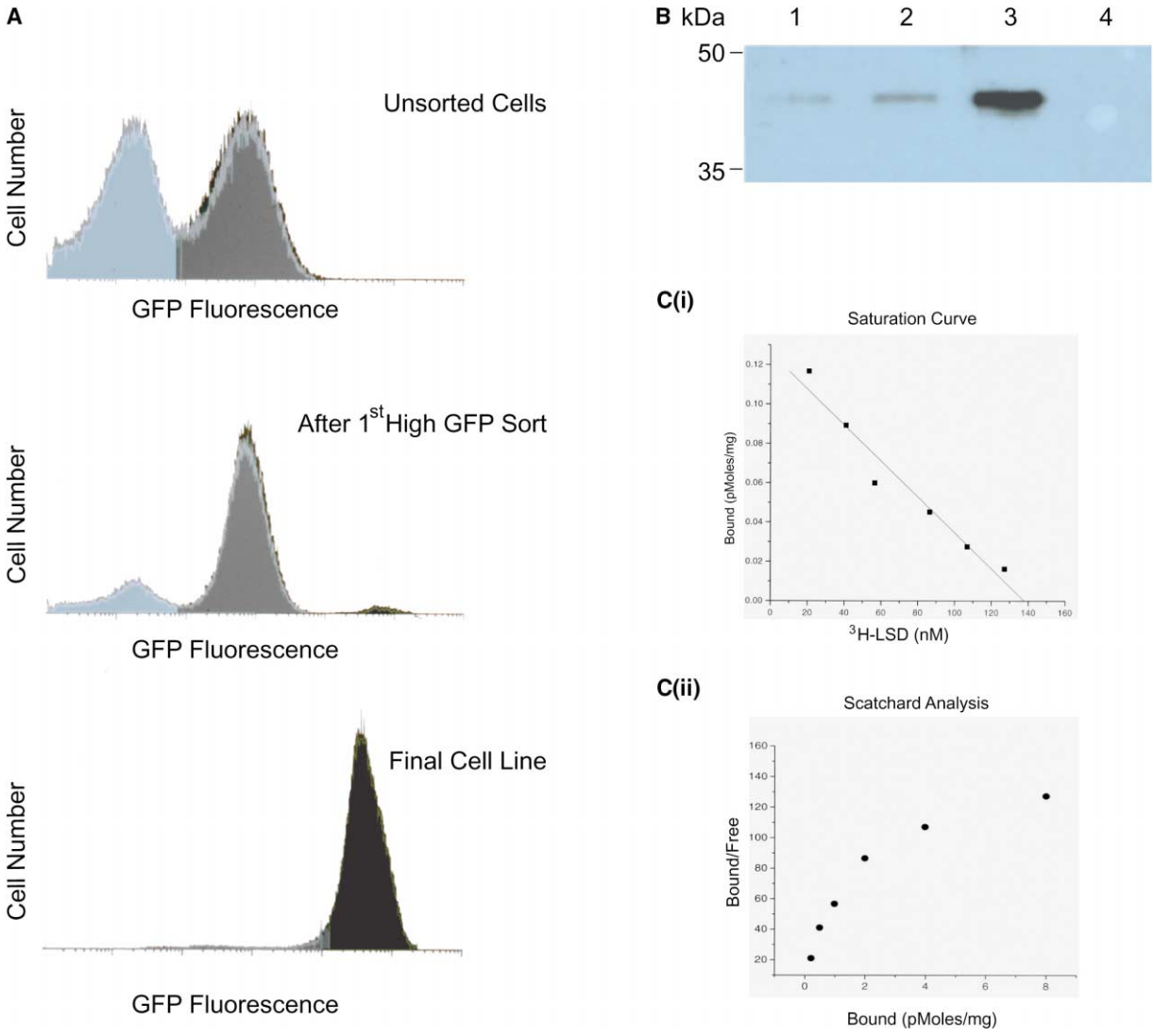

Figure 2. Expression of the Serotonin Receptor Using the GFP Selection Method

(A) Flow cytometry sorting of GFP-serotonin receptor expressing HEK-293T cells. The populations are represented progressively darker in accordance with increasing levels of GFP expression.

(B) Western blot analysis of cells at different stages of selection. Lane 1 represents cells 48 hr after transfection; lane 2 represents cells after puromycin selection; lane 3 represents cells after GFP selection; and lane 4 represents untransfected cells. Twenty thousand cells were run on each lane, and the samples were deglycosylated for 1 hr on ice with endoglycosidase F prior to loading. The membrane was probed with an anti-5HT2c rabbit polyclonal antibody generously provided by Dr. Jon Backstrom (Backstrom et al., 1995).

(C) Ligand binding to membranes isolated from HEK-293T cells enriched for the expression of serotonin-receptor. (i) Saturation curve for tritiated lysergic acid diamine (LSD). (ii) Scatchard plot of data shown in panel (i).

immunological detection. The GFP-based selection method, implemented with FACS sorting, provides an efficient means for identifying and isolating highly expressing cells. For most proteins, no direct method of detection is available. Direct linkages between a protein of interest and a fluorescent marker are not generally suitable for structural studies. The separation of the fluorescent marker from the protein of interest through the use of an IRES element obviates the need to modify the protein while maintaining the correlation between expression level and fluorescence within each cell.

Isolation of the most highly fluorescent cells, corresponding to the highest expressors, can be accomplished in a number of ways. First, visual inspection of fields of colonies enables rapid identification of the most suitable candidates, which can be manually isolated. Alternatively, colonies can be pooled and subjected to FACS analysis. Although, as shown here, the generation of clonal cell lines is not a requirement for achieving high-level expression, the current generation of cell sorters do allow for single-cell cloning. Thus, clonal cell lines can be produced with equivalent ease.

Other fluorescent markers, most notably conjugated antibodies, can also be used to select highly expressing cells (Mirzabekov et al., 1999). Although such markers provide direct correlation to protein expression levels, their use is limited to membrane-attached proteins with extracellular epitopes. In contrast, although the GFP

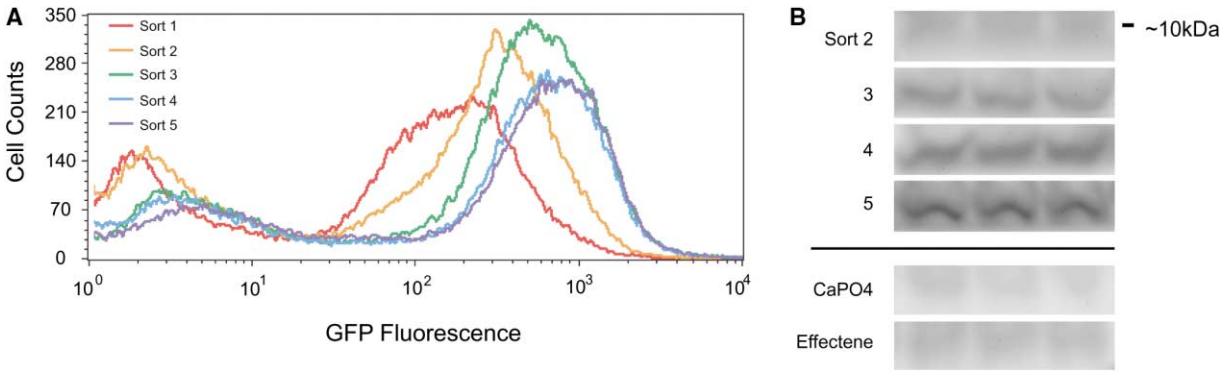

Figure 3. Expression of the Secreted Protein Resistin Using the GFP Selection Method

(A) Flow cytometry sorting of Resistin/GFP-expressing HEK-293T cells. Five sequential cell-sorting runs were performed in accordance with the scheme shown in Figure 1B. Fluorescence traces are shown for cells from each sorting run. After each sort, the top 0.6% of the most fluorescent cells were pooled and expanded. With each sort, the average fluorescence per cell increases, until reaching a plateau at sort five. (B) Resistin protein levels monitored by Coomassie blue staining of SDS gels of cell culture supernatants. Resistin levels of cell supernatants from each sort increase in concordance with GFP fluorescence. Each of the three lanes represents equivalent samples taken from three independently plated culture dishes from each sort, showing the reproducible nature of the increase in protein production. Only the section of the gel corresponding to resistin is shown. The lower two panels show similar analyses of supernatants from cells transiently transfected using either calcium phosphate or Effectene (Qiagen, Inc.). Each lane was loaded with cell supernatant concentrated by a 60% ammonium sulfate cut, which is known to precipitate resistin. The load of each lane corresponds to ~300μl of conditioned serum-free medium. Cells were grown in 75 mm dishes with 10 ml medium per dish; sorted cells were transferred to serum-free media at 80% confluence, and conditioned medium was collected after 3 days. For transiently transfected cells, transfection was performed at 80% confluence, and the media was changed to serum free medium 24 hr post-transfection. Supernatants were collected after 3 days, and treated as above for gel analysis.

selection method correlates fluorescence with expression at the mRNA level, it is not restricted to a limited class of proteins, nor does it depend on the availability of fluorescent markers that bind the protein of interest.

The GFP selection system has potential for future enhancements. These include the possibility for coexpression of multiple genes by constructing bicistronic messages for each, with a different fluorescent protein such as YFP or CFP. This can enable sorting at multiple wavelengths in order to select cells that express all of the proteins highly. Toxic proteins can often be tolerated by cells only under tightly controlled inducible expression. Inducible mammalian expression systems have recently become widely available and have proven extremely valuable for high-level expression of proteins that negatively impact cell viability (Reeves et al., 2002a, 2002b). Although the system described here provides constitutive expression, in principle it can be modified to provide inducible expression by changing the promoter element.

As structural biology progresses toward the elucidation of increasingly complex macromolecular structures, so too will the need for abundant supplies of the appropriately assembled recombinant molecules. This may lead to an increasing utilization of mammalian expression systems. The GFP selection method presented here provides an alternative method of accomplishing one critical step, that of the identification and isolation of highly expressing cells. In the future, we can expect additional substantial developments that will place mammalian expression among the standard techniques of structural biology.

Acknowledgments

We are grateful to Richard Axel for helpful discussions and support throughout the course of this work, Hubi Amrein and Thomas Livelli for carrying out the majority of the work involved in the construction of the pFM vectors, and for their generosity in allowing us their use. We thank Peter Chen for help with bioinformatics analysis of PDB structures and Paul Lee for helpful discussions. F.M. was supported by fellowships from the European Molecular Biology Organization and the Human Frontiers Science Program. This work was supported by grants from NIH (GM68671, W.A.H. and F.M.; GM68671 and GM62529, L.S.; R01-DK55758, P.E.S.; Medical Scientist Training Grant T32-GM97288, M.W.R.), and grants from the American Diabetes Association and Foundation for Research to Prevent Blindness (L.S.). Part of this work was performed within the New York Structural Genomics Research Consortium (NYSGXRC).

Received: April 9, 2004
Revised: June 4, 2004
Accepted: June 4, 2004
Published: August 10, 2004

References

Aruna, B., Ghosh, S., Singh, A.K., Mande, S.C., Srinivas, V., Chauhan, R., and Ehtesham, N.Z. (2003). Human recombinant resistin protein displays a tendency to aggregate by forming intermolecular disulfide linkages. Biochemistry *42*, 10554–10559.

Backstrom, J.R., Westphal, R.S., Canton, H., and Sanders-Bush, E. (1995). Identification of rat serotonin 5–HT2C receptors as glycoproteins containing N-linked oligosaccharides. Brain Res. Mol. Brain Res. *33*, 311–318.

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. Curr. Opin. Biotechnol. *10*, 411–421.

Chalfie, M. (1995). Green fluorescent protein. Photochem. Photobiol. *62*, 651–656.

Galbraith, D.W., Anderson, M.T., and Herzenberg, L.A. (1999). Flow cytometric analysis and FACS sorting of cells based on GFP accumulation. Methods Cell Biol. *58*, 315–341.

Geisse, S., Gram, H., Kleuser, B., and Kocher, H.P. (1996). Eukaryotic expression systems: a comparison. Protein Expr. Purif. *8*, 271–282.

Grisshammer, R., and Tate, C.G. (1995). Overexpression of integral membrane proteins for structural studies. Q. Rev. Biophys. *28*, 315–422.

Hendrickson, W.A. (2000). Synchrotron crystallography. Trends Biochem. Sci. *25*, 637–643.

Holcomb, I.N., Kabakoff, R.C., Chan, B., Baker, T.W., Gurney, A., Henzel, W., Nelson, C., Lowman, H.B., Wright, B.D., Skelton, N.J., et al. (2000). FIZZ1, a novel cysteine-rich secreted protein associated with pulmonary inflammation, defines a new gene family. EMBO J. *19*, 4046–4055.

Ibrahim, S.F., and van den Engh, G. (2003). High-speed cell sorting: fundamentals and recent advances. Curr. Opin. Biotechnol. *14*, 5–12.

Juan, C.C., Kan, L.S., Huang, C.C., Chen, S.S., Ho, L.T., and Au, L.C. (2003). Production and characterization of bioactive recombinant resistin in Escherichia coli. J. Biotechnol. *103*, 113–117.

Julius, D., MacDermott, A.B., Axel, R., and Jessell, T.M. (1988). Molecular characterization of a functional cDNA encoding the serotonin 1c receptor. Science *241*, 558–564.

Liu, X., Constantinescu, S.N., Sun, Y., Bogan, J.S., Hirsch, D., Weinberg, R.A., and Lodish, H.F. (2000). Generation of mammalian cells stably expressing multiple genes at predetermined levels. Anal. Biochem. *280*, 20–28.

Makrides, S.C. (1996). Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol. Rev. *60*, 512–538.

Meng, Y.G., Liang, J., Wong, W.L., and Chisholm, V. (2000). Green fluorescent protein as a second selectable marker for selection of high producing clones from transfected CHO cells. Gene *242*, 201–207.

Mirzabekov, T., Bannert, N., Farzan, M., Hofmann, W., Kolchinsky, P., Wu, L., Wyatt, R., and Sodroski, J. (1999). Enhanced expression, native purification, and characterization of CCR5, a principal HIV-1 coreceptor. J. Biol. Chem. *274*, 28745–28750.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., et al. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. Science *289*, 739–745.

Patel, S.D., Rajala, M.W., Rossetti, L., Scherer, P.E., and Shapiro, L. (2004). Disulfide-dependent multimeric assembly of resistin family hormones. Science *304*, 1154–1158.

Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E.M., and Shichida, Y. (2002). Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. Proc. Natl. Acad. Sci. USA *99*, 5982–5987.

Rajala, M.W., Obici, S., Scherer, P.E., and Rossetti, L. (2003). Adipose-derived resistin and gut-derived resistin-like molecule-beta selectively impair insulin action on glucose production. J. Clin. Invest. *111*, 225–230.

Reeves, P.J., Callewaert, N., Contreras, R., and Khorana, H.G. (2002a). Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. Proc. Natl. Acad. Sci. USA *99*, 13419–13424.

Reeves, P.J., Kim, J.M., and Khorana, H.G. (2002b). Structure and function in rhodopsin: a tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. Proc. Natl. Acad. Sci. USA *99*, 13413–13418.

Sarramegna, V., Talmont, F., Demange, P., and Milon, A. (2003). Heterologous expression of G-rotein-coupled receptors: comparison of expression systems fron the standpoint of large-scale production and purification. Cell. Mol. Life Sci. *60*, 1529–1546.

Steppan, C.M., Bailey, S.T., Bhat, S., Brown, E.J., Banerjee, R.R., Wright, C.M., Patel, H.R., Ahima, R.S., and Lazar, M.A. (2001). The hormone resistin links obesity to diabetes. Nature *409*, 307–312.

Thomsen, D.R., Stenberg, R.M., Goins, W.F., and Stinski, M.F. (1984). Promoter-regulatory region of the major immediate early gene of human cytomegalovirus. Proc. Natl. Acad. Sci. USA *81*, 659–663.

Tsien, R.Y. (1998). The green fluorescent protein. Annu. Rev. Biochem. *67*, 509–544.

Vagner, S., Galy, B., and Pyronnet, S. (2001). Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites. EMBO Rep. *2*, 893–898.

Wuthrich, K. (2003). NMR Studies of Structure and Function of Biological Macromolecules (Nobel Lecture). Angew. Chem. Int. Ed. Engl. *42*, 3340–3363.