ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/50834518

A Practical Teaching Course in Directed Protein Evolution Using the Green Fluorescent Protein as a Model

ARTICLE *in* BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION · JANUARY 2011

Impact Factor: 0.65 · DOI: 10.1002/bmb.20430 · Source: PubMed

CITATIONS
2

READS
22

5 AUTHORS, INCLUDING:

Roberto Ruller
Laboratório Nacional de Ciência e Tecnologi…
47 PUBLICATIONS **619** CITATIONS
SEE PROFILE

Rafael Silva-Rocha
University of São Paulo
45 PUBLICATIONS **388** CITATIONS
SEE PROFILE

Artur Silva
Federal University of Pará
181 PUBLICATIONS **1,409** CITATIONS
SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Rafael Silva-Rocha
Retrieved on: 24 January 2016

© 2011 by The International Union of Biochemistry and Molecular Biology

BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION
Vol. 39, No. 1, pp. 21–27, 2011

Special Section: Innovative Laboratory Exercises

A Practical Teaching Course in Directed Protein Evolution Using the Green Fluorescent Protein as a Model*,†

Received for publication, March 8, 2010, and in revised form, May 6, 2010

Roberto Ruller‡, Rafael Silva-Rocha§, Artur Silva§, Maria Paula Cruz Schneider§, and Richard John Ward‡¶

From the ‡Departamento de Química, Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade São Paulo, São Paulo, Brazil, §Departamento de Genética, Laboratório de Polimorfismo de DNA, Universidade Federal do Pará, Campus Universitário do Guamá, Rua Augusto Correa 01, 66075-970 Belém-PA, Brazil

Protein engineering is a powerful tool, which correlates protein structure with specific functions, both in applied biotechnology and in basic research. Here, we present a practical teaching course for engineering the green fluorescent protein (GFP) from *Aequorea victoria* by a random mutagenesis strategy using error-prone polymerase chain reaction. Screening of bacterial colonies transformed with random mutant libraries identified GFP variants with increased fluorescence yields. Mapping the three-dimensional structure of these mutants demonstrated how alterations in structural features such as the environment around the fluorophore and properties of the protein surface can influence functional properties such as the intensity of fluorescence and protein solubility.

Keywords: Directed evolution, error-prone PCR, green fluorescent protein, random mutagenesis, recombinant protein expression.

This report presents a description of the first experimental teaching course on the directed evolution (DE)[1] of proteins destined for Brazilian postgraduate students, which focuses on a general perspective of recent methodologies in protein engineering for the study of protein structure and function. The practical course is offered each year to postgraduate students (masters and doctorate) in biological and pharmaceutical sciences at the Federal University of Pará in the Brazilian state of Pará. A prerequisite for selection of the students is a general, graduate-level knowledge of biochemistry, cellular biology, molecular biology, and microbiology. The course intends to offer a vision of the current status of protein engineering with a particular focus on the biochemical techniques associated with DE strategies for students with a general interest in the area of structural biology or related fields.

Advances in genetic engineering technologies permit ever more sophisticated control of the functional properties of recombinant proteins [1–3] and have proven their potential in the development of novel biological applications for these macromolecules. Recent progress in recombinant DNA technology, together with the techniques of high-throughput selection, has allowed the development of new biocatalysts. In particular, the strategy of DE has emerged as a powerful tool for engineering proteins of biotechnological interest with applications in areas such as medicine, the pharmaceutical and food industries, and in the production of leather, paper, and other products [3–5].

BACKGROUND THEORY AND PRELABORATORY PREPARATION

Rational design strategies are based on the analysis of amino acid sequences and 3D structures to identify the determinants of a given protein function that may then be altered by site-directed mutagenesis. By contrast, DE relies on random changes in DNA sequences followed by the selection of desired phenotypes from a library of random mutants. When necessary, repeated cycles of random mutagenesis and selection may be carried out. The DE of proteins is thus an approach that mimics the process of natural selection in the laboratory [6–8]. However, the Darwinian evolution of a molecule occurs over a timescale of millions of years, whereas directed *in vitro* evolution dramatically accelerates the appearance of novel phenotypes.

The DE strategy has been successfully applied to the manipulation of a wide range of protein functions, including thermal stability and catalytic efficiency [1–3]. In this practical course, the green fluorescent protein (GFP) is used as the model to demonstrate DE using error-prone

*Additional Supporting Information may be found in the online version of this article.

†This work is supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) PROCAD program (grant number 2580/2008).

‡ To whom correspondence should be addressed. Department of Chemistry, FFCLRP, São Paulo University, Ribeirão Preto, São Paulo, Brazil. Tel.: +55 16 3602 4384; Fax: +55 16 3602 4838. E-mail: rjward@ffclrp.usp.br.

[1]The abbreviations used are: DE, directed evolution; epPCR, error-prone polymerase chain reaction; pGFPuv, plasmid carrying a GFP variant excited by ultra-violet light.

 DOI 10.1002/bmb.20430

polymerase chain reactions (epPCRs). The GFP is produced by the jellyfish *Aequorea victoria* and is intrinsically fluorescent, emitting a brilliant green light when exposed to ultraviolet or blue light [9]. The gene that encodes this protein has been isolated and is useful for the construction of fusion proteins in many organisms, including bacteria, insects, fish, plants, and mammals, as well as the monitoring of cells, protein–protein interactions, or reporting on specific gene expression [9, 10]. The GFP possesses several favorable properties for its use as a gene reporter including cytoplasmatic localization, low toxicity, and simple visualization and quantification [11, 12] that can be easily monitored by the observation of fluorescent emission using a compatible combination of low-cost optical filters.

Improvement in the fluorescent properties of the GFP molecule has broadened the use of this protein as a reporter gene [10–12]. These modifications have resulted from several mutagenesis approaches that have sought to improve the fluorescent properties of the protein. The GFP contains 283 amino acids and presents an 11-stranded β-barrel architecture in the center of which is located a short α-helix that carries the 4-(*p*-hydroxybenzylidene)-5-imidazolinone (*p*-HBI) fluorophore, which is formed by the spontaneous cyclization of the amino acids Ser65, Tyr66, and Gly67. Other amino acid residues (Gln69, Gln94, Arg96, His148, Ile167, Thr203, Ser205, and Glu222) are located in the immediate environment of the p-HBI and contribute to the modulation of the spectral properties of the fluorescent emission [9]. Crameri *et al*. [13] used random PCR mutagenesis and DNA shuffling of the GFP to create and recombine mutants, and after three cycles of DNA shuffling, the triple mutant (F99S, M153T, V163A) was obtained. This mutant presented improved folding at 37°C and higher solubility at high protein concentrations [13]. In addition, this triple mutant presented a shifted excitation spectrum with an optimum in the near ultraviolet region and was thus named GFPuv. The GFPuv has an optimum excitation wavelength of 394 nm, with maximum emission at 509 nm and can be readily visualized under the ultraviolet light (360–450 nm) provided by a laboratory transilluminator.

GENERAL NOTE

The objective of this study is to describe a laboratory exercise that demonstrates the improvement of the fluorescence intensity of the GFPuv using random mutagenesis PCR (error prone) techniques, followed by the identification and analysis of mutants exhibiting increased emission efficiency when excited by ultraviolet light. The same random mutagenesis procedure may be applied to unmodified GFP, using a standard light box covered with an appropriate filter as a light source. In addition, with given time limitations, we have found it convenient to have a pre-prepared random mutant library available in the case of failure in the epPCR or cloning of the amplified fragments.

MATERIALS AND METHODS

Experimental Procedures

Equipment—The following equipments were required for the practical classes: a thermocycler, a minigel DNA electrophoresis system, a microtube centrifuge, a manual ultraviolet light (or laboratory transilluminator), two optical filters (520 and 580 nm), a system for the capture of digital images (a common 3 megapixel resolution digital camera is adequate), a bacterial incubator (37–45°C), two water baths (42 and 100°C), an orbital shaker set at 37°C for bacterial culture, micropipettes, a sterile area for microbiological manipulation, and an autoclave to sterilize the culture broth and materials. Ideally, nucleotide sequencing capability would be necessary for the exact reproduction of the course as described in this report, but in the absence of such a capacity, electropherograms of results are included here, which is given in the Supporting Information, and can be used for mutant analysis and mapping of the three-dimensional structure of the GFPuv.

DNA Manipulation—Error-prone PCR. One hundred femtomoles of linear plasmid carrying a GFP variant excited by ultraviolet light (pGFPuv) vector (carrying the specific open-reading frame of enhanced GFP developed by Crameri *et al*. [13]; Clontech Laboratories, CA) is used as the DNA template for the epPCR reaction [3, 14]. The reaction mixture has a total volume of 100 μL containing 10 pmol of sense (5′-TCGACTCTAGAGGATCC CCGGGT-3′) and antisense (5′-GCGCTCAGTTGGAATTCATTA-3′) primers that are complementary to the 5′ and 3′ multiple cloning site sequences in the pGFPuv vector, and which include restriction sites for the endonucleases *Xba*I and *Eco*RI (underlined in the sense and antisense primers, respectively). The mixture includes 10% v/v of 10× concentrated buffer (70 mM of $MgCl_2$, 500 mM KCl, 100 mM Tris HCl pH 8.3, and either 10% glycerol or bovine serum albumin). A 10× deoxynucleotide stock (2 mM dGTP, 2 mM dATP, 10 mM dTTP, and 10 mM of dCTP in sterile water) is added to 10% v/v. The mixture is completed by adding $MnCl_2$ to a final concentration of 55 mM from a 10× concentrated (550 mM) stock solution and 10 U of Taq DNA polymerase. The Taq DNA polymerase lacks a 3′ to 5′ exonuclease proofreading activity, and therefore care should be taken not to use commercially available DNA polymerases in which proofreading activity is present. The PCR reaction is performed using 40 cycles at 94°C for 1 minute, 45°C for 1 minute, and 72°C for 1 minute with a final extension of 3 minutes at 72°C.

Preparation of the random mutant libraries. The product of the epPCR reaction is analyzed by 2% agarose gel electrophoresis, and the amplified 720-bp fragment corresponding to mutagenized GFPuv is identified, excised, and purified using the Qiagen DNA purification kit (or another appropriate method). The purified fragment is quantified using standard spectrophotometry methods, and 1 μg of DNA is digested with 10 U of *Xba*I and *Eco*RI for 2 hours at 37°C, followed by thermal inactivation at 80°C for 15 minutes. The fragment is precipitated by adding 10% v/v 3 M sodium acetate (pH 4.8) and 300% v/v of ethanol at −20°C, incubated for 10 minutes in a dry ice/ethanol mixture, and centrifuged at 12,000 × *g* for 30 minutes. The supernatant is discarded and the pellet is washed with 70% of ethanol, dried, and resuspended in 20 μL of TE (10 mM of TrisHCl pH 8.0, 5 mM of EDTA).

Plasmid Preparation. One microgram of the pGFPuv vector (Clontech Laboratories, CA) is digested with 10 U of the enzymes *Xba*I and *Eco*RI for 2 hours at 37°C, and the linearized form of the vector is separated by 0.8% agarose gel electrophoresis and purified as described in the previous section. The DNA vector is quantified by standard spectrophotometric methods and dephosphorylated with alkaline phosphatase (calf intestinal alkaline phosphatase; Fermentas Inc. Glen Burnie, MA).

DNA ligation and transformation of the library in Escherichia coli. Three hundred nanogram of the dephosphorylated vector is mixed at a molar ratio of ~1:5 vector/fragment and ligated using 5 U of T4 DNA ligase in appropriate buffer and is then incubated at 22°C for 2 hours. *Escherichia coli* DH5α is transformed with the ligation mixture using competent cells prepared by washing in Ca^{2+} solutions as described previously [15]. Around 10^4–10^5 transformants should be obtained per microgram of circular DNA. After transformation, the cells are plated in solid Luria-Bertani medium (1% tryptone, 0.5% yeast extract and 1% NaCl, 1.5% bacteriological agar at pH 7.0) with ampicillin (final concentration 100 μg/mL; Sigma-Aldrich, CA) in 15-cm-diameter Petri dishes. All dishes are further inoculated with two 25 μL drops containing 2×10^3 CFU per milliliter of *E. coli* DH5α transformed with the unmodified pGFPuv. These drops are left undisturbed and result in the growth of colonies expressing nonmutated GFPuv in a limited region of the Petri dish, which serves as an internal control during the selection procedure. The inoculated Petri dishes are incubated for 16 hours at 37°C.

Initial selection (screening on solid media). After growth, the colonies are counted and all the dishes are illuminated with ultraviolet light for the identification of the enhanced fluorescent phenotype. Two digital images of each Petri dish are recorded through 520-nm and 580-nm filters and analyzed using Photoshop (Adobe Systems Incorporated, Seattle, WA) or similar image-processing software. The image pairs are processed by subtracting the "green" channel of the 580-nm filtered image from the inverted "green" channel of the 520-nm filtered image. The threshold of the resulting subtracted image is adjusted manually until the intensity of the internal control colonies (expressing nonmutated GFPuv) disappears. The colonies still visible in this final image present more intense fluorescence in comparison with the control GFPuv and can be identified and selected for the second stage of the screening procedure.

Secondary selection (screening in liquid media). The colonies selected from the first screening stage are individually inoculated into 1 mL of selective liquid Luria-Bertani (LB) medium (1% tryptone, 0.5% yeast extract, 1% NaCl, at pH 7.0, and ampicillin 100 μg/mL) in 96 "deep-well" format multiwell plates. The cultures are grown under continuous aerobic agitation (200 rpm) for 16 hours at 37°C, after which the cell densities in the individual cultures are estimated by measuring the OD_{600} with a spectrophotometer (Hitachi, Tokyo, Japan), and normalized to a value of 0.6 using LB medium. The diluted cells are transferred to 96-well plates and photographed under ultraviolet light through 520-nm and 580-nm filters. The image pairs are processed as described above, and the cultures presenting improved fluorescence are selected for final analysis.

DNA sequencing and protein structure analysis. The selected cultures are centrifuged at 12,000 × *g* for 5 minutes, and the plasmid DNA is extracted from the bacterial pellets by the alkaline lysis method [15], quantified using spectrophotometry, and analyzed by 0.8% agarose gel electrophoresis. Nucleotide sequences of the extracted DNA are obtained with the DYEnamic ET Dye Terminator Kit using sense (5′-ATGACCATGATTA-CGCCAAGCTT-3′) and antisense (5′-CACCAGACAAGTTG GTAA TGG-3′) primers that are complementary to regions located 50 bp upstream and 30 bp downstream of the GFPuv ORF in the pGFPuv vector. Electropherograms were recorded using a MegaBace 1000 DNA Analysis System (GE Life Sciences, Fairfield, CT). The nucleotide sequences are analyzed using the free software BioEdit version v7.0.5 (www.mbio.ncsu.edu/BioEdit/bioedit.html).

Mutant mapping and protein 3D structure analysis. The sequences of the GFP mutants are used to create structural models of these proteins. The models are built automatically with the Swiss-Pdb Viewer program [16] using the coordinate structure of wild-type GFP as the template (PDB code 1GFL).

RESULTS AND DISCUSSION

The course has been offered to classes of between 9 and 15 students, typically divided into working groups each composed of three students. In our practical examples, we present a work schedule and results from a class of nine students divided into three groups. All three groups (GA, GB, and GC) prepared a random mutant library, analyzed the epPCR reactions (Fig. 1), and confirmed the amplification of the GFPuv coding sequence with a minimum size of approximately 720 bps. All three groups observed a high-molecular-weight DNA smear that is typical of epPCR, which is due to the increased nonspecific amplification resulting from the excess of manganese and magnesium chlorides in the reaction mixture. These gels provide a useful diagnostic tool for estimating the yield of the epPCR reaction.

The 720 bp bands were excised from the gel, purified, digested, and ligated into the pGFPuv. About 1 μL, 5 μL, and 10 μL of the ligation reaction were used to transform competent *E. coli* DH5α. Although the DH5α strain was used as the host cell, tests with other *E. coli* lineages (DH10B and XL1-blue) produced similar results. The transformation frequency should be around 10^4 or 10^5 cells per microgram of DNA to guarantee the use of small volumes of transformed cells during screening of the random mutant library. Each group inoculated 10 Petri dishes with 700–900 cells to produce a total of 7,000–8,000 colonies, which was sufficient to screen a random library generated by epPCR using a DNA template of approximately 720 bp. An analysis of the mutation frequency during preparation of each random mutant library is important for the estimation of the efficiency of the epPCR. Ten clones were thus selected randomly from each plate, and the GFPuv coding region was sequenced. Analysis of the nucleotide sequences revealed that around 10% of the mutations resulted in loss of the fluorescent phenotype in the transformed cells.

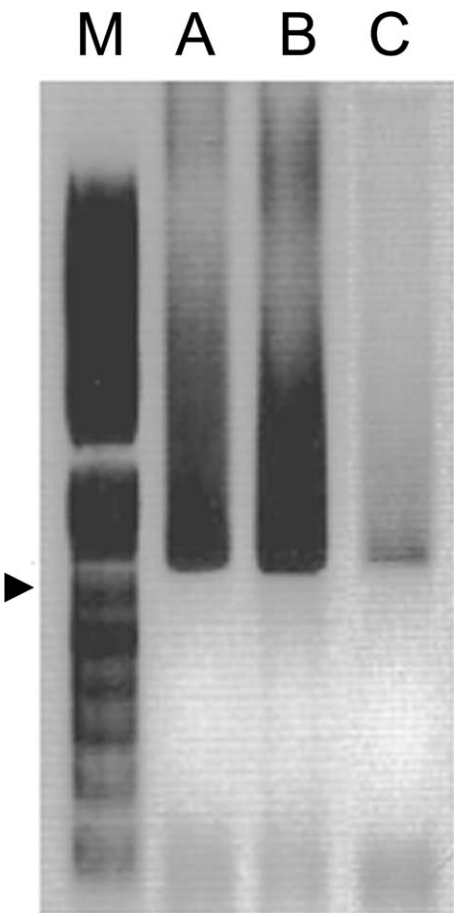

FIG. 1. **Gel electrophoresis of the epPCR reactions.** The 2% agarose gel of the epPCR reactions achieved by Groups A, B, and C (lanes GA, GB, and GC, respectively). The arrow indicates the 700-bp band in the DNA size marker (*lane M*).

The Petri dishes containing the transformed colonies were then submitted to the first round of selection (Fig. 2). For each Petri dish, two color images were taken using a standard 3 megapixel digital camera with 520-nm (Fig. 2*A*) and 580-nm (Fig. 2*B*) filters. Figure 2*C* presents the image of the Petri dish after processing and subtraction of the image pairs, and Fig. 2*D* shows the final contrast-adjusted image, which permits the rapid identification and isolation of colonies with enhanced fluorescence.

The colonies identified in this first screening step were inoculated in liquid media for growth and expression and the second phase of screening and selection. The bacterial cultures were incubated for 12 hours at a temperature of 37°C (note that aerobic conditions are essential for high levels of GFP expression). The cultures were grown in 96-well plates (deep-well format, 1 mL cultures per well), and each student group screened up to 94 mutants (the remaining two wells in each plate were inoculated with the cells transformed with the nonmutated GFPuv as internal controls used for image analysis). Figures 2*E*–2*H* present typical results of the second selection process, in which the 96-well plates were illuminated with ultraviolet light, photographed using the same filters, and processed using the same image combinations as in the initial selection. Twenty clones from Group A, 27 clones from Group B, and 25 clones from Group C that returned the brightest emissions in comparison with the internal controls after image processing were selected for the final analysis.

Plasmid DNA was directly extracted from the deep-well plate cultures and was quantified, and then the nucleotide sequence was determined by automated sequencing. The groups were then instructed on how to perform a comparative analysis of the nucleotide sequences in relation to the nonmutated GFPuv sequence using the software package BioEdit version v7.0.5. The nucleotide sequences of the cloned PCR inserts were translated *in silico*, and the positions of the

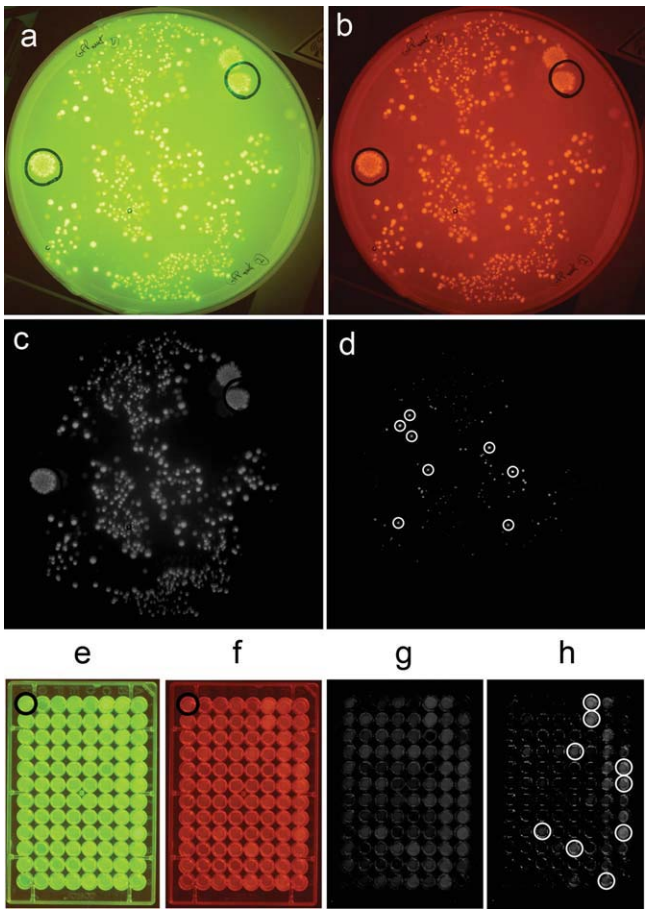

FIG. 2. ***Panels* (*a*)–(*d*)**: **Initial selection of Petri dishes**. Images of transformed colonies growing on a single Petri dish recorded with 520-nm (*a*) and 580-nm (*b*) filters. The colonies within the circles are internal controls expressing the nonmutated GFPuv. Image (*c*) shows the subtraction of images (*a*) and (*b*), and *panel* (*d*) shows the final image in which the threshold has been adjusted to eliminate the signal from the nonmutated GFPuv. The colonies outlined by the white circles passed onto the second selection. *Panels* (*e*)–(*h*): Second selection in 96 deep-well plates. After normalization of the cell density of cultured colonies from the initial selection, images were recorded with 520-nm (*e*) and 580-nm (*f*) filters. The black circles show the control well in which nonmutated GFPuv was expressed. Image (*g*) shows the result of the subtraction of images (*e*) and (*f*), and *panel* (*h*) shows the final image after threshold adjustment above the signal of the nonmutated GFPuv. The wells outlined by the white circles were selected for nucleotide sequence analysis.

TABLE I
List of GFPuv mutants

Clones	Mutation				
GA1	D21E	Y200F			
GA2	I128V				
GA3	G67S	Q184L			
GA4	K52I	Q157K			
GA5	H25R				
GA6	N105T	K126I	Y237F		
GB1	E34V				
GB2	S65T				
GB3	F223Y	I229F			
GB4	D76E				
GB5	E90D				
GB6	N149I				
GB7	I167T				
GB8	N212H				
GB9	S65T	H139L			
GC1	L125F				
GC2	Y39H	T62S			
GC3	N149K				
GC4	K107E	K166I			
GC5	H81Q	I188M	A206G		
GC6	D21V	K156Q			
GC7	V22I	K52I	I188M	A206G	H217P

The changes in amino acids generated by the epPCR experiments on the green fluorescence protein are shown. The mutants GA, GB, and GC were observed by Groups A, B, and C, respectively (see text for details).

amino acid substitutions were mapped onto the 3D structure of the GFPuv. Of the 20 clones sequenced by Group A, 18 presented mutations in the nucleotide sequence and 11 changes in the amino acid sequence (the difference being due to silent mutations). The frequency of gene mutation generated by Group A was 0.126% (18 mutations in a total of $20 \times 717 = 14{,}340$ nucleotides). The observed mutation frequencies in the selected clones from Groups B and C were 0.108% (21 mutations in a total of 19,359 nucleotides in 27 sequences) and 0.112% (20 mutations in 17,925 nucleotides in 25 sequences), respectively. The pattern of mutant incorporation varied significantly among groups (Table I). Group B recorded the highest number of single mutants (seven) and two double mutants, whereas Group C observed two single, three double, one triple, and one quintuple mutant.

The students were encouraged to compare their results, from which it was immediately clear that Group B had obtained a higher frequency of a single mutant than Group A, which in turn compares with Group C where double, triple, and even quintuple mutants were obtained. This demonstrates the sensitivity of the epPCR technique, where even subtle variations in the conditions of the epPCR reaction between the three groups resulted in significant differences in mutation frequencies. A more detailed analysis of the amino acid substitutions reveals that some mutants are repeated. For example, Group C recorded the double I188N and A206G substitutions twice, which may be the result of an accumulation of mutants during the later rounds of epPCR. Perhaps, even more interesting is the observation of multiple substitutions at a particular position, for example, D21 may be substituted by either glutamic acid (as observed in mutant GC6) or valine (in mutant GA1) and N149 may be substituted by either lysine (mutant GC3) or isoleucine (mutant GB6). Finally, the K52I mutant was observed in two different mutants (GA4 and GC7) derived from independent experiments. The appearance of identical mutants and multiple substitutions at the same position in independent experiments demonstrates the power of the DE technique. However, the majority of the mutants were unique, indicating that there are multiple paths to improved fluorescence yield in the GFPuv.

The three panels in Fig. 3 show the localization of all the mutants identified by each group, and these results prompted a discussion with regard to the structural basis of the increase in the fluorescence signal in the selected GFPuv mutants. The amino acids S65 and G67 are key residues in fluorophore formation in the GFPuv, and thus it might be expected that substitutions at these position would influence the environment around the fluorophore and thereby influence the fluorescence yield. The G67S substitution was identified in the GA3 mutant from Group A, again confirming the observation that independent

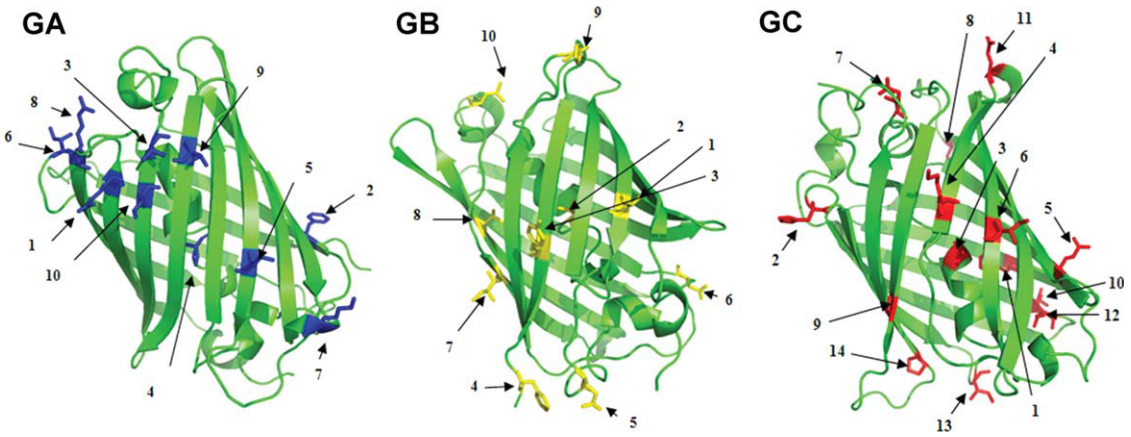

FIG. 3. **Mapping the mutants onto the 3D structure of GFPuv**. Structural models showing the locations of the mutants observed by Groups A (GA), B (GB), and C (GC). The numbers represent the amino acid substitutions (Table I) according to single-letter nomenclature of IUPAC GA: (1) E21, (2) F200, (3) V128, (4) S67, (5) L184, (6) I52, (7) K157, (8) R25, (9) T105, (10) I126, (11) F237 not shown; GB: (1) V34, (2) T65, (3) Y223, (4) F229, (5) E76, (6) D90, (7) I149, (8) T167, (9) H212, (10) L139; GC: (1) F125, (2) H39, (3) S62, (4) K149, (5) E107, (6) I166, (7) Q81, (8) M188, (9) G206, (10) D21, (11) Q156, (12) I22, (13) I52, (14) P217. The structural models were made using the program SwissPDB viewer.

experiments using DE may result in consistent and reproducible amino acid substitutions of the structural determinants of specific phenotypes. Mapping of the other substitutions revealed that in addition to G67S, the S65T mutant is located immediately before the *p*-HBI fluorophore. In addition to these residues that lie immediately adjacent to the GFPuv fluorophore, other substitutions were identified (T62S and I167T in mutants GC2 and GB7) with amino acid side chains that make contacts with the fluorophore atoms. These results demonstrate that the DE technique is a powerful method for identifying mutants in the immediate vicinity of the fluorophore.

From a total of 36 substitutions identified in the practical class, only six are buried in the interior of the protein, and as explained above, four of the six substitutions are located in the immediate vicinity of the fluorophore. The two remaining residues, at positions 22 and 125, are both located in a cluster of hydrophobic residues that include L18, V22, F27, V55, L60, Y106, L125, and F130. Both of the observed substitutions (V22I in mutant de GC7 and L125F in mutant GC1) increase the volume of the amino acid side chain, indicating that side-chain packing in the hydrophobic residues of this cluster is improved. Improved hydrophobic packing results in improved stability of the protein and suggests that protein unfolding and/or protein stability are important for the manifestation of fluorescence in the heterologous GFPuv. Biochemical studies have demonstrated that the formation of the GFPuv fluorophore is a spontaneous process that involves autocatalytic conversion of the Y66 and G67 residues to form the *p*-HBI fluorophore. Thus, protein fluorescence is only observed after correct folding of the protein, and it has been previously demonstrated that on expression of heterologous GFPuv in bacteria, only a fraction of the total expressed protein adopts the native conformation, while the rest forms inclusion bodies composed of non-native protein aggregates which are not fluorescent. This explanation is consistent with the results observed in this practical class, and it is to be expected that mutations that influence protein stability or protein folding will be detected by the fluorescence screening method used here.

The high proportion of surface residues that appeared in all three experiments is a striking result and has provoked discussion among the students. Although the present data do not permit definitive conclusions, the discussion served to place the results of the practical class in the context of research into the structure–function relationships of proteins and stimulated ideas on pathways for further investigation. The first proposal was that the high proportion of surface substitutions is a trivial consequence of the difficulty of improving hydrophobic core packing, and it is far more likely to identify substitutions of surface residues where steric limitations are far more relaxed. It was suggested that the results could be compared with those of DE studies in other proteins, and it might be expected that the accumulation of amino acid substitutions of surface residues is a common feature in proteins engineered through DE techniques. Another proposal suggested that the high frequency of surface residue substitution is related to improved solubility of the recombinant GFPuv. In this case, it might be expected that this type of substitution would increase the polarity of the protein surface, which prompted the students to analyze the physicochemical properties of the amino acid substitutions observed in all three experiments. Somewhat surprisingly, of the 30 substitutions of surface located residues, 16 were considered to be conservative, 10 resulted in the exchange of polar or charged residues for apolar residues, three substitutions of polar residues by charged residues, and only in a single case (Y39H in mutant GC2), an aromatic residue was substituted by a polar side chain. Therefore, this analysis indicates that there is a slight overall tendency to decrease the polarity of the protein surface, which is although inconsistent with the proposal for improved solubility, it is not in itself sufficient to eliminate the possibility of improvements in protein folding *in vivo*. To test this hypothesis more stringently, one of the groups proposed an experimental strategy in which the fraction of soluble protein was calculated as a fraction of the total expressed protein, given by the sum of the inclusion bodies and the soluble protein.

In conclusion, protein engineering of the GFPuv using a strategy based on random mutagenesis by epPCR was used to improve a desired functional feature of the protein. A random mutant library generated from single cycle of epPCR was submitted to a straightforward two-stage screening, which permitted the identification of a series of mutant proteins that demonstrated increased fluorescence emission intensity. Mapping of these mutants on the three-dimensional structure of the GFPuv allowed the interpretation of the results in terms of the location and physicochemical characteristics of the amino acid substitutions, providing insights into the structure–function relationship of the protein.

LABORATORY TIME

The DE course described here is a 40-hour course taught over a period of 5 days. This report is based on the experience of nine students divided into three groups. Each day was divided into two periods, and in the first period of 2 hours, theoretical classes introduced the principal themes of the structure/function relationship of proteins and the theoretical bases of current techniques used in DE such as epPCR and DNA shuffling. In the second period of 6 hours, the students were supervised in laboratory experiments based on a pre-established protocol for DE using the GFPuv.

The final activity of the course was a questionnaire regarding the content and presentation, which allowed the students to make suggestions for improvements. In the first version of the course, many students felt that their lack of previous experience in basic molecular biology procedures left them insecure in the laboratory. Therefore, on the basis of a self-evaluation at the start of the course, the students are now divided into groups in which at least one member has a reasonable level of molecular biology experience. With time, the number of students in the host institution who have passed through the course has increased, and we have a positive experience including these “veterans” as monitors to help the less experienced students. In addition, several students

commented on the rhythm of theoretical classes that left "too little time" to complete the practical classes. In light of these comments, the more recent versions of the course have fewer theoretical classes, which allows more time for informal discussion in the laboratory. Questions to initiate such discussion might include:

1. Why is it important to use a Taq DNA polymerase that does not have "proofreading" activity?
2. How do the buffer conditions influence the error rate of the Taq DNA polymerase? Is the epPCR truly random? How might bias in the nucleotide substitutions influence the final result?
3. How the random mutagenesis approach to protein engineering can provide insights as to the structural basis of protein function?

Acknowledgments— The authors thank the technical staff and postgraduate students in the DNA Polymorphism Laboratory of the Federal University of Pará, especially Silvanira Barbosa, for their commitment and assistance during the course.

REFERENCES

[1] F. H. Arnold, A. A. Volkov (1999) Directed evolution of biocatalysts, *Curr. Opin. Chem. Biol.* **3**, 54–59.
[2] F. H. Arnold, J. C. Moore (1997) Optimizing industrial enzymes by directed evolution, *Adv. Biochem. Eng. Biotechnol.* **58**, 1–14.
[3] R. Ruller, L. Deliberto, T. L. Ferreira, R. J. Ward (2008) Thermostable variants of the recombinant xylanase A from Bacillus subtilis produced by directed evolution show reduced heat capacity changes, *Proteins* **70**, 1280–1293.
[4] J. D. Bloom, F. H. Arnold (2009) In the light of directed evolution: Pathways of adaptive protein evolution, *Proc. Natl. Acad. Sci. USA* **106** (Suppl 1), 9995–10000.
[5] Z. Shao, H. Zhao, L. Giver, F. H. Arnold (1998) Random-priming in vitro recombination: an effective tool for directed evolution, *Nucleic Acids Res.* **26**, 681–683.
[6] E. T. Farinas, T. Bulter, F. H. Arnold (2001) Directed enzyme evolution, *Curr. Opin. Biotechnol.* **12**, 545–551.
[7] J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan, F. H. Arnold (2005) Evolving strategies for enzyme engineering, *Curr. Opin. Struct. Biol.* **15**, 447–452.
[8] W. P. Stemmer (1994) Rapid evolution of a protein in vitro by DNA shuffling, *Nature* **370**, 389–391.
[9] M. Chalfie (1995) Green fluorescent protein, *Photochem. Photobiol.* **62**, 651–656.
[10] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, D. C. Prasher (1994) Green fluorescent protein as a marker for gene expression, *Science* **263**, 802–805.
[11] W. Chiu, Y. Niwa, W. Zeng, T. Hirano, H. Kobayashi, J. Sheen (1996) Engineered GFP as a vital reporter in plants, *Curr. Biol.* **6**, 325–330.
[12] J. D. Plautz, R. N. Day, G. M. Dailey, S. B. Welsh, J. C. Hall, S. Halpain, S. A. Kay (1996) Green fluorescent protein and its derivatives as versatile markers for gene expression in living *Drosophila melanogaster*, plant and mammalian cells, *Gene* **173**, 83–87.
[13] A. Crameri, E. A. Whitehorn, E. Tate, W. P. Stemmer (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling, *Nat. Biotechnol.* **14**, 315–319.
[14] R. C. Cadwell, G. F. Joyce (1994) Mutagenic PCR, *PCR Methods Appl.* **3**, S136–S140.
[15] J. Sambrook, E. F. Fritsch, T. Maniatis (1989) Molecular Cloning: A Laboratory Manual, 3rd ed., New York, pp. E.3–E.11.
[16] N. Guex, M. C. Peitsch (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723.