Fluorescent microplate-based analysis of protein-DNA interactions II: immobilized DNA

Zhan-Ren Zhang[1], Marcus D. Hughes[1], Leonie J. Morgan[1], Albert F. Santos[2], and Anna V. Hine[1]

[1]Aston University, Birmingham and [2]Amersham Biosciences, Amersham plc, Cardiff, UK

A simple protein-DNA interaction analysis has been developed using both a high-affinity/high-specificity zinc finger protein and a low-specificity zinc finger protein with nonspecific DNA binding capability. The latter protein is designed to mimic background binding by proteins generated in randomized or shuffled gene libraries. In essence, DNA is immobilized onto the surface of microplate wells via streptavidin capture, and green fluorescent protein (GFP)-labeled protein is added in solution as part of a crude cell lysate or protein mixture. After incubation and washing, bound protein is detected in a standard microplate reader. The minimum sensitivity of the assay is approximately 0.4 nM protein. The assay format is ideally suited to investigate the interactions of DNA binding proteins from within crude cell extracts and/or mixtures of proteins that may be encountered in protein libraries generated by codon randomization or gene shuffling.

INTRODUCTION

In the preceding paper (1), we describe a novel assay format that is amenable to the high-throughput screening of protein-DNA interactions. This technique involves the direct immobilization of protein in microplate wells. Subsequent incubation with fluorescently labeled DNA enables the protein-DNA interaction to be quantified directly by residual fluorescence after a brief washing procedure. Because the assay conditions may be perturbed (i.e., by increasing the stringency of the washing procedure) and the results reevaluated, the assay is suitable for iterative analyses. After having successfully developed this format for use with purified proteins/ligands, we then wanted to exploit it in the analysis of protein mixtures against specific DNA targets. The end point application is in the field of combinatorial biology for the analysis of protein libraries.

In general, high-throughput screening of randomized protein libraries is achieved by biopanning. Typically, this involves the ligand of interest being immobilized onto the surface of a microplate, and randomized proteins in a display format, such as phage (2), yeast (3), bacterial (4) or ribosome proteins (5), are added free in solution. After the expressed libraries have been incubated with the immobilized ligand, the bound protein conjugates are washed repeatedly, eluted, and amplified. Sequential repetition of the process (an arbitrary number of times) yields multiple, relatively high-affinity receptor(s) that may be isolated and then identified by sequencing. Because display libraries involve the physical linkage between the phenotypes of protein libraries and their genotypes, deconvolution (identification of active proteins) can only be achieved by this final sequencing step at the very end of the process.

Although phage display of randomized proteins/peptides and subsequent biopanning have been used very effectively to examine the protein-DNA interactions of zinc fingers and to identify newly engineered proteins (6–10), the techniques used in biopanning are necessarily nondynamic in nature. A mixture of displayed proteins is added to the immobilized ligand, and no binding kinetics can be measured. Moreover, the ideal end point of the analysis, the point at which only the most strongly interacting protein remains bound to the ligand, cannot be determined. Thus, the end point of a biopanning experiment typically results in multiple clones with differing binding affinities for the specific ligand. The binding characteristics of each of these clones must then be determined individually, typically by ELISA-based methods (11), by surface plasmon resonance (SPR) (12), or most recently, by measuring bacteriophage binding to DNA binding site microarrays (13).

Here we report an assay that may be used to examine the interactions of mixtures of DNA binding proteins with a target DNA at any stage in the screening process. The proteins of interest (here, high and low-affinity/specificity zinc fingers) are fused to green fluorescent protein (GFP) and added to microplate wells in which the target DNA of interest has been immobilized. The interaction profile of the population of protein molecules with the immobilized DNA can then be investigated directly in an iterative manner. Ultimately, we plan to exploit this assay format in a display-free (i.e., protein only), sequencing-free deconvolution strategy for use with randomized protein libraries.

MATERIALS AND METHODS

Materials

Glutathione-S-transferase (GST)-ZFH and GST-ZFL proteins were expressed and purified as previously described (1). Enzymes used in DNA manipulation were purchased from New England Biolabs (Hitchin, UK) or MBI Fermentas (Vilnius, Lithuania). The GFP gene was amplified from the pGFPuv vector (BD Biosciences, Oxford, UK). Poly(dA)•poly(dT) DNA was obtained from Sigma (Gillingham, UK). Oligonucleotides for gene synthesis were obtained from Applied Biosystems (Warrington, UK). Binding site oligonucleotides BSH, 5′-Biotin-T_{10}GGGGCGGCTT$_{10}$-3′; BSHc, 5′-A$_{10}$AGCCGCCCCA$_{10}$-3′; BSL 5′-Cy™3-T_{10}GGGGGGGGGGT$_{10}$-3′; and BSLc, 5′-Cy3-A$_{10}$CCCCCCCCCA$_{10}$-3′ were obtained from MWG-Biotech (Milton Keynes, UK). High binding capacity black 96-well plates precoated with streptavidin were obtained from Perbio Science (Tattenhall, UK).

Synthesis of the ZFL Gene

The DNA sequence illustrated in Figure 1A was divided into six fully overlapping oligonucleotides (LJM 01–06) that were first annealed in pairs, and then the pairs themselves were ligated together prior to cloning. The fully assembled gene, which additionally encoded a histidine 6 (H_6) N-terminal motif (data not shown), was inserted into plasmid pT7-7 (14), and the sequence was verified. To express GST-ZFL, the gene, lacking the N-terminal histidine tag, was amplified by PCR and inserted into *Bam*HI-digested pGEX2TK. The resulting clone, pGEX-ZFL, encodes the low-affinity zinc finger protein as a C-terminal fusion to GST.

Preparation of Crude Protein Extracts Containing Zinc Finger-Green Fluorescent Protein Fusions

The GFP gene was amplified by PCR and inserted in-frame as a C-terminal fusion to the zinc finger genes in both pGEX-ZFL and pGEX-ZFH (1). Genes encoding the resulting zinc finger/GFP fusion proteins (lacking the N-terminal GST motif) were then amplified by PCR and subcloned into the vector pETcoco-1 (EMD Biosciences, Madison, WI, USA), which additionally encodes an N-terminal H_6 motif. Sequences were verified and the plasmids transformed into *Escherichia coli* Tuner™ (DE3) cells (EMD Biosciences). For the expression of H_6-ZFH-GFP and H_6-ZFL-GFP, one fresh colony was inoculated into 5 mL LB medium containing 12.5 µg/mL chloramphenicol and 0.2 % (w/v) glucose. The resulting culture was incubated with shaking overnight at 37°C, and then 1 mL was inoculated into 200 mL of the same medium, prewarmed to 37°C, and then further incubated at 37°C until an A_{600} of 0.4–0.5 was reached. Isopropyl-β-D-thiogalactoside (IPTG) was added to 0.5 mM, and the cells were cultured at 30°C overnight. The cells were harvested by centrifugation at 2,500× *g* for 7 min and resuspended in one-tenth volume of phosphate-buffered saline (PBS). The cells were then treated with 100 µg/mL lysozyme for 1 h at 25°C. Expressed proteins were released by six cycles of freezing and thawing. Cell debris was removed by centrifugation at 48,400× *g* for 10 min, and the clarified lysates were stored on ice. The relative concentration of zinc finger-GFP fusion proteins within the extracts was estimated by the quantification of GFP fluorescence against a calibration curve constructed using purified GFP.

Inverted Protein-DNA Interaction Assay

Double-stranded biotin-labeled BSH/BSHc DNA was prepared as previously described (1). The DNA (100 nM) was then added to the wells of a black streptavidin-coated microplate (100 µL/well), and the plate was incubated for 1 h with gentle shaking at room temperature. Unbound DNA was removed by the rapid inversion and subsequent washing of each well in 3× 200 µL PBS. Crude protein extract was diluted in PBS as required and added to the microplate
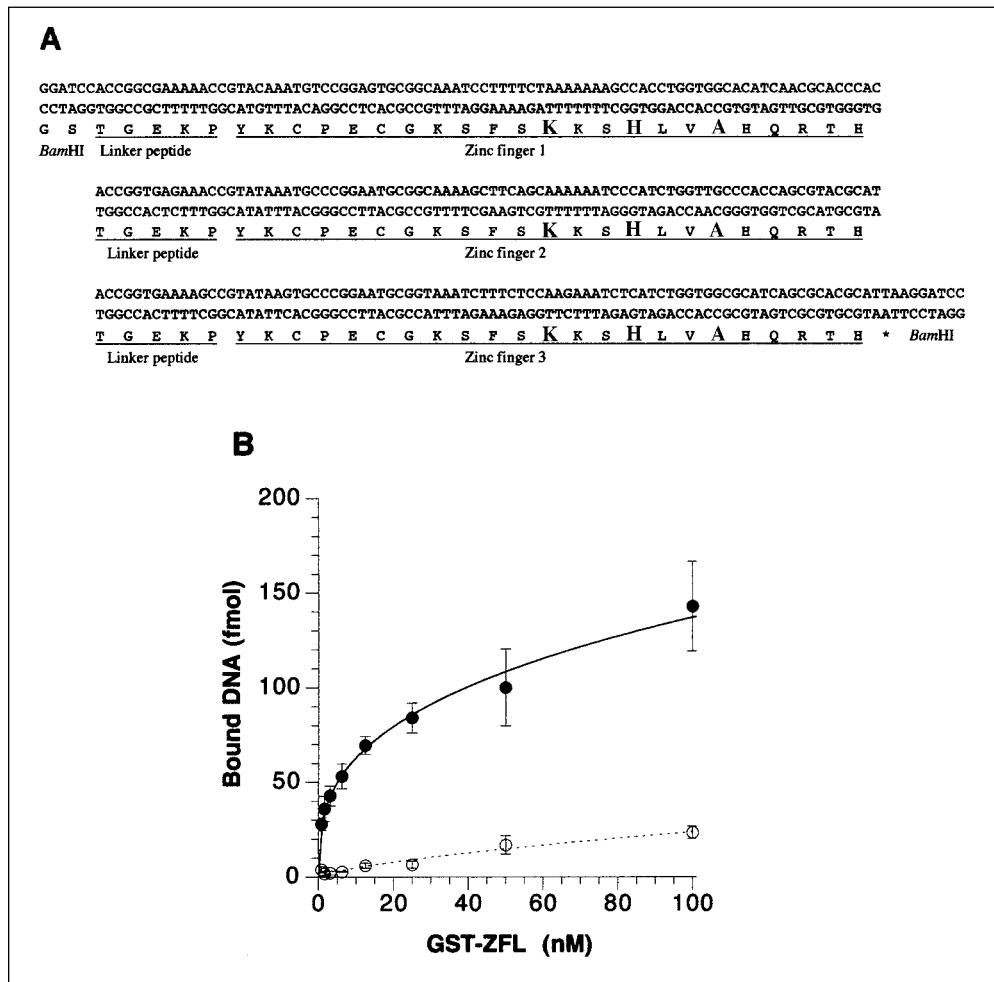

Figure 1. Zinc finger protein ZFL. (A) Peptide and derived DNA sequences. Residues at the potential DNA-contacting sites -1, +3, and +6 in the zinc finger α helices are indicated by the larger letters. To facilitate gene expression, the DNA sequence derived from the peptide includes *Escherichia coli* codon preferences (21), employing either first preference codons or second preference codons where necessary to incorporate restriction sites or to limit sequence repetition (maximum repeat length = 8 bp). (B) DNA binding by GST-ZFL. DNA binding to BSL/BSLc was analyzed as previously described (1) in either the absence (—●—) or presence (—○—) of 10 µg/mL poly(dA)•poly(dT) competitor DNA. The error bars represent standard deviation.

wells in 100 μL volumes. The plate was incubated as described above and then washed with 3× 200 μL PBST (10 mM sodium phosphate, pH 7.3, 140 mM NaCl, 0.1% Tween® 20). After the removal of the final wash by inversion, bound fluorescence was then measured in a SPECTRAmax® Gemini XS plate reader (Molecular Devices, Sunnyvale, CA, USA) with excitation and emission wavelengths of 390 and 510 nm, respectively, and a cut-off filter of 495 nm. All assays were performed in triplicate.

RESULTS AND DISCUSSION

A model zinc finger protein was designed to mimic the multiple low-affinity/low-specificity proteins that might typically be generated within a protein library, after the randomization of key DNA-contacting residues within a zinc finger protein. The low-affinity protein, named ZFL, consists of three repeats of a consensus zinc finger scaffold (15), with each zinc finger preceded by a Kruppel-type linker peptide (16). Key DNA-contacting residues in each zinc finger were those of the first zinc finger of transcription factor Sp1 (17,18). This zinc finger is known to have low affinity/specificity for the sequence 5′-GGG-3′ (19). We therefore predicted that ZFL would bind nominally to the sequence 5′-GGGGGGGGG-3′. Figure 1A illustrates the resulting peptide and encoding DNA sequences. The encoded protein ZFL was then used alongside ZFH, a high-affinity/high-specificity zinc finger protein (20), to develop an assay suitable for the quantitative screening of mixtures of DNA binding proteins.

In the preceding paper (1), a protein-DNA interaction assay was developed, in which unlabeled protein molecules were immobilized onto the surface of microplate wells, and fluorescently labeled DNA was added free in solution. This assay format was employed to verify the DNA binding characteristics of purified GST-ZFL protein. DNA binding of the predicted target sequence 5′-GGGGGGGGG-3′ was performed in both the absence and presence of unlabeled poly(dA)•poly(dT) competitor

DNA. Although ZFL does indeed bind to its predicted target DNA sequence, poly(dA)•poly(dT) DNA effectively competes for binding to ZFL, confirming its low sequence specificity (Figure 1B). ZFL also bound to the ZFH target DNA sequence 5′-GGGGCGGCT-3′ at levels that were comparable with ZFL binding to 5′-GGGGGGGGG-3′ in the presence of poly(dA)•poly(dT) competitor DNA (data not shown).

After establishing that ZFL could serve as an effective mimic of low-specificity DNA binding proteins, we then sought to develop an assay format suitable for screening protein mixtures such as libraries of DNA binding proteins or even crude protein extracts. We reasoned that the immobilization of proteins onto microplate wells, while ideal for studying single purified proteins, may not be so useful for screening protein mixtures because the overall binding capacity of the microplate wells would be expected to limit the quantity of the individual proteins immobilized. This in turn would lead to low binding of the fluorescently labeled DNA ligand and consequently to a low fluorescent signal. To circumvent this potential limitation, we elected to invert the assay format, immobilizing the DNA target (via biotin/streptavidin capture) and adding GFP-labeled DNA

binding proteins free in solution.

Preliminary studies employing the shallow-well plates described previously (1) revealed, after streptavidin immobilization, high background fluorescence at wavelengths required for GFP detection (data not shown). Consequently, standard microplates in which high levels of streptavidin had been immobilized commercially were employed in the development of the inverted assay. Using these plates, the linearity of the fluorescent response was established with respect to GFP concentration, giving a minimum sensitivity of approximately 0.4 nM GFP (Figure 2A). Crude *E. coli* extract containing an H_6-tagged GFP fusion of ZFH (1) was then employed to study the ZFH-DNA interaction using the inverted assay format. Figure 2B illustrates that the inverted assay format can indeed detect the protein-DNA interactions of a GFP-labeled zinc finger protein in the presence of myriad unlabeled proteins and demonstrates that protein binding saturates at approximately 20 nM total H_6-ZFH-GFP protein.

Would this detection remain achievable in the presence of other fluorescently labeled proteins? Such detection would be imperative if the assay is to be used to study protein libraries. Therefore, to mimic randomized protein libraries, we then sought to
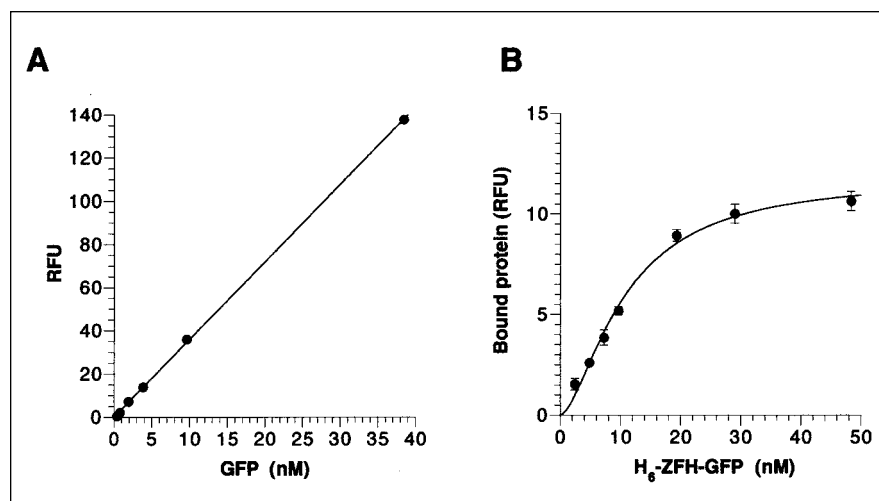

Figure 2. Protein DNA interactions monitored by green fluorescent protein (GFP) tagging. (A) GFP calibration curve. Purified GFP (BD Biosciences) was diluted in phosphate-buffered saline (PBS) to concentrations as indicated, and the fluorescence of single 100-μL samples was measured in plain black microplate wells as described in Materials and Methods. (B) Inverted protein-DNA interaction assay. Crude *Escherichia coli* lysate containing free histidine 6 (H_6)-ZFH-GFP (concentrations as indicated) was added to black microplate wells in which biotinylated target DNA had been immobilized via streptavidin capture. The assay was performed as described in the text. The error bars represent standard deviation. RFU, relative fluorescence units.

exploit the inverted assay format in the interrogation of mixtures of the high and low-specificity DNA binding proteins. Crude *E. coli* extracts containing either H_6-ZFL-GFP or H_6-ZFH-GFP were prepared, and the concentration of GFP-labeled protein within these extracts was estimated by quantification of the GFP fluorescence. Constant quantities of ZFL-containing extract were then combined with increasing quantities of ZFH-containing extract to give crude protein extracts containing ZFL:ZFH protein ratios of between 1000:1 and 50:1. The protein-DNA interactions of protein mixtures were then investigated using the inverted protein-DNA interaction assay (Figure 3). H_6-ZFL-GFP was kept constant in all measurements so that the background DNA binding by this protein could be estimated (Figure. 3, first data point). The contribution to DNA binding by the H_6-ZFH-GFP protein is barely detectable at 1000:1 and 750:1 protein ratios but is clearly discernable at 500:1 ZFL:ZFH ratios and below (Figure 3), which demonstrates the high-affinity DNA binding by H_6-ZFH-GFP. As in the absence of the ZFL protein (Figure 2B), binding again saturates at approximately 20 nM total H_6-ZFL-GFP protein.

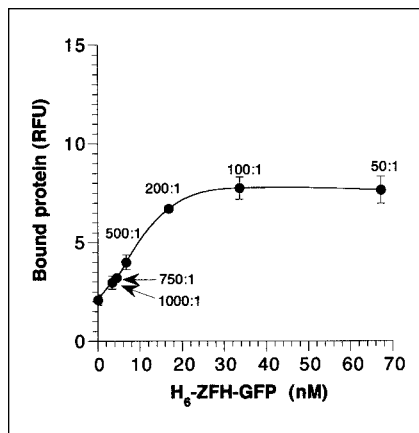As demonstrated in Figure 3, the

Figure 3. Screening protein mixtures. Green fluorescent protein (GFP)-labeled ZFH and ZFL proteins were combined at the molar ratios indicated in the presence of all soluble *Escherichia coli* proteins. The interactions of the GFP-labeled proteins with the ZFH target DNA sequence were analyzed using the inverted protein-DNA interaction assay as described in the text. The error bars represent standard deviation. RFU, relative fluorescence units; H_6, histidine 6 .

qualitative detection of the ZFH-DNA interaction (at the protein concentrations used in this experiment) is possible using the inverted assay format in the presence of up to 500-fold molar excess of the GFP-labeled low-specificity protein ZFL. Above this ratio, there is insufficient ZFH to produce a readable fluorescent signal.

For more quantitative measurements, an upper sensitivity limit must also be taken into account because quantitative estimations cannot be made once the immobilized ligand has become saturated. At the protein concentrations used in these experiments (Figure 3), the protein-DNA interactions of H_6-ZFH-GFP may be quantified in the presence of between approximately 200 and 500-fold molar excess of the ZFL-containing protein. [Because the H_6-ZFL-GFP concentration was constant for each data point, increased binding can be attributed directly to H_6-ZFH-GFP (Figure 3).] Below the lower ratio of 200:1, protein binding to the immobilized DNA is saturated, and for the quantification of mixtures in this range, reduced quantities of both proteins would be required to avoid saturation. Because ZFL shows low-affinity binding to several DNA sequences, we believe that the mixtures of ZFH and ZFL at varied ratios serve as an effective model for libraries of DNA binding proteins. Therefore, our results suggest that under the assay conditions described, it should be possible to quantify the combined interactions of mixtures containing between 200 and 500 different DNA binding proteins. Indeed, we are currently using the assay successfully to study the interactions of various mixtures of 400 zinc finger/GFP fusion proteins with single target DNA sequences. We are able to quantify the collective binding of such mixtures and see the differences between mixtures containing strongly and weakly binding proteins (Z.-R. Zhang, M.D. Hughes, A.F. Santos, and A.V. Hine, unpublished results). Similar to our assay for purified proteins (1), this quantification could be achieved at any time within the screening process and can therefore be used to make informed decisions about both washing conditions and when to switch from washing away unbound contaminants to the elu-

tion of bound protein.

Finally, this assay is not limited to the analysis of DNA binding proteins. The format should be directly applicable to the analysis of any protein-ligand interaction provided that (*i*) the ligand may be immobilized onto microplate wells; (*ii*) the protein can be expressed successfully as a GFP fusion; and (*iii*) the GFP fusion does not interfere with ligand binding.

REFERENCES

1. **Zhang, Z.-R., D. Palfrey, D.A. Nagel, P.A. Lambert, R.A. Jessop, A.F. Santos, and A.V. Hine.** 2003. Fluorescent microplate-based analysis of protein-DNA interactions I: immobilized protein. BioTechniques *35*: 980-986.
2. **Scott, J.K. and G.P. Smith.** 1990. Searching for peptide ligands with an epitope library. Science *249*:386-390.
3. **Kieke, M.C., B.K. Cho, E.T. Boder, D.M. Kranz, and K.D. Wittrup.** 1997. Isolation of anti-T cell receptor scFv mutants by yeast surface display. Protein Eng. *10*:1303-1310.
4. **Samuelson, P., E. Gunneriusson, P.-A. Nygren, and S. Stahl.** 2002. Display of proteins on bacteria. J. Biotechnol. *96*:129-154.
5. **Hanes, J. and A. Pluckthun.** 1997. In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. USA *94*:4937-4942.
6. **Choo, Y. and A. Klug.** 1994. Toward a code for the interactions of zinc fingers with DNA: selection of randomized zinc fingers displayed on phage. Proc. Natl. Acad. Sci. USA *91*:11163-11167.
7. **Isalan, M., A. Klug, and Y. Choo.** 2001. A rapid, generically applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. Nat. Biotechnol. *19*: 656-660.
8. **Rebar, E.J. and C.O. Pabo.** 1994. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science *263*: 671-673.
9. **Segal, D.J., B. Dreier, R.R. Beerli, and C.F. Barbas III.** 1999. Toward controlling gene expression at will: selection and design of zinc finger domains recognising each of the 5′-GNN-3′ DNA target sequences. Proc. Natl. Acad. Sci. USA *96*:2758-2763.
10. **Wolfe, S.A., E.I. Ramm, and C.O. Pabo.**

2000. Combining structure-based design with phage display to create new Cys$_2$His$_2$ zinc finger dimers. Structure *8*:739-750.

11. **Choo, Y. and A. Klug.** 1994. Selection of DNA binding sites for zinc finger using rationally randomized DNA reveals coded interactions. Proc. Natl. Acad. Sci. USA *91*:11168-11172.

12. **Wu, H., W.-P. Yang, and C.F. Barbas III.** 1995 Building zinc fingers by selection: toward a therapeutic application. Proc. Natl. Acad. Sci. USA *92*:344-348.

13. **Bulyk, M.L., X. Huang, Y. Choo, and G.M. Church.** 2001. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. Proc. Natl. Acad. Sci. USA *98*: 7158-7163.

14. **Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl.** 1990. Expression using the T7 RNA polymerase promotor system, p. 16.2.1-16.2.11. *In* Current Protocols in Molecular Biology. Wiley-Interscience, New York.

15. **Krizek, B.A., B.T. Amman, V.J. Kilfoil, D.L. Merkle, and J.M. Berg.** 1991. A consensus zinc finger peptide: design, high-affinity metal binding, a pH-dependent structure, and a His to Cys sequence variant. J. Am. Chem. Soc. *113*:4518-4523.

16. **Choo, Y. and A. Klug.** 1993. A role in DNA binding for the linker sequences of the first three zinc finger proteins of TFIIIA. Nucleic Acids Res. *21*:3341-3346.

17. **Berg, J.M.** 1992. Sp1 and the subfamily of zinc finger proteins with guanine rich binding sites. Proc. Natl. Acad. Sci. USA *89*: 11109-11110.

18. **Shi, Y. and J.M. Berg.** 1995. A direct comparison of the properties of natural and designed zinc-finger proteins. Chem. Biol. *2*:83-89.

19. **Yokono, M., N. Saegusa, K. Matsushita, and U. Sugiura.** 1998. Unique DNA-binding mode of the N-terminal zinc finger of transcription factor Sp1. Biochemistry *37*: 6824-6832.

20. **Desjarlais, J.R. and J.M. Berg.** 1993. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. Proc. Natl. Acad. Sci. USA *90*:2256-2260.

21. **Nakamura, Y., T. Gojobori, and T. Ikemura.** 2000. Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucleic Acids Res. *28*:292.

Address correspondence to Anna V. Hine, School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham, B4 7ET, UK. e-mail: a.v.hine@aston.ac.uk

Nitrogenase genes in PCR and RT-PCR reagents: implications for studies of diversity of functional genes

Jonathan P. Zehr, Lori L. Crumbliss, Matthew J. Church,
Enoma O. Omoregie, and Bethany D. Jenkins
University of California, Santa Cruz, Santa Cruz, CA, USA

*Studies of the diversity of microorganisms in the environment have been facilitated by use of PCR and reverse transcription PCR (RT-PCR). Inhibition of the PCR by complex sample matrices and low abundance of some target microorganisms require the use of high-sensitivity amplification procedures, involving a large number of cycles or nested PCR methods. Using these methods, we frequently observed contamination of the amplification reagents, including polymerases, by genomic DNA containing nitrogenase (*nifH*) and rRNA genes. Contaminating genes were sequenced and found to belong to a variety of rRNA clades, but only three major *nifH* clades. These sequence types included a few *nifH* sequences reported in previous studies of the environment. Contamination could be reduced by restriction digestion and ultrafiltration of PCR reagents, but efficiency of amplification was also reduced. Our results suggest that studies relying on large numbers of PCR amplification cycles to assess environmental gene diversity should take precautions to assure that clone libaries generated from amplified PCR products are not the result of contaminated PCR reagents.*

INTRODUCTION

The study of the diversity of microorganisms in the environment has been facilitated by the development of DNA and RNA amplification procedures (1). The diversity of microbial assemblages can be investigated by amplification of rRNA genes (2,3) or genes involved in specific metabolic capabilities, such as ammonia oxidation (4), dissimilatory nitrate reduction (5,6), and nitrogen fixation (7,8).

Environmental applications of the PCR are often hampered by inhibition of amplification by the sample matrix, and target microorganisms are often present in extremely low abundance. As a result, high-sensitivity methods such as nested PCR and reverse transcription PCR (RT-PCR) are often used. Target microorganisms often have not been previously cultivated (9–11), making it difficult to determine if PCR-amplified sequences were obtained from the environment or from low-level contamination from sample handling or reagents.

In environmental studies that target previously uncultivated microorganisms, there are many potential sources of contamination with microorganisms from reagents or laboratory surfaces. Genomic DNA contamination of thermal-stable polymerases and other reagents has been reported in rRNA studies (12–14). Contamination often remains problematic even though a number of procedures can be used to eliminate or reduce contamination (12,15–18).

Targeting genes that are not universally distributed among microorganisms, such as nitrogenase, ammonia monooxygenase, or methane monooxygenase (4,19,20), would be expected to be less prone to contamination from laboratory or reagent sources, since these genes may not be present in many of the microorganisms that are typically found in laboratory reagents and enzyme preparations. However, we found that commercial reagents are routinely contaminated with low levels of nitrogenase genes (*nifH*, encoding the Fe protein) as well as 16S rRNA genes.

MATERIALS AND METHODS

PCR Reagents

Only commercial water (nuclease-free; Ambion, Austin, TX, USA) and