High-throughput Pyrosequencing™ of a phage display library for the identification of enriched target-specific peptides

Ahad Rahim, Charles Coutelle, and Richard Harbottle
Imperial College, London, UK

BioTechniques 35:317-324 (August 2003)

Gene therapy clinical trials have highlighted the importance of specific cellular/tissue targeting of gene delivery vectors. Phage display libraries are powerful tools for the selection of novel peptide ligands as targeting moieties because of their high-throughput screening potential. However, a severe rate-limiting step in this procedure in terms of time, numbers, and cost is the sequence identification of selected phages. Here we describe the application of Pyrosequencing™ technology for sequencing phage isolates after panning a random 7-mer peptide expressing phage library against the A549 bronchial epithelial cell line to search for enrichment of possible targeting peptides. Pyrosequencing allows sequencing of 96 phages at one time in approximately 45 min at only a sixth of the cost of conventional sequencing methods. Using this technology, we have identified four sequences of interest. A phage binding assay revealed that three of the four sequences show a significant increase in binding abilities and specificity for A549 cells when compared to an unrelated cell line.

INTRODUCTION

Vectors developed for gene therapy have distinct advantages and disadvantages, and targeting has been suggested as a vital step towards overcoming some of the current drawbacks (1). Specific targeting to cellular receptors would increase therapeutic efficiency because of enhanced potency, reduction in dosage, and avoidance of gene delivery to nontarget cells. This would also exploit the cellular mechanism of receptor-mediated endocytosis and has therefore become the focus of much research (2–8).

Phage display technology has been used extensively to search for novel ligands to cellular receptors. A genetic fusion with the minor coat protein pIII of filamentous phage allows the display of foreign proteins on the surface of phage (9,10) and has become a powerful tool in the form of high-complexity combinatorial libraries. Phage displaying peptides that have the greatest affinity to receptors on the target of interest are enriched against nonspecific binding by consecutive rounds of panning and enrichment under increasing stringency. Enriched phages need then

to be sequenced to identify the displayed peptide for investigation as a potential cell-specific targeting ligand. However, because of the high complexity of the library and the presence of phages with different affinities to the target cells, an abundance of phage isolates needs to be sequenced to identify the enriched population. Sequencing by conventional means is time-consuming and expensive and is also usually tailored for longer sequences, whereas in the context of phage analysis, only very short sequence analysis is required.

Pyrosequencing™ (Pyrosequencing AB, Uppsala, Sweden) is a nonelectrophoretic, real-time method of DNA sequencing that typically allows 20–30 bases to be sequenced, although greater stretches have been reported (11–13), and has been used in a wide range of research applications (11,14,15). The procedure begins with the immobilization of a PCR-generated biotinylated template derived from the analyzed DNA to streptavidin-coated Sepharose® beads and the dissociation of its DNA strands to allow a primer sequence to anneal. This is an automated system that uses a 96-well plate then adds nucleotides sequentially together with an enzyme sys-

tem and substrate. The addition of nucleotides to the primer by Klenow DNA polymerase produces pyrophosphate. Pyrophosphate triggers an enzyme cascade system involving apyrase and ATP-sulfurylase, resulting in the luciferase-driven production of light that is detected by a charge-coupled device (CCD) camera, and the obtained information is converted to a graphical form. Pyrosequencing is quick and cost-effective, thus allowing high-throughput DNA sequencing as essential for rapid phage library analysis.

Here we describe the application of Pyrosequencing technology to sequence phages in a 96-well plate following three rounds of panning upon the A549 bronchial epithelia cell line using a 7-mer amino acid phage display library to identify potential ligands to target vectors for treatment of cystic fibrosis.

MATERIALS AND METHODS

Biopanning

A549 human bronchial airway epithelial cells were grown to confluency in a 6-well plate. Cells were maintained in Dulbecco's modified Eagle's minimal (DMEM) with L-glutamine and supplemented with 10% fetal bovine serum (FBS) (both from Invitrogen, Paisley, UK) and 5% penicillin-streptomycin. Growth conditions were maintained at 37°C and 5% CO₂. Cell medium was replaced with serum-free DMEM and incubated for 2 h. The medium was replaced with 1% bovine serum albumin (BSA; Sigma, Dorset, UK) made up in phosphate-buffered saline (PBS; Invitrogen) and incubated for 1 h. Cells were washed with Tris-buffered saline (TBS), and 2 × 10¹¹ phages were added from the Ph.D.-7 phage display library (New England Biolabs, Beverly, MA, USA) in 1 mL TBS and left on cells for 60 min at 4°C before washing away nonspecific phages in 10 washes using

TBS. Bound phage were rescued by adding 1 mL 0.2 M glycine-HCl, pH 2.2, for 10 min and then neutralized by adding 150 µL 1 M Tris-HCl, pH 9.1.

The eluted phage were amplified and titered according to the manufacturer's protocol using *Escherichia coli* strain ER2738. Phages (2 × 10¹¹) were reappplied to A549 cells in subsequent panning rounds. Second and third rounds of panning were performed with increasing stringency by decreasing binding times to 15 and 10 min, respectively. The washing stringency was also progressively increased by the addition of 0.1% and 0.2% Tween® 20 detergent (Sigma) in TBS. Ninety-six phage infected colonies were then picked and individually amplified in 1 mL early log-phase ER2738 and left to incubate at 37°C in a shaking incubator for 4 h. The cultures were microcentrifuged at 13,000 rpm for 30 s, and the super-

natant was collected and respun. The upper 80% of the supernatant was collected and stored at 4°C. Each of the 96 tubes represents a stock of the individual phage picked after the third round of panning.

Preparation of Sample Template

The first step of phage preparation for sequencing involved the production of a biotinylated PCR product of the randomized region of the pIII gene. The following PCR primers were designed to anneal with the constant sequence flanking the variable region and synthesized with the reverse primer including a 5' biotin label: 5'-ATTCGCAATTCCTTTAGTGGTA-3' and 5'-biotin-GGGATTTTGCTAAACAACCTT-3'. The optimized PCR in a total of 50 µL containing GeneAmp® 1× PCR Gold Buffer (Applied Biosystems, Fos-

Figure 1. Pyrogram and electropherogram analysis of phage DNA sequence. An isolated phage that showed enrichment after Pyrosequencing (A) was resequenced using a BigDye Terminator cycle sequencing kit, as presented by the electropherogram reading (B). The predicted DNA sequence generated by the Pyrosequencing program (which can be validated by manual reading of the pyrogram, the sequence of which is indicated above the peaks) and the sequence derived from the electropherogram match exactly and translate to the peptide sequence HSMQLST (C).

ter City, CA, USA), 2 mM MgCl₂ (Applied Biosystems), 0.2 mM dNTP mixture (Invitrogen), 0.2 μM forward primer (Sigma-Genosys, Cambridge-shire, UK), 0.2 μM reverse primer (Sigma-Genosys), 1.25 U AmpliTaq Gold® (Applied Biosystems) was carried out at 95°C for 5 min, followed by 45 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 60 s, with a final extension at 72°C for 10 min.

Immobilization of PCR products was carried out using 384 μL streptavidin-coated Sepharose beads (Amersham Biosciences, Piscataway, NJ, USA) mixed with 3840 μL binding buffer (10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA, 0.1% Tween 20). Forty microliters of the streptavidin binding buffer solution were added to 40 μL of each PCR in a 96-well plate and incubated at room temperature for 10 min, with shaking at 12,000 rpm. The samples were transferred to a 96-well filter plate (Millipore, Bedford, MA, USA) and placed on a vacuum manifold where vacuum was applied. Fifty microliters of denaturing solution (0.2 M NaOH) were added and left to incubate for 1 min, after which vacuum was applied. The beads were washed twice with 150 μL washing buffer (10 mM Tris-acetate, pH 7.6), applying the vacuum during the process. Five microliters of the following 3 μM sequencing primer were dispensed into each well of a PSQ™ 96 Plate Low (Pyrosequencing AB): 5'-TGGTACCTTTCTATTCTCAC-3' (Sigma-Genosys). Fifty microliters of annealing buffer (20 mM Tris-acetate, pH 7.6, 2 mM MgAc₂) were used to resuspend the beads on the filter plate before transferring 40 μL from each well to the PSQ 96 Plate Low wells containing the sequencing primer. The plate was heated to 80°C for 2 min using a PSQ Sample Prep Thermoplate Low (Pyrosequencing AB) and then left at room temperature for 10 min.

Pyrosequencing

The PSQ 96 Plate Low containing the primer annealed to the template in each well was loaded into an automated PSQ 96MA, and Pyrosequencing was carried out using a PSQ SNP 96 reagent kit and dispensation cartridge (Pyrosequencing AB). Nucleotide dis-

pensation was entered as TCT followed by cyclic rounds. The software was set to analyze all 96 wells, and progress was monitored in real time with the option to stop or pause the reaction.

Alternative Automated Sequencing

Sequences produced by Pyrosequencing were verified using an ABI PRISM® 377 DNA sequencer using a BigDye™ Terminator cycle sequencing kit (both from Applied Biosystems).

Phage Binding Assay

Binding studies were carried out on A549 cells, and the cervical epithelium-derived HeLa cells using ELISA. The cells were grown to confluency in 96-well plates and blocked with 1% BSA for 1 h. The blocking buffer was removed, and wells were gently washed 6 times with TBS. Phages (2×10^{10}) in 100 μL TBS were added to each well and incubated for 1 h at 4°C. Unbound phages were removed by washing wells 6 times with TBS containing 0.2% Tween 20. Anti-M13 antibody (200 μL) conjugated to horseradish peroxidase (HRP) (Amersham Biosciences) diluted 1:5000 in blocking buffer was added to each well and incubated for 1 h, followed by washing 6 times. Detection was carried out by adding 200 μL 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (Sigma) and measurement of absorbance at 420 nm after 1 h.

RESULTS

PCR Amplification of Randomized pIII Region for Sequencing Template

To apply Pyrosequencing analysis of individual phages selected after biopanning enrichment, we produced biotinylated PCR fragments. These included the randomized region of the pIII gene encoding the displayed peptide and a short region to which a sequencing primer can anneal. The sequencing primer was designed to start DNA synthesis 3 bp preceding the randomized region. These 3 bp preceding the randomized sequence and the known sequence following it can act as a common sequence and a guide to check the

accuracy and reliability of the data.

The addition of intact phage to the PCR mixture and processes as described in Materials and Methods section resulted in the reaction producing an amplified fragment of the expected size.

Pyrosequencing and Identification of Potential Ligands

Pyrosequencing was performed on 96 PCR samples in one plate in approximately 45 min. Peaks representing nucleotide incorporation were displayed during this process on the computer screen in real time with the option to either pause or stop the reaction. The data are represented as a Pyrogram™ (Pyrosequencing AB) as shown in Figure 1A in comparison to Sanger-sequencing (Figure 1B). The DNA sequence was determined and displayed by the software package and confirmed by manual reading the Pyrograms. In Figure 1A, the letters below the Pyrogram represent the sequential dispensation of the nucleotides A, C, T, or G to the wells of the 96-well plate in the course of the Pyrosequencing procedure. Successful incorporation of a particular nucleotide to the synthesized strand (as dictated by the complementary sequence) results in a light signal produced by the enzyme cascade, as mentioned previously. This is represented as a peak on the Pyrogram. If no addition of the dispensed nucleotide to the synthesized strand takes place, then no peak is seen. The height of the peak is proportional to the number of adjacent nucleotides incorporated into the synthesized strand. The letters shown above the peaks of the Pyrogram designate how the Pyrogram should be read (i.e., the obtained sequence). The Pyrosequencing software is programmed specifically in this case to dispense T, C, and T as the first three nucleotides since they represent the end of the known constant sequence, whereas thereafter cyclic dispensations of nucleotides A, T, C, and G occurs, allowing for sequencing of the randomized region coding for the peptide from this point onwards.

The obtained 96 sequences were analyzed for enrichment, and four reoccurring sequences were identified. They were HAIYPRH (3/96), HSMQLST (4/96), STQAMFQ (9/96), and a se-

quence that was not complete, difficult to read, and contained many spurious peaks (21/96). However, this sequence could still be recognized as enriched because the individual pattern of all 21 sequences was the same.

Figure 2 shows the frequency at which these sequences appeared as a percentage of the 96 phages sequenced and also failed sequences and those that appeared only once. Failed sequences represent those sequences that produce no Pyrogram or contain too many spurious peaks to be read.

The enriched sequences were also determined by ABI automated Sanger sequencing and found to match exactly (Figure 1). This also confirmed the enrichment of incomplete sequence and allowed full identification as FLSKPP. The failure of Pyrosequencing to give a complete sequence in this case may be due to a long 5-nucleotide homopolymer T stretch.

Determination of Binding Efficiency and Specificity of Peptides Using Phage Binding Assay

The isolated phage displaying the STQAMFQ, HSMQLST, HAIYPRH, and FLSKPP were further investigated for their binding properties to target cells. A phage binding assay was used to show that phage displaying these peptide sequences had increased binding properties for A549 cells when compared to a wild-type M13 phage that displays no peptide on its coat protein. A549 cells were grown to confluency in a 96-well plate to which the phage were added and incubated at 4°C to prevent any receptor-mediated internalization. Unbound phages were washed away, followed by the addition of an anti-M13 HRP conjugate and subsequent washing. Addition of the substrate for HRP allows quantification of binding by colorimetric means. The assay was also carried out on the unrelated HeLa cells as a control to look for cell-type specificity. Figure 3 shows that phages expressing the peptides HAIYPRH, STQAMFQ, and FLSKPP show significant-

ly increased binding efficiency to A549 cells when compared to wild-type phage. When binding efficiencies of the phages were compared on A549 and HeLa cells, phages expressing HAIYPRH, STQAMFQ, and FLSKPP showed a significantly increased specificity to A549 cells, while the wild-type phages showed almost identical levels of binding to both cell types. Phages expressing the HAIYPRH peptide clearly show the greatest levels of binding to the A549 cells and are approximately twice that shown by the wild-type phages.

DISCUSSION

Phage display technology has progressed rapidly, having been extended from in vitro experiments to in vivo use (16) as well as by development of new amino acid recognition software (17). However, little progress has been made in the development of sequencing the selected phage. Current methods of sequencing such as the manual or automated Sanger method are costly, time-consuming, and laborious.

We applied a novel, non-

Figure 2. Enriched phages identified after Pyrosequencing. Ninety-six phages were sequenced using Pyrosequencing technology to identify potential ligands to A549 cells. Four recurring sequences were found, three of which were sequenced in full and one that could not be fully sequenced represented in the figure as incomplete sequence.

electrophoretic, high-throughput method of sequencing called Pyrosequencing to phage display screening. Following panning of a phage display library on A549 bronchial epithelium-derived cells, Pyrosequencing of 96 phage isolates determined four enriched DNA sequences, of which three were fully and one incompletely identified. The four enriched DNA sequences represented the peptide sequences, HAIYPRH, FLSKPP, HSMQLST, and STQAMFQ, as potential ligands. Our data show that the majority of the 96 phages used for Pyrosequencing was sequenced without problem. Approximately 15% of the sequences failed. We speculate this to be partially attributed to the difficulty that the Pyrosequencing software has with stretches of homopolymer nucleotides, which has also been encountered in other studies using Pyrosequencing (14). The enriched DNA sequence that could only be partially identified also included a homopolymer

Figure 3. Phage binding assay comparing levels of binding of isolated phages and wild-type phages on A549 and HeLa cells. The isolated phage displaying HAIYPRH, STQAMFQ, and FLSKPP show significantly increased levels of binding to A549 cells compared with a control wild-type phage displaying no peptide on its surface coat. Binding of the isolated phage to HeLa cells was lower and significantly so in the case of HAIYPRH, STQAMFQ, and FLSKPP. Binding of the wild-type phage was approximately equal in both cell types. The asterisks highlight statistical significance of differences in phages binding to the different cell types and compared with the wild-type phages. Statistical analysis of data was carried out using a one-tailed *t* test assuming equal variance with Bonferroni correction ($P < 0.05$).

T stretch. However, it could be identified as enriched by manual inspection of the Pyrograms seen by repeated occurrence of identical sequence patterns. In the case of such sequences, automated Sanger sequencing would be required for complete characterization of the sequence. The number of failed sequences obtained by Pyrosequencing is comparable to failures with other sequencing methods such as Sanger sequencing (18). The accuracy of the Pyrosequencing data was shown by comparison with data from a conventional sequencing method.

The potential for Pyrosequencing to sequence-screen other phage libraries may depend on the length of the randomized region to be sequenced. The library used in this study contains a randomized region of 21 bp coding for 7 amino acids, but a variety of libraries are available with longer peptides displayed. They will require longer readings of sequence, which has been reported for pyrosequences up to 100 bp (13). No other sequencing technology currently available allows for the sequencing of 96 phages in less than 1 h. We also estimate that the cost of each phage sequenced using Pyrosequencing is about one-sixth of those associated with conventional automated sequencing.

The enriched peptide sequences were investigated by an ELISA-based binding assay to test for increased binding efficiency to A549 cells. This showed that phages expressing the peptides HAIYPRH, STQAMFQ, and FSLSKPP display significantly increased binding to A549 cells compared with control wild-type phages. The phages displaying the HAIYPRH peptide showed the highest levels of binding, approximately twice that of the control phages. When the assay was conducted on the unrelated cervical epithelium-derived HeLa cells, binding of HAIYPRH, STQAMFQ, and FSLSKPP was significantly lower with no significant difference to wild-type phages, which was approximately the same on both cell lines. By increasing the stringency through decreasing the time allowed for phage to bind, it is possible that background binding could be further reduced.

Peptide sequences exhibiting increased binding and specificity to airway epithelium where receptors gener-

ally used by conventional gene therapy vectors are not present are highly desirable. We believe that such peptides can be conjugated to delivery vectors, resulting in greater efficiency of gene transfer to the airway epithelium for treatment of cystic fibrosis.

ACKNOWLEDGMENTS

This work was supported by the March of Dimes Birth Defects Foundation. We wish to thank Karen Walker and Jenny Dunker from Pyrosequencing AB for technical advice and the staff from the Hammersmith Hospital, London for the use of equipment.

REFERENCES

1. Somia, N. and I.M. Verma. 2000. Gene therapy: trials and tribulations. *Nat. Rev. Genet.* 1:91-99.
2. Wu, G.Y. and C.H. Wu. 1986. Targeted inhibition of transferrin-mediated iron uptake in Hep G2 hepatoma cells. *J. Biol. Chem.* 261:16834-16837.
3. Haensler, J. and F.C. Szoka, Jr. 1993. Synthesis and characterization of a trigalactosylated bisacridine compound to target DNA to hepatocytes. *Bioconjug. Chem.* 4:85-93.
4. Sosnowski, B.A., A.M. Gonzalez, L.A. Chandler, Y.J. Buechler, G.F. Pierce, and A. Baird. 1996. Targeting DNA to cells with basic fibroblast growth factor (FGF2). *J. Biol. Chem.* 271:33647-33653.
5. Merwin, J.R., E.P. Carmichael, G.S. Noell, M.E. DeRome, W.L. Thomas, N. Robert, G. Spitalny, and H.C. Chiou. 1995. CD5-mediated specific delivery of DNA to T lymphocytes: compartmentalization augmented by adenovirus. *J. Immunol. Methods* 186:257-266.
6. Hart, S.L., R.P. Harbottle, R. Cooper, A. Miller, R. Williamson, and C. Coutelle. 1995. Gene delivery and expression mediated by an integrin-binding peptide. *Gene Ther.* 2:552-554.
7. Buschle, M., M. Cotten, H. Kirlappos, K. Mechtler, G. Schaffner, W. Zauner, M.L. Birnstiel, and E. Wagner. 1995. Receptor-mediated gene transfer into human T lymphocytes via binding of DNA/CD3 antibody particles to the CD3 T cell receptor complex. *Hum. Gene Ther.* 6:753-761.
8. Wagner, E., M. Zenke, M. Cotten, H. Beug, and M.L. Birnstiel. 1990. Transferrin-polycation conjugates as carriers for DNA uptake into cells. *Proc. Natl. Acad. Sci. USA* 87:3410-3414.
9. Sidhu, S.S. 2001. Engineering M13 for phage display. *Biomol. Eng.* 18:57-63.
10. Noren, K.A. and C.J. Noren. 2001. Construction of high-complexity combinatorial phage display peptide libraries. *Methods* 23:169-178.
11. Garcia, C.A., A. Ahmadian, B. Gharizadeh, J. Lundeberg, M. Ronaghi, and P. Nyren. 2000. Mutation detection by pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene. *Gene* 253:249-257.
12. Andreasson, H., A. Asp, A. Alderborn, U. Gyllensten, and M. Allen. 2002. Mitochondrial sequence analysis for forensic identification using Pyrosequencing™ technology. *BioTechniques* 32:124-133.
13. Gharizadeh, B., T. Nordstrom, A. Ahmadian, M. Ronaghi, and P. Nyren. 2002. Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer. *Anal. Biochem.* 301:82-90.
14. Pourmand, N., E. Elahi, R.W. Davis, and M. Ronaghi. 2002. Multiplex Pyrosequencing. *Nucleic Acids Res.* 30:e31.
15. Gharizadeh, B., M. Kalantari, C.A. Garcia, B. Johansson, and P. Nyren. 2001. Typing of human papillomavirus by pyrosequencing. *Lab. Invest.* 81:673-679.
16. Pasqualini, R. and E. Ruoslahti. 1996. Organ targeting in vivo using phage display peptide libraries. *Nature* 380:364-366.
17. Arap, W., M.G. Kolonin, M. Trepel, J. Lahdenranta, M. Cardo-Vila, R.J. Giordano, P.J. Mintz, P.U. Ardel, et al. 2002. Steps toward mapping the human vasculature by phage display. *Nat. Med.* 8:121-127.
18. Agaton, C., P. Unneberg, M. Sievertzon, A. Holmberg, M. Ehn, M. Larsson, J. Odeberg, M. Uhlen, and J. Lundeberg. 2002. Gene expression analysis by signature pyrosequencing. *Gene* 289:31-39.

Received 2 May 2003; accepted 27 March 2003.

Address correspondence to Ahad Rahim, Gene Therapy Research Group, Biomedical Sciences, Imperial College Faculty of Medicine, Sir Alexander Fleming Building, Exhibition Road, South Kensington, London, SW7 2AZ, UK. e-mail: ahad.rahim@imperial.ac.uk