Retargeting of pre-set regions on chromosome for high gene expression in mammalian cells

Peng Jiao, Christine Chang, Kelly Kral, Jonathan Rogger, Nicolas Wyhs, Daniel Wang
Biotechnology Process Engineering Center, Massachusetts Institute of Technology

Abstract-- **We have developed a system to hunt and reuse special gene integration sites that allow for high and stable gene expression. A vector, named pRGFP8, was constructed. The plasmid pRGFP8 contains a reporter gene, *gfp*2 and two extraneous DNA fragments. The gene *gfp*2 makes it possible to screen the high expression regions on the chromosome. The extraneous DNA fragments can help to create the unique loci on the chromosome and increase the gene targeting frequency by increasing the homology. After transfection into Chinese hamster ovary cells (CHO) cells, the linearized pRGFP8 can integrate into the chromosome of the host cells and form the unique sites. With FACS, 90 millions transfected cells were sorted and the cells with strongest GFP expression were isolated, and then 8 stable high expression GFP CHO cell lines were selected as candidates for the new host cell. Taking the unique site created by pRGFP8 on the chromosome in the new host cells as a targeting locus, the *gfp*2 gene was replaced with the gene of interest, human *ifn-gamma*, by transfecting the targeting plasmid pRIH-IFN. Then using FACS, the cells with the dimmest GFP fluorescence were selected. These cells showed they had strong abilities to produce the protein of interest, IFN-gamma. During the gene targeting experiment, we found there is positive correlation between the fluorescence density of the GFP CHO host cells and the specific production rate of IFN-gamma. This result shows that the strategy in our expression system is correct: the production of the interesting protein increases with the increase fluorescence of the GFP host cells. This system, the new host cell lines and the targeting vector, can be utilized for highly expressing the gene of interest. More importantly, by using FACS, we can fully screen all the transfected cells, which can reduce the chances of losing the best cells.**

Key words: Gene retargeting, GFP, high expression, mammalian cells

I. INTRODUCTION

Genetically engineered mammalian cells play an essential role in pharmaceutical protein production because the mammalian cells are close to cells found in the human body and their products are similar to human's(1). Multiple factors affect recombinant protein production in mammalian cells. Factors, such as the specific cell line used, the expression vector, chromosomal integration site, copy number of the integrated recombinant gene, selection procedures, cell-culture conditions and medium employed can all have a large impact on overall production(2).

Transcription is the first and most important step of recombinant gene expression. Two basic strategies are used to maximize gene expression at the transcriptional level: multiple interesting gene copy number and finding a special gene integration site that allows for high and stable gene expression.

The copy number of the interesting gene can have a significant influence on the expression characteristics. On average, higher gene copy number correlate with higher expression levels. The overexpression of recombinant protein in Chinese Hamster Ovary (CHO) cells has been achieved by gene amplification. In this method, the dihydrofolate reductase (*dhfr*) gene is cotransfected with the gene of interest, then selected with the amplification drug methotrexate. (3,4). The final selected cell may contain hundreds or even thousands transgene copy numbers. However, the procedure to isolate high copy number clones is extremely time-consuming. The initial expression level is not meaningful, and amplifiable clones cannot be identified in any other way. Therefore, a large number of clones must be handled initially. Additionally, the cells with multiple copies of the transgene suffer another critical defect: genetic instability. To prevent the loss of productivity, the cells must be grown continuously in the presence of the respective selection drug.

Based on these considerations, in recent years, efforts have been made to use cell clones with single or low transgene copy numbers to overcome instability without a loss in

productivity. Many reports show that single-copy integration of a transgene can produce a high and stable expression. (5, 6, 7)

Here we report a strategy to hunt and create the region(s) on the chromosome that allows high and stable gene expression. The specific region(s) are modified to create a unique locus for gene retargeting in order to introduce the gene of interest.

II. MATERIALS AND METHODS

Genes:

gfp and *gfp*2 genes: *gfp* gene was generously donated from Hastings' Lab, Department of Biology, Harvard University. The gene *gfp*2 has been modified by removing the *Spe*I site from *gfp* by changing ACTAGT to ACTTGT through site-mutagenesis. The expressing product of *gfp*2 still had strong fluorescence intensity.

Plasmids:

Original plasmid: pIND is a high expression vector with a CMV promoter and enhancer but without a selective marker. Hence, a 1.1 kb fragment, which includes the 794 bp neo^r gene as well as the SV40 ploy(A) sequence, was inserted into the mammalian expression plasmid pIND. The novel recombinant plasmid is named pRN.

The primary hunting plasmid, pRGFP8, was constructed by inserting *gfp*2 gene and two other DNA fragments, termed F1(927bp) and F2 (1344bp) into pRN. The plasmid pRGFP8, 8.8kb, contains one recognition site for the restriction endonuclease *Spe* I for linearization (Fig1.).

The second targeting plasmid, pRIH-INF. Using the pRGFP8 as a backbone, pRIH-INF contains the gene of interest, human interferon-gamma (*ifn*-γ) gene (Gene bank accession number GI: 184638). A new selective gene, hyg^r, is linked behind the ifn-gamma through an internal ribosomal entry site (IRES) for reducing the background of the selection. The plasmid pRIH-IFN, size 10.56kb, contains only one recognition site of *Spe* I for linearization.

Cell lines:

The non-recombinant CHO K1 cell line, obtained from the American Type Culture Collection (ATCC number CCL-61), was derived as a subclone from the parental CHO cell line initiated from a biopsy of an ovary of an adult Chinese hamster. (8)

γ-CHO cell line was obtained from Dr. Walter Fiers. (9) This cell line was created from a dihydrofolate reductase deficient CHO cell line by cotransfecting the cells with *dhfr* gene and human *ifn-γ* gene.

Cell culture and transfection:

The γ-CHO cells were cultured in DMEM medium supplemented with 0.25μM methotrexate and 5% or 10% fetal bovine serum (FBS). (10)

The CHO K1 cells were cultured in DMEM medium containing 10%FBS in 6-well plates at the cell density of 0.3 million/well. After an overnight culture, the cell confluence reached about 60%. The transfection reagent was FuGene 6. By using 2μg *Spe* I linearized plasmids mixed with 6μl FuGene 6 in 100μl serum-free DMEM medium per well, the cells were incubated for 12 hours. Then the relative antibiotic, such as G418 or Hygromycin, was added for the selection of stable cell lines. Media changes performed every 24 hr. After the obvious cell death and 3 or 4 more days of growth, small colonies containing about 10 cells were apparent.

FACS sorting for GFP expressing cells:

The cells from the 6-well plates were washed with PBS, trypsinized and resuspended in serum-free DMEM medium. After centrifugation (10min, 200*g*), cells were dispersed in DMEM at the density of 10 millions cells/mL. The suspensions were transferred to FACS tubes and kept on ice until the test. After FACS sorting, the selected cells were collected in DMEM/10%FBS, and then transferred into 6-well plates with DMEM medium supplemented with 10%FBS and relative antibiotics: G418 or Hygromycin or both. The GFP fluorescence tests were also processed with a fluorescence microscope.

Human IFN test:

The concentrations of IFN-γ were determined using an ELISA kit (HyCult Biotechnology b.v., Uden, the Netherlands). (11)

Specific production rate, q

The specific production rate (q) was calculated using the following equation:

$$q=[\Delta C/t]\times[(\ln Nf-\ln Ni)/(Nf-Ni)]$$

Where ΔC is the change in component concentration over the time period t. N_f and N_i are the final and initial viable cell densities, respectively.

III. RESULTS AND DISCUSSION

A. Hunting for the high expression regions on the chromosome and setting a targeting locus

The transcriptional activity of transgenes in mammalian cells critically depends on the sites of their integration which is modulated by interactions between the promoter and surrounding chromatin structures. (7) To exploit the high expression regions on the chromosome, a designed linearizing vector (hunting plasimd, pRGFP8) with *gfp*2 gene and several extraneous DNA fragments is transfected into CHO cells. Theoretically, the transfected DNA can integrate in the chromosome randomly, and the integration site decides the *gfp*2 expression abilities. Therefore, the *gfp*2-transfected cells have different gene expression abilities due to the different integrating locations on the chromosome.

After exposure to G418, the transfected cells are selected with FACS. The cells with the strongest GFP expression are selected as the new host cells, which should indicate the *gfp* gene has inserted into the high expression regions on the chromosome. At the same time, these extraneous DNA fragments, F1 and F2 (nonendogeneous with host cell's) may integrate into the host cell's chromosome as well. These DNA fragments will help create a unique locus on the chromosome. During the retargeting process, the retargeting plasmid, pRIH-INF will easily locate to target these loci and this will increase the gene targeting frequencies.

Taking the *gfp* gene in the new host cells as a target locus, after transfecting pRIH-INF into the new host cells, the cells whose *gfp*2 gene is knocked out by the interesting gene based on the homologous recombination can be easily selected with FACS. Therefore, the selected stable cells should have strong abilities to express the protein of interest. On the whole, by using *gfp*2, the high expression sites on the chromosome of cells are marked, and the nonendogeneous DNA fragments are introduced to create the specific locus in order to increase the frequency of homologous recombination.

B. Strategy for the plasmids constructs

Homologous recombination in animal cells is now a routine technique that is used to modify the cells' genome at any chosen locus. (12, 13) Two basic types of targeting vectors, insertion and replacement, remain in current use.

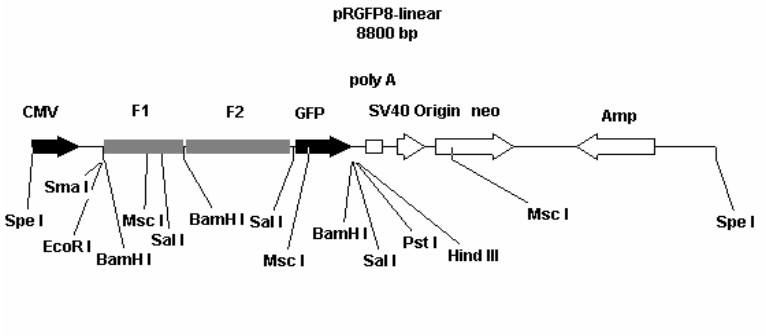

Fig.1. the linearlized plasmid pRGFP8. Gene-targeting vectors are routinely introduced in cells in a linearized form. Thus a unique restriction enzyme site must be present as a site for linearization. For a replacement gene-targeting vector, the linearization site must be outside of the regions of homology. To ensure the recombination vector has only one restriction enzyme site, we removed the *spe* I site of *gfp*, the new *gfp* gene is termed *gfp*2.

With both types of vectors, increasing the length of homology improves the targeting frequency (14,15). Successful targeting is difficult to achieve with a replacement vector when the total length of homology drops below 2 Kb, although 500 bp of homology on the short

arm has proved sufficient. An insertion vector is linearized within the region of homology and homologous recombination will lead to a duplication of genomic sequences. Thus, the replacement vector is more commonly used.

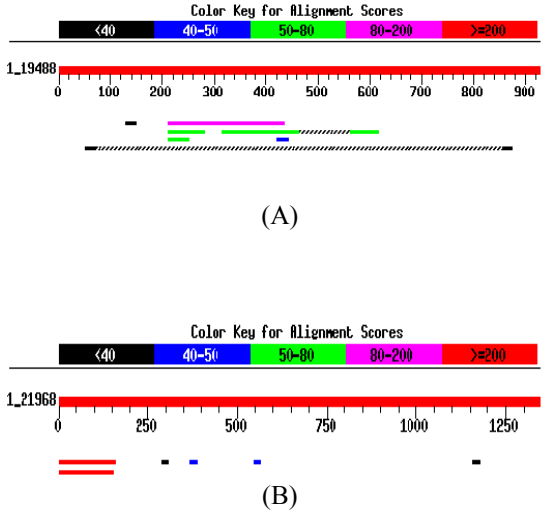

Fig. 2. BLAST tests for F1 (extraneous fragment 1) and F2 (extraneous fragment 2). (A), Extraneous fragment 1, Distribution of 9 Blast hits on the fragment 1, the largest fragment is 182bp, with 81% identities. (B), Extraneous fragment 2, Distribution of 8 Blast hits on the fragment 2, the largest fragment is 160bp, with 98% identities. Because the homologous recombination happens when the homologous sequence is at least 500bp, the F1 and F2 cannot influence the endogeneous chromosome of CHO cell.

The strategy for designing the hunting plasmid is to use special extraneous DNA to modify the host cell: not only to mark the high expression sites on the chromosome of the host cell, but also to set a unique locus for gene targeting and replacement. A replacement vector typically contains two separate arms of homology to the target locus (*gfp* gene), separated by a region of nonhomologous DNA. A nonreciprocal homologous recombination event between each arm of the vector and the target locus results in the replacement of part of the target locus with vector sequences. Thus, we introduced two extraneous DNA fragments in the short arm in order to increase the homology. These extraneous DNA fragments should have unique sequences that are different with the endogeneous DNA of host cell. In our study, we chose CHO as the host cell. Since there is not information about the entire CHO genomic sequence in the gene bank, we used mouse EST for the blast tests to select the suitable DNA fragments. Two extraneous fragments, termed F1 (927bp) and F2 (1344bp), were selected, and BLAST tests showed both of these fragments have low similarity after comparing them with the sequence of the entire GenBank Mouse EST (2,793,398 sequences; 1,254,021,342 total letters). (Fig.2)

C. gfp2 high expression CHO cell lines

After being transfected with pRGFP8, the cell population has become a cell library, which contains thousands, or even millions of cells with different *gfp*2 locations. To get a high-expression stable cell line, a two-round expression selection was performed. After G418 stable selection, a total of about 90 millions transfected cells were sorted in the first round on a Becton Dickinson FACScan flow cytometer and the top 1% of fluorescence cells were collected and continuously cultured in DMEM mediun/10%FBS containing G418 for two weeks. The same FACS sorting was carried out again and the top 1% high-fluorescent population was collected. The average fluorescence intensity of this population was 207.21, and comparied to the original *gfp*-transfected cell population, the fluorescence had increased about 38 times.

After being sorted with FACS, the selected cells were then picked up for single colony selection: each cell line was culture about 30-50 generations, and their offspring were tested on the fluorescence performance. Those cells that were able to express the inserted gene stably were eligible to be used as potential new host cells.

D. Targeting the pre-setting loci

All of the *gfp*2 stable high-expression CHO cells selected are potential new host cells and gene targeting should be performed in order to find the cell whose *gfp*2 gene can be replaced easily. We chose the 8 stable GFP CHO cell lines as the targeting cells. After 50 generations' culture, all of these cell lines show at least 90% positive fluorescence cells (from 90.87% to 95.91%) in a 10,000-events FACS sorting test.

We chose the human IFN-gamma gene as a model gene to replace the *gfp* in the GFP CHO cells. Upon the DNA transfected into GFP CHO cells, random integration of the plasmid can lead to hygromycin resistance, while homologous recombination with the endogenous *gfp* gene causes both hygromycin resistance and GFP negative. After transfection with linearized pRIH-INF and sorted with FACS, the dimmest 10% cells were collected and then cultured in DMEM medium supplemented with 10% FBS and G418 and hygromycin. In the second round of FACS sorting, the dimmest 1% of cells were collected and cultured in the same conditions and then single cell colonies were picked up to test their GFP and IFN-gamma productions.

After gene targeting, five of the eight cell lines were found to not be GFP$^-$-hygromycin^r-IFN$^+$. One of reasons for this is that they may contain a few *gfp* gene copies in different positions and these *gfp*2 cannot be all knocked-out at the same time. The cells found that meet the above condition were A12, A16 and B4.

Compared to most of the other usages of gene targeting, nonhomologous recombination does not have significant negative influence on our expression system. In fact, the interesting gene inserting randomly into chromosome may be good for increasing the production of protein of interest in most of cases.

When we isolated the *gfp*2 knock-out cells, these cells were cultured in DMEM medium and their specific IFN-gamma productivities were determined. The results show most of the cell lines have a high expression level of IFN-gamma compared to the control γ-CHO cells. Interestingly, there is a positive correlation between the fluorescence densities of GFP CHO cells and the expression levels of IFN-gamma. (Fig.3.)

On average, the productivity of newly targeted cells is predictable based on the fluorescence density of the original GFP cell lines. However, when individual clones are compared, this correlation does not work well because of the drastic differences in the expression level of individual clones. Nonhomologous recombination may be one of important reasons for the variance of the productivities among the individual clones. If we neglect the possibility that the nonhomologous recombination may cause fatal mutants, nonhomologous recombination is good for increasing gene expression.

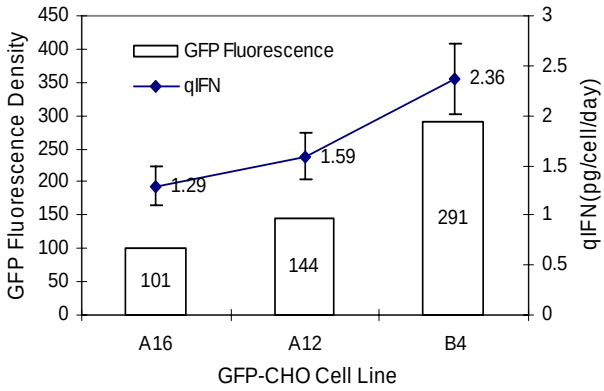

Fig. 3. The correlation between the fluorescence densities of the origianal GFP-CHO cell lines and the interferon-gamma production abilities of the *gfp*2 knock-out cell lines. The qIFN data were determinate at the day 3's culture in the DMEM/5%FBS medium. The fluorescence density of the GFP-CHO cell lines were determined before they were transfected with pRIH-INF. The qIFN of γCHO was 1.12pg/cell/d.

E. Influence of two extrous DNA fragments on the gene targeting frequency

For increasing the gene targeting frequency, we introduced two extraneous DNA fragments, F1 and F2 in the vectors, which would have integrated into chromosome to create the unique locus. For the cell lines A16 and B4, the gene targeting frequency was 1.2×10^{-5} and 2.0×10^{-5} respectively. Systematic studies of the targeting ratio at the APRT locus in CHO cells as a function of transfection method have shown that the targeting ratio varies from 1:15 for microinjection to 1:370,000 for Fugene-6. (16-18). The transfection method can influence the targeting frequency dramatically. As we used FuGene-6 as the transfection reagent, the gene targeting frequency reached to the similar level shown above, this means the extraneous DNA fragments were helpful. Moreover, the same targeting tests for A16 cells were performed with the plasmid without F1 and F2 fragments (data not shown). Without F1 and F2 fragments, the homology of the short arm of the vector decreases from 3.2 Kb to 0.9Kb, and this caused at least a 100-fold decrease in the targeting frequency.

This result supports that there is just one *gfp*2 gene copy in both the A16 and B4 cell lines. However, for the transfected vector, whether it is initially circular or linear, the ends may become truncated during the integration process.(14,15) Thus, the pre-set locus in these two different cells may not be same , neither are the total homology between the targeting locus and the linearized pRIH-INF. This makes the targeting frequency for these cell lines, A16 and B4, different.

The targeting frequency for A12 was lower, about 1×10^{-7}. We consider that there may be at least two *gfp*2 gene copies in A12. As for the other 5 cell lines that were not found the purpose cells after gene targeting, they may contain even more than two *gfp*2 gene copies. Based on statistical estimation, we may obtain only one cell where all the copies of the *gfp*2 gene are knocked-out for every 2 billion transfected cells which contain 3 copies of expressed-*gfp*2 gene.

The system described here is a good method for high expression of the interesting gene. As opposed to the common targeting method, nonhomology recombination is not a issue anymore. With the extraneous DNA fragments F1 and F2, the pre-set loci are easily targeted. The positive correlation between fluorescence density of the GFP-CHO cells and protein productivity of the purpose cells makes the targeting results predicated. However, considering that the genomic DNA of CHO cells may be more than 1 billion base pairs and any base pair difference in integration may cause the difference of gene expression level, more work should be done in hunting for the high expression regions. A much larger transfected cell number for sorting may be a good solution in obtaining the cell line which contains just one transgene copy and has a stable and high gene expression level.

ACKNOWLEDGMENT

We are thankful for Dr. Liyun Liu in Hastings' lab, Department of biology, Harvard Universty, for his generous gift of genes. We also thank Mr. Glenn Paradis, flow cytometry group, cancer center, MIT, for his technical assistance.

REFERENCES

[1] V. A. Vinci and S.R. Parekh, *Handbook of Industrial Cell Culture.*Totowa, NJ: Humana Press, 2003, pp.21-42.

[2] H. Hauser and R. Wagner, *Mammalian Cell Biotechnology in Protein Production.* DeGruyter, New York, 1999, pp.1-190.

[3] R.J. Kaufman, L.C. Wasley, A.J. Spiliotes, S.D. Gossels, S.A. Latt, G.R. Larsen, et al. " Coamplification and coexpression of human tissue-type plasminogen activator and murine dihydrofolate reductase sequence in Chinese hamster ovary cells," Mol. Cell. Biol., vol. 5, pp1750-1759, 1985.

[4] J.E. Looney and J.L. Hamlin, " Isolation of the amplified dihydrofolate reductase domain from methotrexate-resistant Chinese hamster ovary cells," Mol. Cell. Biol., vol. 7, pp569-577, 1987.

[5] N.J. Kilby, M.R. Snaith, and J.A.H. Murray, " Site-specific recombinases: tools fro genome engineering," Trends Genet., vol. 9, pp413-421, 1993.

[6] D. Schubeler, K. Maass, and J. Bode. " Retargeting of retroviral integration sites for the predictable expression of transgenes and the analysis of cis-acting sequences, " Biochemistry, vol 37, pp. 11,907-11914, 1998.

[7] E. Verhoeyen,H.Hauser, and D. Wirth, " Evaluation of retrovial vector design in defined chromosomal loci by Flp-mediated cassette replacement," Hum. Gene Ther., vol 12, pp933-944, 2001.

[8] T. T. Puck, S. J. Cieciura, A. Pobinson, "Genetics of somatic mammalian cells III. Long-term cultivation of euploid cells from human and animal subjects," J. Exp. Med. , vol 6 pp945-956, 1958.

[9] S. J. Scahill, R. Devos, J. Van der Heyden and W. Fiers, " Expression and characterization of the product of a human immune interferon cDNA gene in Chinese hamster ovary cells", Proc. Natl. Acad. Sci. USA, vol 15, pp4654-4658, 1983.

[10] I. H. Y. Yuk, "Protein expression and glycosylation in CHO cells," Ph.D. dissertation, Dept. Chem. Eng., MIT, Cambridge, MA, 2001.

[11] S. R. Fox, "Active hypothermic growth: a novel means for increasing total recombinant protein production by CHO

cells," Ph.D. dissertation, Dept. Chem. Eng., MIT, Cambridge, MA, 2003.

[12] U. Muller, "Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis," Mechanisms Develop., vol 82, pp3-21, 1999.

[13] P. Hasty, J. R. Perez, C. Chang and A. Bradley, " Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells," Mol. Cell. Biol., vol11, pp4509-4517, 1991.

[14] P. Hasty, J. R. Perez, C. Chang and A. Bradley, " The length of homology required for gene targeting in embryonic stem cells," Mol. Cell. Biol., vol11, pp5586-5591, 1991.

[15] K.M. Vasquez, K, Marburger, Z. Intody and J.H. Wilson, "Manipulating the mammalian genome by homologous recombination," PNAS, vol 98, pp8403-8410, 2001.

[16] B.C Waldman, J.R O'Quinn, and A. S. Waldman, "Enrichment for gene targeting in mammalian cells by inhibition of poly(ADP-ribosylation)," Biochim Biophys Acta., vol 3, pp241-250, 1996.

[17] R. S. Nairn, G. M. Adair, T. Porter, S. L. Pennington, D. G. Smith, J. H. Wilson JH, M. M. Seidman, "Targeting vector configuration and method of gene transfer influence targeted correction of the APRT gene in Chinese hamster ovary cells", Somat. Cell. Mol. Genet., vol4, pp363-375, 1993.

[18] R. J. Yanez, and A. C. Porter. "Influence of DNA delivery method on gene targeting frequencies in human cells", Somat. Cell. Mol. Genet., vol 1, pp27-31, 1999.