

The Engineering of Chinese Hamster Ovary Cells to Achieve More Efficient Gene
Amplification for Improving Biopharmaceutical Development

Jonathan Cacciatore

Submitted in Partial Fulfillment of the
Requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2012

© 2012
Jonathan Cacciatore
All rights reserved

ABSTRACT

The Engineering of Chinese Hamster Ovary Cells to Achieve More Efficient Gene Amplification for Improving Biopharmaceutical Development

Jonathan Cacciatore

This dissertation addresses the issue of the long development times of obtaining Chinese hamster ovary (CHO) cells capable of producing high quantities of therapeutic proteins. It addresses the specific time bottlenecks associated with developing high producing CHO cell lines and reviews various methods that are employed to alleviate these bottlenecks. The specific time consuming process of gene amplification is the focus of this work. Gene amplification is the process of selecting CHO cells which have been genetically modified to contain many copies of a therapeutic transgene, and therefore has the ability to produce a high amount of therapeutic protein. Two separate projects are described which decrease the time necessary to obtain a high producing cell. The first project describes a novel process developed which can measure and quantify the amplification rate of a transgene in CHO cells. This process was used to successfully isolate a CHO cell clone with the capability of amplifying a transgene targeted to a specific location in the genome and thus produce higher quantities of protein in a shorter time period. Site-specific recombination (SSR) technology was utilized to target the transgene to this location which was deemed capable of amplifying a transgene at a high rate. The second project also utilizes SSR technology to integrate many copies of a transgene into many recombination sites in the CHO genome. A cell line containing several thousand integration sites was isolated, however only about twenty of these sites successfully integrated a transgene after optimizing cell transfection conditions. Efforts towards engineering an improved recombinase for this purpose has led to the result that

DNA sequences flanking recombination sites have the ability to greatly improve this integration process. Potential future experiments are described which may isolate such sequences and ultimately increase the number of transgenes integrated into the CHO cell genome. Overall, these improvements to CHO cells have the ability to ultimately isolate a higher producing cell line faster, thus decreasing the time to get a potential drug candidate to market.

Table of Contents

Chapter 1 – Introduction.....	1-9
Chapter 2 - Gene Amplification and Vector Engineering to Achieve Rapid and High-Level Therapeutic Protein Production using the Dhfr-based CHO Cell Selection System	10-49
Chapter 3 - Recombinant protein production in CHO cells from a site selected for a high amplification rate.....	50-78
Chapter 4 – Fluctuation Analysis Theory.....	79-96
Chapter 5 - Achieving immediate high transgene copy numbers using pre-amplified site-specific recombination sites.....	97-132
Chapter 6 – Conclusions.....	133-137

List of Charts, Graphs, and Illustrations

Figure 2-1 - Laser-enabled analysis and processing of IgG secreting cells.....	23
Figure 2-2 - Enhanced amplification of <i>dhfr</i> using leaky splicing.....	27
Table 2-1 - Summary of cell and vector development methods.....	40
Figure 3-1 - DNA for measuring <i>ada</i> amplification rates	60
Figure 3-2 - Amplification rates of transfectant clones	62
Figure 3-3 - Site specific recombination scheme	64
Figure 3-4 - Gene amplification of transgenes in recombinant clones	65
Figure 3-5 - Recombined <i>dhfr</i> becomes amplified in clone 4.....	66
Figure 3-6 - SEAP secretion during gene amplification	68
Table 3-1 - Calculated amplification rates using the Po, mean and median methods	73-75
Figure 4-1 - Schematic of CHO gene amplification process.....	83
Figure 4-2 - Comparison of L&D mutation study vs. CHO gene amplification study	87
Table 4-1 - Raw colony count data for clone 4 used to calculate amplification rate...	93
Figure 5-1 - Schematic of cre/lox recombination system.....	101
Figure 5-2 - Schematic of ϕ C31 integrase system.....	102
Figure 5-3 - Schematic of amplified attP genomic DNA.....	104
Figure 5-4 - Method for creating attP concatemers.....	108
Figure 5-5 - Stained transfectants with and without ϕ C31 integrase vector.....	115
Figure 5-6 - Endpoint PCR of Genomic DNA from Transfectant Pools.....	116
Figure 5-7 - Relative <i>dhfr</i> copy number measured using real time PCR.....	117

Table 5-1 - *Dhfr* copy number across different transfection conditions.....118

Figure 5-8 - Relative *dhfr* copy number using calcium phosphate DNA co-precipitation
.....119

Figure 5-9 - Schematic of replicating plasmid experiments121

Figure 5-10 - Schematic of ϕ C31 integrase/zinc finger chimera.....123

Figure 5-11 - *In vitro* recombination assay.....124

Figure 5-12 - *In vitro* compartmentalization (IVC) process for isolating efficient ϕ C31
enzymes.....127

Acknowledgments

I would like to thank several people for their support during the writing of this thesis. I would like to thank my parents for supporting my education and their emphasis on the importance of hard work. I would like to thank my sister, grandparents, and uncles for their support throughout these years. I would like to thank my girlfriend Nikki, a fellow graduate student, for her input and advice. I also thank her family for their support over the last few years.

A special thanks to Mauricio Arias, Shengdong Ke, Vincent Anquetil, Dennis Weiss, Laurens Moore van Tienen, and Mrinalini Gururaj for their scientific advice and assistance throughout this project.

Chapter 1

Introduction

Biological molecules have emerged over the past decade as effective treatments for various conditions. This has led to biotechnology and pharmaceutical companies shifting more efforts toward the research and development of such molecules as opposed to synthetic small molecule based drugs. The demand for biologics has increased rapidly over the past decade and has nearly doubled in the last five years in terms of worldwide sales. Although worldwide sales of both types of drugs have increased, the increase in demand for biologics is much larger relative to total worldwide sales and this trend is expected to continue (Lehman, 2007). The majority of these molecules are recombinant proteins such as antibodies, hormones and enzymes with very specific cellular functions.

A subdivision of these biologics whose demand is increasing at a rapid pace constitutes monoclonal antibodies (Mab's). Mab's are currently the best selling and highest growing class of biologics with a market of about 18.5 billion dollars and a growth rate of about 10% as of 2010 (Aggarwal, 2011). Mab's are engineered biomolecules which specifically recognize and bind antigens or receptors in/on various cells (mimicking antibodies naturally produced in the body) to perform specific functions in treating varying conditions. This class of molecules is particularly effective in treating various types of cancer. Mab's have been engineered to bind ligands on certain cancerous cells making them more visible to the immune system for their destruction. Mab's have the ability to trigger both complement-dependent cytotoxicity (CDC) and antibody-dependent cellular cytotoxicity (ADCC) upon binding the surface of a cancer cell. Other proteins bind the membrane bound Mab which recruits the membrane attack complex (MAC) for cellular destruction. Receptors on immune effector cells may bind the F_c

region of the bound Mab which triggers destruction of the cancer cell by that particular effector (Sawada et al., 2011). The binding of Mab's may also destroy cancer cells using other mechanisms upon binding antigens associated with tumors. For example, the drug rituximab (Rituxan) is engineered to specifically bind the protein CD20 on the surface of B cells. This allows the immune system to destroy cancerous B cells associated with certain lymphomas more effectively (Wood, 2001). Mab's have also been engineered to bind to receptors on cancer cells that block certain growth signals. This inhibition prevents cancer cells from dividing. One drug currently on the market that acts in this manner is cetuximab (Erbix) which binds receptors on cancerous cells that block the binding and action of epidermal growth factor (Mendelsohn and Baselga, 2000). This drug is currently used to treat colon and head and neck cancer. Another mode of action of Mab's is inhibiting the formation of blood vessels which provide nutrients to cancerous cells. The drug bevacizumab (Avastin) binds to vascular endothelial growth factor (VEGF) which inhibits a cancer cell's ability to initiate the formation of new blood vessels (Ranieri et al., 2006). Mab's may also be combined with other molecules for the direct targeting of those molecules to cancer cells through specific binding of the Mab to tumor antigens. Mab's have been linked to radioactive particles as a means of delivering radiation more specifically to cancer cells. Ibritumomab (Zevalin) is used to treat non-Hodgkin's lymphoma by specifically binding receptors on cancerous blood cells and delivering low levels of radiation (Lin and Iagaru, 2010). Another example of combining Mab's with other molecules for treatment is the process of antibody-directed enzyme prodrug therapy (ADEPT). This method links an enzyme to a Mab which specifically binds cancer cells. After this occurs a prodrug is administered to the patient that will

only be converted to its active form by the enzyme on the cancer cells bound by the Mab (Syrigos and Epenetos, 1999). This is an effective way of targeting another drug to cancer cells through the specificity of Mab binding.

Initially, murine Mab's were isolated using hybridoma technology which involves fusing an antibody producing cell with a myeloma cell. These Mab's often caused problems because the human immune system would reject the molecule since it was produced in a mouse (Stern and Herrmann, 2005). This has led to the development of chimeric antibodies which contain murine variable regions (for antigen binding) connected to human constant regions. This has significantly reduced immunogenicity in humans. Humanized Mab's have also been developed by substituting portions of the murine variable domain onto a fully human antibody (Presta et al., 1993). Fully human antibodies have also been developed by introducing human immunoglobulin genes into transgenic mice for their production after vaccination of the mice (Hudson and Souriau, 2003). Each of these types of antibodies is represented in the current portfolio of Mab's on the market.

The co-emergence of recombinant DNA technology has allowed for development of a successful platform for the manufacture of Mab's. Using recombinant DNA technology allows for the modification of immunoglobulin genes to discover and test large numbers of Mab's for their efficacy in binding certain targets associated with many disease types. The current platform for producing these molecules involves the use of host cells transfected with Mab genes for expression, translation, and secretion of the molecule.

The complexity of Mab's requires that they be produced in mammalian cells as opposed to bacterial cell lines. Mammalian lines allow proper folding and glycosylation, necessary to the activity of these proteins. This current platform acts as a time bottleneck when considering the number of Mab candidates that need to be produced and tested for pre-clinical study. Furthermore, unlike other recombinant proteins previously manufactured such as erythropoietin and tissue plasminogen activator (tPA), therapeutic Mab's are needed in relatively large amounts per dose which makes meeting the ever increasing demand for these molecules even more challenging. Although there has been considerable work done to improve processes for the development and manufacture of such molecules, the pace of increasing demand seems to be surpassing current methods.

The ever increasing demand for Mab's has made it necessary to rapidly produce many candidates for testing in order to quickly get these drugs to market. Even for preclinical studies, significant amounts of the drug candidate are needed. The process of transfecting Mab genes into host cells and isolating productive transfectants for producing these molecules is a major time bottleneck. Since these molecules are produced as transfected transgenes into a host cell, there is a limitation on production based on how well the Mab gene is expressed, translated, and ultimately secreted into the cell culture medium. Each of these steps may act as a bottleneck to producing more Mab; however the main focus of this thesis is to improve the amount of Mab gene expression. There have already been numerous studies which have improved expression of the Mab gene by modifying the expression vector with enhancing sequences, insulators, and superior promoters. Since these vector sequences are already somewhat established, our

work focuses on achieving a cell line with many copies of this transgene in a shorter period of time; thus increasing expression and production of the Mab per cell. The current protocol for achieving high copy numbers of a Mab transgene in the host cell is a very time consuming process and takes about 6 months. The goal of this thesis is to develop a method for shortening this process, thus ultimately decreasing the time in which a Mab candidate can reach the market.

In Chapter 2 the current production method used to produce therapeutic proteins such as Mab's in the Chinese hamster ovary (CHO) cell is discussed. This cell line is the most common host for various reasons described in this chapter. The various methods of transgene transfection and establishment of cell lines with many copies of that transgene through a process known as gene amplification are also discussed in detail. Various ways for making the production process more efficient are reviewed in detail including improved selection of high producing clones, modified gene amplification protocols, modifying the expression vector for improved expression, and utilizing site-specific recombination (SSR) technology to target the transgene to favorable genomic locations.

Chapter 3 focuses on alleviating the time bottleneck associated with isolating CHO cells with many copies of the Mab transgene through the gene amplification process. Gene amplification refers to the random duplication of a portion of the CHO genome during cell division. Isolation of a cell that has undergone this random process many times at the locus containing the Mab gene is possible using a specific selection process described in more detail in Chapters 2, 3 and 4. In Chapter 3 we develop a method for isolating

clones of CHO cells that have the ability to amplify DNA at a specific genomic location at a high rate, thus decreasing the time needed to isolate a high Mab gene copy number clone. Our method utilizes fluctuation analysis developed by Luria and Delbruck to measure amplification rates in 100 CHO clones. Once high rate clones are identified and isolated, we target a transgene for the assayable protein secreted embryonic alkaline phosphatase (SEAP) to those locations deemed highly amplifiable using SSR technology. We specifically use the ϕ C31 recombination system which is explained in detail in Chapter 2. We confirm that using our most highly amplifiable clone, we can improve the reliability of achieving a high copy number cell line.

Chapter 4 focuses on the theory associated with the analysis employed in Chapter 3. The derivation of the equations used for determining gene amplification rates are explained in detail. Parallels are made between Luria and Delbruck's use of these methods in determining bacterial mutation rates and our adaptation to calculate gene amplification rates in CHO cells.

Chapter 5 also focuses on alleviating the time bottleneck associated with isolating CHO cells with many copies of the Mab transgene. Unlike the work described in Chapter 3, we aim to develop a method for integrating many copies of a transgene into a host CHO cell without the need for a lengthy gene amplification protocol. Whereas in Chapter 3 we discuss our successful isolation of clone with the ability to undergo the amplification process faster, in Chapter 5 we describe our work towards a single experimental protocol utilizing site-specific recombination technology to quickly isolate a high copy number

cell line. We used two different SSR methods to achieve this goal; the cre/lox system and the aforementioned ϕ C31 recombination system. We describe various ways to improve the integration of transgenes into many host genomic locations since SSR is inherently inefficient. These methods include increasing the number of sites for SSR, alternative DNA transfection techniques, increasing incoming DNA concentration using a viral replication scheme, and improving the efficacy of the ϕ C31 enzyme to promote more efficient DNA recombination into a large number of genomic locations.

Chapter 6 details our overall conclusions from two separate yet connected efforts; both of which aim to decrease the development time of a CHO cell containing many copies of a transgene leading to a high protein production per cell. The application is specifically geared toward the quick production of Mab candidates for preclinical and clinical study for the treatment of cancer and other diseases. We summarize our main conclusion for Chapter 3 which is the isolation of a CHO clone capable of fast transgene amplification and the development of a method which can be employed to isolate a fast amplifying cell line of a different lineage. We also draw conclusions based on the results described in Chapter 5 which implies that modifying the ϕ C31 enzyme and the DNA recognition sequences it binds may be necessary to develop a way which completely abrogates the need for lengthy gene amplification. We describe future experiments which can be conducted to this extent. Finally we describe the overall application to the current methods employed by the pharmaceutical industry to isolate high producing cell lines and how our work improves upon those methods.

References

- Aggarwal, S., 2011. What's fueling the biotech engine--2010 to 2011. *Nature biotechnology*. 29, 1083-1089.
- Hudson, P.J., Souriau, C., 2003. Engineered antibodies. *Nature medicine*. 9, 129-134.
- Lehman, (2007) *Biologic Drugs and Vaccine Sales to 2011E*. Bio IT World. Lehman Brothers.
- Lin, F.I., Iagaru, A., 2010. Current concepts and future directions in radioimmunotherapy. *Current drug discovery technologies*. 7, 253-262.
- Mendelsohn, J., Baselga, J., 2000. The EGF receptor family as targets for cancer therapy. *Oncogene*. 19, 6550-6565.
- Presta, L.G., Lahr, S.J., Shields, R.L., Porter, J.P., Gorman, C.M., Fendly, B.M., Jardieu, P.M., 1993. Humanization of an antibody directed against IgE. *Journal of immunology*. 151, 2623-2632.
- Ranieri, G., Patruno, R., Ruggieri, E., Montemurro, S., Valerio, P., Ribatti, D., 2006. Vascular endothelial growth factor (VEGF) as a target of bevacizumab in cancer: from the biology to the clinic. *Current medicinal chemistry*. 13, 1845-1857.
- Sawada, R., Sun, S.M., Wu, X., Hong, F., Ragupathi, G., Livingston, P.O., Scholz, W.W., 2011. Human monoclonal antibodies to sialyl-Lewis (CA19.9) with potent CDC, ADCC, and antitumor activity. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 17, 1024-1032.
- Stern, M., Herrmann, R., 2005. Overview of monoclonal antibodies in cancer therapy: present and promise. *Critical reviews in oncology/hematology*. 54, 11-29.
- Syrigos, K.N., Epenetos, A.A., 1999. Antibody directed enzyme prodrug therapy (ADEPT): a review of the experimental and clinical considerations. *Anticancer research*. 19, 605-613.
- Wood, A.M., 2001. Rituximab: an innovative therapy for non-Hodgkin's lymphoma. *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists*. 58, 215-229; quiz 230-212.

Chapter 2

Gene Amplification and Vector Engineering to Achieve Rapid and High-Level Therapeutic Protein Production using the Dhfr-based CHO Cell Selection System

(Published in *Biotechnology Advances*, 2010, 28:673-681)

Jonathan J. Cacciatore^a, Lawrence A. Chasin^{b,*}, Edward F. Leonard^a

^a Department of Chemical Engineering, Columbia University, New York, NY 10027
United States

^b Department of Biological Sciences, Columbia University, New York, NY 10027 United
States

Corresponding Author: * 912 Fairchild Center, Mail Code 2433, Columbia University,
New York, NY 10027, United States Tel.:+ 1 212 854 4645
Email: lac2@columbia.edu (Lawrence Chasin)

Jonathan J. Cacciatore: Performed literature search and writing

Lawrence A. Chasin: Assisted in writing and editing, created figures and legends

Edward F. Leonard: Assisted in writing and editing

Abstract

Demand is increasing for therapeutic biopharmaceuticals such as monoclonal antibodies. Achieving maximum production of these recombinant proteins under developmental time constraints has been a recent focus of study. The majority of these drugs are currently produced in altered Chinese hamster ovary (CHO) cells due to the high viability and the high densities achieved by these cells in suspension cultures. However, shortening the process of developing and isolating high-producing cell lines remains a challenge. This article focuses on current expression systems used to produce biopharmaceuticals in CHO cells and current methods being investigated to produce biopharmaceuticals more efficiently. The methods discussed include modified gene amplification methods, modifying vectors to improve expression of the therapeutic gene and improving the method of selecting for high producing cells. Recent developments that use gene targeting as a method for increasing production are discussed.

Keywords: Biopharmaceutical, amplification, methotrexate, expression, enhancers

Introduction

The high specificity of protein-based drugs such as monoclonal antibodies (MAbs) has made them particularly effective for the treatment of various diseases including several types of cancer. This realization has shifted the efforts of pharmaceutical companies toward the manufacture of more therapeutics that are protein based (Pearson, 2007). The production of such therapeutics requires the use of mammalian cell lines and not yeast or bacteria, since mammalian cells allow the folding and glycosylation necessary for the activity of these drugs. A majority of these proteins are currently produced using Chinese hamster ovary (CHO) cells, for several reasons. First, this cell line, similar to other mammalian cells used for production, has the ability to modify its protein products with glycosylation patterns similar to those in humans (Jayapal et al., 2007). Second, these cells have the ability to grow to high densities in bioreactors. Third, they pose a low risk for the transmission of human viruses (Wiebe, 1989). Fourth, this cell line is known to have a very unstable genome relative to other candidate cell lines considered for production, making them good candidates for gene amplification and other genetic manipulations. Fifth, CHO cells, similar to other cell candidates, can be easily transfected with gene vectors that are expressible to produce therapeutic proteins of choice. Finally, industrial CHO production processes involving gene transfection and selection are well characterized. Contravening these many advantages, the quick development of cell lines which maximize protein production remains a challenge. Other cells currently being used for biopharmaceutical production include NS0 (mouse myeloma-derived), BHK, HEK-293, and PER-C6 (human retina-

derived) (Wurm, 2004). CHO cells are currently the most widely used cells for production and they will be the main focus of this review.

Current production platform

Transfection

A majority of platforms utilize mutant strains of CHO cells that have both alleles of the dihydrofolate reductase (*dhfr*) gene mutated (Urlaub and Chasin, 1980) or deleted (Urlaub et al., 1983). DHFR catalyzes the conversion of folic acid to tetrahydrofolate and is necessary for biosynthetic pathways that produce glycine, purines, and thymidylc acid. These auxotrophic cell lines require growth medium containing glycine, a source of purines such as hypoxanthine, and thymidine (GHT) for their survival. A typical scheme for production of a protein-based drug begins with co-transfection of the gene for the protein of choice and a *dhfr* minigene. The gene for the protein of choice and the *dhfr* minigene may be present on the same vector or two separate vectors as both are integrated at the same location regardless (Chen and Chasin, 1998). DHFR is used as a selection marker for cells that have successfully integrated the vector. Thus, CHO cells that have successfully integrated the expression vector for the drug of interest will have also integrated the *dhfr* gene, and will survive in medium deficient in GHT. The resulting pool of cells will have the gene integrated into their genomes at various locations, where each may exhibit different levels of expression as determined by the surrounding sequence into which the vector has been randomly integrated. Thus, the

ability to isolate mutants lacking DHFR allows easy selection of stable clones. This selection of surviving clones has no effect on productivity of the gene of choice since only a low level of *dhfr* expression is required for survival.

An alternative to generating stable cell lines for recombinant protein production is to perform transient transfection. Transient expression of the therapeutic protein gene is capable of producing up to 1 g/L of product within a span of about a week (Backliwal et al., 2008a; Morrow, 2008). Although this is lower than the maximum yields achievable with stably transfected cell lines, the short time required to harvest protein makes this an attractive option (Wurm, 2004). Advantages of transient transfection include simpler expression vectors since no selective marker is used, a short time frame for production, applicability to a wide range of cell lines, and genetic stability of cells due to the short production time frame (Wurm and Bernard, 1999). In many scenarios these advantages can outweigh the lower yields of transient transfection compared to permanent transfection and the repeated requirement for producing large amounts of plasmid DNA for each run.

Permanent cell lines have the advantage of utilizing the highest expressing clones from a pooled population, allowing optimization of culture conditions for those particular clones to provide the highest yield, and the availability of a standard workhorse for repeated use. This process may take several months, but it continues to be the most commonly used industrial process.

MTX amplification

At this point in the development phase, a method is often used to obtain cell lines that have a high copy number of the integrated transgene due to gene amplification, a low frequency event which occurs during cell division in CHO cells. Among the isolated stable clones are a small number that have a high copy number of the integrated transgene. A step frequently used to increase expression of a transfected gene exploits *dhfr* gene amplification. Transfected cells carrying both the *dhfr* minigene and a gene for a protein of interest are exposed to a low concentration of methotrexate (MTX). Methotrexate is a specific inhibitor of DHFR and cells that acquire resistance are usually found to have undergone *dhfr* gene amplification at the site of the integrated *dhfr* vector (Nunberg, 1978). The result will be cells with an increased copy number for both *dhfr* and the protein of interest, since these are usually integrated together even when introduced on separate plasmids (Chen and Chasin, 1998). The majority of cells, which have not undergone amplification, will be unable to survive the inhibitory action of MTX. This procedure is repeated using gradually increasing concentrations of MTX in order to increase the selection pressure until only cells with very high gene copy numbers prevail. The result is a pool of cells with varying integration sites and increased gene copy numbers. MTX selection of amplifiants is a major bottleneck in the production of biopharmaceuticals since many rounds are required to obtain cells with high gene copy numbers, a process that typically takes several months. Moreover, increasing MTX concentrations are only capable of selecting high copy number clones to a certain extent. At a certain concentration, higher MTX concentrations fail to yield clones with

increasing copy number. This time bottleneck prevents quick production of a protein product for pre-clinical evaluation. The saturation effect of using MTX to select for amplified cell lines also limits the extent to which high producers can be isolated.

GS amplification

A similar amplification protocol exists using glutamine synthetase (GS) mediated gene amplification. GS-mediated amplification utilizes the GS enzyme, which catalyzes the synthesis of glutamine from glutamate and ammonia (Bebbington et al., 1992). This system typically requires the use of a cell line other than CHO, specifically the NS0 murine myeloma cell type. Analogous to the *dhfr*-deficient CHO cells, NS0 cells are auxotrophic for glutamine and can be selected for in glutamine free medium (Jun et al., 2006). Utilizing this system in CHO cells requires the concurrent addition of methionine sulfoximine (MSX), a specific inhibitor of GS at a low level (20 μ M) for the selection of transfectants since CHO cells possess endogenous glutamine synthetase (Brown et al., 1992). Through the use of MSX during the selection of transfectants, a mutant cell line is not required for transfection unlike the *dhfr* selection system. Another advantage of this system is that isolated transfectants using the GS system have higher production levels before gene amplification when compared to transfectants using the *dhfr* system (Cockett et al., 1990). Despite these advantages over the *dhfr* system, there is data to suggest that production is unstable using GS-mediated amplification. High-producing subclones of recombinant CHO cells producing humanized antibody isolated at various MSX concentrations showed a significant decrease in production over the first six passages

(Jun et al., 2006). Moreover, cytogenetic studies revealed unstable chromosome structure that may contribute to production instability (Jun et al., 2006). Due to these drawbacks, the *dhfr* system continues to be the most popular within the pharmaceutical industry despite its lengthy gene amplification regimen. For this reason, the *dhfr* system will be the focus of the methods reviewed here.

Clone selection

A second bottleneck in the isolation of a high-producing cell line is the selection of the clones from an amplified pool of cells that exhibit the best productivity and growth rate, properties that depend upon but are not wholly determined by copy number. Standard methods involve isolating individual colonies through limiting dilution (Puck and Marcus, 1955) or cloning cylinders (Davis, 2002). Each clone is then evaluated for both its growth rate and its rate of productivity of the protein of interest. Both processes are very time consuming and thus hinder the development of new biopharmaceuticals.

Two different strategies are used for the selection of high-producing clones. The first involves isolating individual clones from the first level of MTX selection, subjecting each clone to a higher level of MTX, again isolating single clones, and so on for each level of selection. Subcloning can also be carried out to optimize homogeneity (Kim et al., 1998). The second strategy involves pooling clones at each stage of the MTX selection and isolating individual clones only from the final MTX resistant pool. A study was performed to compare the efficacy of these two strategies. Jun *et. al.* compared the antibody productivity of 30 parental clones to 10 parental cell pools after subjecting both

to the MTX amplification procedure. High-producing clones were isolated from the cell pools within 15 weeks at an antibody titer of approximately 5 $\mu\text{g/ml}$. High producers were isolated from the 30 parental clones in about 17 weeks with the highest subclone achieving a titer of approximately 17 $\mu\text{g/ml}$ (Jun et al., 2005). The individual clone strategy proved to be labor intensive and more time consuming due to an extra cloning step. This protocol was not improved by including MTX in the initial selection of transfectants. The cell pool strategy was less labor intensive; however the highest producer was about one third as productive as the highest producer isolated using the individual clone strategy. The disparity could be due to high-producing clones within each pool displaying lower growth rates and being overtaken by lower producing clones (Imanaka and Aiba, 1981). Despite being more labor intensive, the method of selection and amplification using individual clones may be more efficient in the end (Jun et al., 2005).

Even after isolation of the most productive clones, their performances must be evaluated in various types of media and in scaled-up scenarios for production in larger vessels. Recent effort has been focused on optimizing media formulations for the growth of these production cell lines. The optimum formulation for one high-producing clone may not be effective in achieving high production rates from another (Gerber et al., 2008). The need to find an optimum medium that may be specific to each clone accentuates the necessity of efficiently isolating high-producing clones upstream of the process of medium selection.

A number of host cell improvements have also been implemented for maximizing production through genetic engineering. Maintaining viability and favorable growth

characteristics remains a key goal in cell line development. It is common that isolated clones with high specific productivity have low growth rates (Arden et al., 2004). Furthermore, cell apoptosis during production acts to decrease viability. To address these problems, cell cycle control genes (cyclins), growth factor genes, and anti-apoptotic genes have been incorporated into cell lines for enhanced viability and optimal growth rate (Arden et al., 2004). In addition to high production, achieving product quality and efficacy is also important. This often requires the improvement of post-translational protein modifications. Protein glycosylation is perhaps the most prominent of these modifications. Examples of genetically engineering the glycosylation pattern in CHO cells include the overexpression of N-acetylglucosaminyltransferase-III (Umana et al., 1999), and the knockout of α -1,6-fucosyltransferase to modify the resulting glycoform for improved mediation of antibody dependent cellular toxicity.

Improved selection technologies

Increasing demand for biopharmaceuticals as well as the problems cited above has triggered the development of high-throughput methods for clone selection. One technology utilized for this process is flow cytometry, which allows the screening of large numbers of transfectants (Browne and Al-Rubeai, 2007). This method has been implemented by immobilizing a scaffold on the surface of cells with a high affinity for the recombinant protein (Holmes and Al-Rubeai, 1999). Using a fluorescent ligand, the cells with the highest fluorescence can be isolated using a fluorescence-activated cell sorter (FACS). The isolated high producers can then be cloned using the methods

described above or by FACS sorting into 96-well dishes and evaluated for productivity and growth rate. Another fluorescence-based selection platform that involves cell sorting utilizes fluorescently labeled MTX. Cells that are expressing high levels of DHFR can be made highly fluorescent by the addition of fluorescently labeled MTX that is specifically bound to the DHFR enzyme (Yoshikawa, 2001). These high DHFR producers are then isolated by flow cytometry on the assumption that they will be high producers of the protein of interest. A more recent method uses vectors with dual fluorescent labels that assess production of each chain of an antibody product. In particular, the reporters eYFP and eGFP are used to evaluate the production of the light chain and heavy chain of an antibody product, respectively. These reporter proteins are translated from an internal ribosome entry site (IRES) placed downstream of the coding regions of the proteins of interest. This method offers a single-step approach for isolating high producers and is a more accurate determinant of MAb productivity since the expression of both antibody chains is evaluated (Sleiman, 2007).

Another FACS-based method utilizes the Single Cell Secretion Assay (Borth et al., 2000) to measure production rates of individual cells by retaining secreted products on the cell surface after binding an externally applied antibody to the cells (Manz et al., 1995). This method was utilized to isolate high producers that also exhibited high production independent of growth rate and increased cell line stability. Sorting was performed on cells grown under conditions of cell density dependent growth inhibition. Cells were also cultivated without selection pressure for 180 days to isolate cells with increased stability. Using this assay with these specific conditions allowed for selection of desired variants after one sort and has the potential to play a key role in cell line optimization (Bohm et al.,

2004). FACS-based methods utilizing detection of a reporter protein have also been investigated. One study uses the gene for cell surface protein CD20 linked to the gene for the therapeutic protein through an IRES. Since the CD20 is co-expressed with the therapeutic protein, expression is detected using a fluorescent anti-CD20 antibody. This is advantageous since an antibody specific for the therapeutic protein is not required. This method is capable of analyzing expression of over 1000 clones per day (DeMaria et al., 2007).

Other methods have been developed that do not require the addition of an immobilization scaffold on the surface of cells. One such method takes advantage of the transient association of any secreted protein on the surface of cells. Since this occurs downstream of secretory bottlenecks, it is an accurate determinant of secreted protein from the cell (Brezinsky et al., 2003). Cold temperatures were used for this assay since endocytosis of protein product aggregates is slowed at low temperatures. Using fluorescently tagged antibodies that bind to the secreted product at low temperature and cell sorting, specific productivity of murine MAb AQC2 directed against human alpha 1 beta 1 integrin was increased up to 120-fold when used in conjunction with MTX amplification (Brezinsky et al., 2003). This cold capture assay has been demonstrated to be a robust process with over 20 secreted proteins tested for improved specific productivity (Brezinsky et al., 2003). Pichler *et. al.* performed a temperature dependency and time course study of this assay and determined that the fluorescent signal is stable for up to 24 h at temperatures ranging from 4 to 20°C (Pichler et al., 2009).

Automated systems (LEAP, Laser Enabled Analysis and Processing) have been developed that are capable of identifying high-producing adherent cells or colonies based on labeling the localized secreted product with fluorescent antibodies (Koller, 2004; Hanania, 2005). One method can be utilized for human and humanized IgG production. Cells are grown on a proprietary matrix that specifically retains secreted IgG. Next, a fluorescent-conjugated anti-IgG detection reagent is added to the culture and binds to the secreted IgG; cells are also generally labeled with a viable cell dye. Upon imaging, an algorithm calculates the intensity surrounding each cell and targets lasers to destroy low-producing cells, sparing high producers which can then proliferate without competition (Cresswell et al., 2009). The LEAP method is depicted in Fig. 1. Another system uses a robotic colony picker to isolate highly fluorescent colonies grown in a semi-solid medium (Lee, 2006).

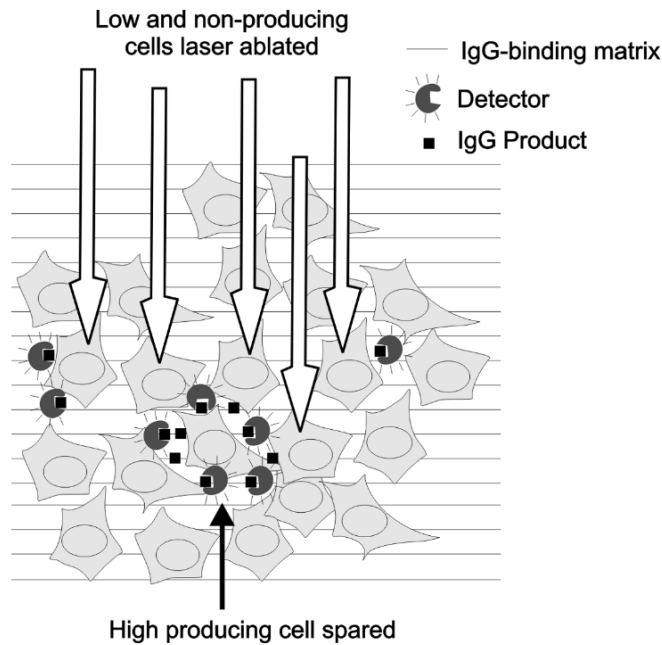


Figure 1. Laser-enabled analysis and processing of IgG secreting cells. IgG molecules that are secreted are captured by binding to a proprietary matrix on which the cells are growing. Accumulated secreted products are then detected using a fluorescently labeled anti-IgG. Low and non-producing cells are automatically laser ablated but the cells passing a threshold of fluorescence are spared and allowed to grow into colonies.

Although these methods of screening high producers in a high-throughput manner decrease development time significantly, the technology is expensive and is often particular to each cell line. Specific antibodies directed against the protein drugs being produced may be difficult to obtain (Browne and Al-Rubeai, 2007).

Improving production using modified amplification methods

The recently developed strategies to be discussed below differ from the selection technologies discussed above in that they modify gene vectors to increase the average production per cell *a priori* as opposed to developing technology for high-throughput

selection of clones from a large pool of cells *a posteriori*. These methods utilize vector engineering techniques and gene recombination reactions to obtain high levels of protein production. These methods aim to decrease the development time associated with standard MTX amplification without sacrificing production levels per cell. They also aim to achieve acceptable product yields without performing MTX amplification. Finally, they aim to decrease development time while minimizing additional development costs. Although the ideal strategy addresses all these goals, any method that achieves one of them can be a general tool for the development of many different biopharmaceuticals in a limited amount of time.

Traditional vectors used to clone the gene that codes for a therapeutic protein product often contain elements to enhance transcription of the gene. The gene is usually under control of a strong mammalian promoter such as the cytomegalovirus (CMV) promoter. The following three novel techniques engineer the *dhfr* vector to allow more stringent selection conditions by increasing the effectiveness of MTX amplification. One method adds a second marker for amplification and the other two methods weaken the *dhfr* marker so as to enhance its range of amplification.

Dual marker amplification

One method to increase the power of MTX amplification depends upon another amplifiable marker analogous to *dhfr* that allows extra amplification steps after the ability of MTX to select amplifiants is exhausted.

A typical approach to the use of more than one marker utilizes dicistronic vectors for the isolation of clones with enhanced production capacity for thyroid-stimulating hormone (TSH) (Peroni et al., 2002). The vector contains one functional unit of the hormone and one unit to be used in the selection process. A standard protocol for the production of this drug involves the co-transfection of a therapeutic protein gene and a *dhfr* gene either on two different plasmids or on the same plasmid, followed by MTX amplification. In this study, vectors were constructed that contained separate amplifiable marker genes. One vector contained the gene for the α subunit of TSH and *dhfr* and the other vector contained a gene for the β subunit and adenosine deaminase (ADA). Co-transfecting these two vectors allowed two sequential selections, first via MTX resistance and then via deoxycoformycin (dCF) resistance based on the amplification of the ADA gene. The dCF drug is a specific inhibitor of ADA. (Kaufman et al., 1986) Using MTX amplification alone, a secretion level of $7.2 \pm 1.3 \mu\text{g}/10^6$ cells per day of TSH was reported. Subsequent selection with dCF yielded a high-producing clone with a secretion level of $17.8 \pm 7.6 \mu\text{g}/10^6$ cells/day (Peroni et al., 2002). The dual amplification strategy yielded a 2-3 fold increase in TSH production in contrast to the yield of a single amplification step. However, this strategy requires a second selection protocol. A higher level of protein production came at the cost of a longer development time.

This protocol has also been applied to production of IgG (Fouser et al., 1992). The light chain gene was linked to the *dhfr* gene and the heavy chain was linked to the ADA gene. CHO cell lines that underwent dual amplification using these two markers produced a chimeric anti-ganglioside GD2 mAb at $80\text{-}110 \mu\text{g}/10^6$ cells/day (Trill et al., 1995).

Improving MTX amplification through weakening of the dhfr marker

The standard MTX amplification protocol can be modified to obtain higher producers by weakening the *dhfr* selection marker. Since gene amplification occurs over a DNA sequence that includes the *dhfr* gene and the therapeutic protein gene, their copy numbers both increase equally through each round of amplification. If *dhfr* activity is weakened, cells resistant to a specific dose of MTX will require a greater number of *dhfr* genes. Thus the amplification can be expanded to yield greater copy numbers for both the marker and production genes.

DHFR expression via incomplete splicing

A different amplification method used an engineered vector for the increased production of chimeric antibodies from CHO cells (Xiong et al., 2005). The vector contained a *dhfr* cDNA sequence upstream of the heavy chain sequence but a cytomegalovirus splice donor was inserted upstream of *dhfr* and an SV40 splice acceptor site downstream of *dhfr*, so that the *dhfr* coding region would be expected to be spliced out of the mature heavy chain mRNA. To the extent that this splicing fails, *dhfr* would be expressed. After transfection of this engineered vector and selection in –GHT medium, each step in MTX amplification resulted in the isolation of cells with a higher copy number than those isolated using transfection followed by amplification with a normal vector. This occurs since resistance to a certain MTX level depends on functional *dhfr*

being expressed. In this case the number of *dhfr* gene copies must be higher than the usual case in which all transcripts yield a functional *dhfr* mRNA since only a fraction of *dhfr* transcripts are properly spliced. Thus, this scheme allows the use of lower MTX concentrations at each amplification step to isolate cells with increased copy number. This strategy allows for more rounds of amplification before reaching the maximum MTX concentration at which selection for high copy numbers is effective and thus leads to the ultimate isolation of cells with higher copy numbers and production levels. This strategy is shown in Fig. 2.

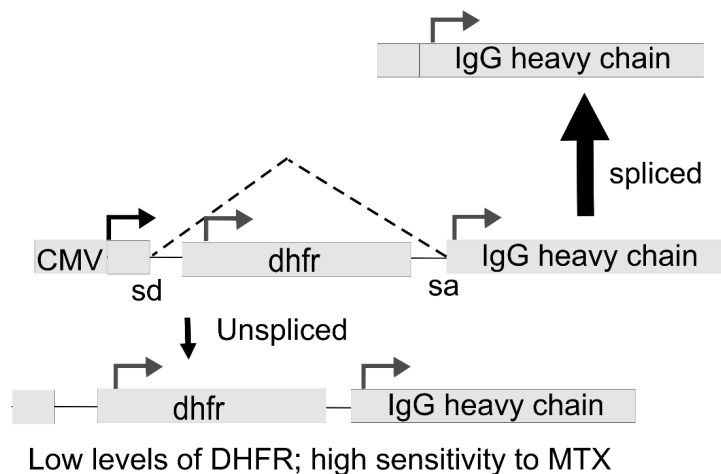


Figure 2. Enhanced amplification of *dhfr* using leaky splicing.

A *dhfr* cDNA gene is inserted into an intron in an IgG gene. In the rare instances in which it is not spliced out, it acts as an mRNA and is translated into DHFR enzyme. The low initial levels of DHFR confer ability to grow in purine-free medium but at the same time a high sensitivity to MTX and thus more dynamic range for amplification.

Production from clones containing this engineered vector were compared to a standard vector containing light and heavy chain genes for both anti-VEGF (vascular endothelial growth factor) antibody and HGAb (human glioma antibody). Using this method, MTX amplification using a standard vector accounted for an 18.4 fold increase in anti-VEGF antibody and a 17.6 fold increase in HGAb antibody production, both

compared to cells that were not subject to amplification. A 1.6 and 1.5 fold increase in antibody secretion was observed using the engineered vector compared to amplified cells using the standard vector for anti-VEGF antibody and HGAb antibody respectively (Xiong et al., 2005). This effort describes engineering of a vector that reduces the effort required to isolate a high-producing clone through a more stringent amplification protocol. Although the end result of this protocol is the isolation of a moderately higher producer, most of the secretion increase is still attributed to amplification and the development time is still time consuming.

Destabilization sequences on dhfr markers

Other techniques have been investigated to reduce selection marker strength in order to isolate higher producing clones. A method was developed that utilizes both mRNA and protein destabilizing sequence elements (Ng et al., 2007) to decrease the yield of DHFR per gene copy. If the *dhfr* marker is engineered to contain elements that act to destabilize its mRNA or the enzymatic activity of its product, it must integrate into the genome at a location favoring high expression for clones to survive (Gross and Hauser, 1995). The result is then an increase in the expression of the gene of choice since the marker and gene for the drug integrate at the same location. Naturally occurring sequences in the 3' untranslated region (UTR) have been found that act to destabilize mRNA. One such sequence is an AU-rich element (ARE), a nonamer containing exclusively adenine and uracil (Guhaniyogi and Brewer, 2001). Another type of destabilizing element, known as the PEST region, is from murine ornithine decarboxylase (MODC) and signals rapid

degradation of the enzyme. The very short half life of MODC is critical to its functionality (Rogers et al., 1986).

Vectors have been engineered containing both the *dhfr* marker gene and the gene for interferon-gamma (IFN γ), a biotherapeutic cytokine. Vectors containing these destabilizing elements for DHFR were compared to standard vectors by analyzing IFN γ production using ELISA (enzyme-linked immunosorbent assay). One engineered vector contained only the ARE element for *dhfr* mRNA destabilization, another contained only the MODC PEST region for DHFR enzyme degradation, and a third contained both destabilizing elements. Production titers of recombinant IFN γ using the vector containing both destabilizing sequences were found to be the highest. The addition of ARE, MODC PEST, and both sequences resulted in a 1.1, 3.7, and 12.6 fold improvement in IFN γ productivity respectively. This result was also seen by performing quantitative polymerase chain reaction (qPCR) measurements on cDNA molecules isolated from the different transfectants. Experiments also showed that this method can be used in conjunction with MTX amplification since these destabilizing elements do not affect the gene amplification process (Ng et al., 2007). These results show that the destabilization of the *dhfr* marker mRNA and enzyme product successfully isolated higher producing clones since many more copies of the therapeutic transgene are required for their survival in selective medium. This technique shows a significant increase in biopharmaceutical production without the use of a lengthy amplification step.

Vector enhancers to promote higher production levels of biopharmaceuticals

Instead of specifically integrating the therapeutic protein gene into a favorable expression location, transcriptional enhancing elements can be included in the gene vector itself. One such element is the gene promoter, which can be modified to maximize the expression of the gene. Another example of these elements is regulatory sequences that are responsible for remodeling chromatin, allowing the promoter of the gene to be accessible to transcription machinery. Two of the latter elements that will be discussed are chromatin opening elements and chromatin attachment regions.

Promoter Engineering

One of the more obvious means to achieve high levels of gene expression is to insert a strong promoter upstream of the therapeutic gene. Previously used strong mammalian promoters include that for the cytomegalovirus (CMV), the beta-actin promoter (Page and Sydenham, 1991), and the EF-1 α promoter (Allison et al., 2003). More recently, methods developed for protein engineering allow modification of promoters to enhance expression strengths. One such method used error-prone PCR to create a diverse library of promoters derived from a constitutive bacteriophage P_L- λ promoter that could be cloned into a vector to drive the expression of GFP in *E. coli* (Alper et al., 2005). The library exhibited a very wide range of fluorescence intensities. The utility of such a promoter library can allow one to isolate a promoter of the greatest strength by assaying for the gene. GFP fluorescence, mRNA levels quantified using

quantitative RT-PCR, and a minimum inhibitory concentration (MIC) of chloramphenicol applied to cells containing a gene vector for chloramphenicol acetyltransferase all showed a strong correlation, indicating that promoter engineering is generally effective. The promoter engineering technique was successfully applied to maximize growth yield and lycopene production in *E. coli* through control of the genes *ppc* and *dxs*, which produce phosphoenolpyruvatecarboxylase and deoxy-xylulose-P synthase respectively (Alper et al., 2005). Creation of a promoter library was extended to mutation of the *TEF1* promoter to drive yellow fluorescent protein (YFP) synthesis in *S. cerevisiae*. This extension showed that promoter engineering can be used to achieve a desired phenotype in both prokaryotic and eukaryotic cells.

Utilization of this method as a means of increasing biopharmaceutical production and shortening development time has not been explored in CHO cells or other candidate mammalian cells. Despite the initial time investment to develop a promoter for CHO cells expressing biopharmaceutical products, the use of an engineered promoter may prove to be a powerful method that could be combined with many of the other strategies described below.

Other promoters that have been investigated for maximizing production of biopharmaceuticals include the G1 phase specific growth-arrest and DNA damage inducible GADD153 promoter and the inducible MMTV promoter. The GADD153 promoter is active under serum-free conditions and in the G1 phase of the cell cycle. Secreted alkaline phosphatase (SEAP) and enhanced green fluorescent protein variant d2EGFP productivity under control of this promoter was highest in CHO cells at the G1 phase in medium without serum (de Boer et al., 2004). It was also shown that production

of both proteins under control of more commonly used promoters CMV and simian virus SV40 promoter was highest in complete medium at the S1 phase (de Boer et al., 2004). Since most cells during production are cultured in serum-free medium and exist in the G1 phase, the GADD153 promoter may be useful in increasing production of biopharmaceuticals in CHO cells. In another study, SEAP production in CHO cells was increased by more than 4-fold when induction of the MMTV promoter occurred at a high cell density when compared to induction at a low density (Lipscomb et al., 2004). By utilizing this inducible promoter at maximum cell density in a reactor, volumetric productivity can be maximized.

Chromatin opening elements

The specific genetic elements responsible for maintaining an open chromatin structure in the region of highly expressed genes are currently unknown. UCOEs (ubiquitous chromatin opening elements) are methylation-free CpG islands that possess a chromatin opening function making a DNA region more accessible to transcription machinery (de Poorter et al., 2007). CpG islands are regions with many unmethylated CpG dinucleotides. Methylation of this region is known to inhibit expression of a gene. One study investigated the use of CpG islands to inhibit transcriptional silencing and concomitant unstable expression of a transfected gene (Antoniou et al., 2003). In this study a vector was designed containing either TBP-PSMB1 (human TATA binding protein proteasome component B1) (Imbert et al., 1994) or hnRNPA2B1-CBX3 (heterogeneous nuclear ribonucleoprotein A2/B1) (Biamonti et al., 1994), ubiquitously

expressed housekeeping genes that naturally contain a methylation-free CpG island that extends over two divergently transcribed promoters (Kozu et al., 1995; Trachtulec et al., 1997). FACS analysis showed that a GFP reporter was stably expressed at high levels when transfected with the dual promoters from these housekeeping genes containing many CpG islands. This stable high expression was also evident when the gene was transfected at a poor location for stable expression such as within centromeric heterochromatin. These results suggest that expression vectors containing therapeutic genes based on dual promoter CpG island regions can provide high level, stable expression for maximum production of a therapeutic protein in mammalian cell lines and can ameliorate sensitivity to the location of vector integration in the genome. This result was confirmed by showing the inclusion of an UCOE enhanced expression of a CMV-driven cDNA cassette in stably transfected CHO cells (Benton et al., 2002). It was also shown that inhibiting histone deacetylation with sodium butyrate caused up to a 16-fold increase in reporter gene expression (de Poorter et al., 2007). Valproic acid can also be used as a cost effective alternative to sodium butyrate (Backliwal et al., 2008b).

Chromatin Attachment Regions

Another study was focused on scaffold or matrix attachment regions (S/MARS) of DNA and how they affect chromatin structure and consequently gene expression (Girod and Mermoud, 2003). S/MARS are genomic DNA sequences at which chromatin is anchored to the nuclear matrix during interphase (Mirkovitch et al., 1984). S/MARS are also known to be linked to histone hyperacetylation which indirectly recruits DNA

demethylase to demethylate DNA in order to make it accessible for transcription (Jost et al., 2001; Zhu et al., 2001). Results of *in situ* hybridization studies revealed that actively transcribed genes tend to be associated with the nuclear matrix. Previous studies revealed that S/MARS augment the expression of a reporter protein in stably transfected mammalian cell lines (Stief, 1989). In standard scenarios where the aim is to achieve a high gene copy number for maximum production, production increases are not always proportional to copy number increases since the amplified genes are in locations with varying expression characteristics. Experiments have shown that when each gene copy was flanked by an S/MAR, genes were expressed at levels proportional to copy number. Additional work has shown that genes proximal to S/MARS are expressed at higher levels than distal genes (Bode et al., 2000).

Protein production in CHO cells was tested using S/MARS to assess whether or not production would be improved through stronger gene expression. Chicken lysozyme (cLys) S/MARs (Bonifer et al., 1997) were added to two expression plasmids coding for heavy and light chain immunoglobulin (IgG). S/MAR elements increased IgG production in CHO cells 5-10 fold when S/MAR sequences flanked the coding gene (Zahn-Zabal et al., 2001).

Although technology using chromatin opening elements and chromatin attachment regions has proven to be useful in the CHO cell platform, the production increases are not as high as those associated with standard MTX amplification. However, further optimization of this method may achieve acceptable levels of protein with fewer rounds of amplification, and thus may decrease development time.

Utilizing gene targeting techniques for improved biopharmaceutical production

An approach fundamentally different from those described above is intended to increase recombinant protein production by specifically integrating a vector at a location with favorable expression characteristics. This goal can be approached by using site-specific recombination systems in the genome, a procedure that is in contrast to standard transfection protocols in which the vector is integrated randomly. Three such systems are described below.

Cre/LoxP recombination

One gene targeting technique utilized a Cre/loxP recombination system for targeting genes to sites in the genome that exhibit high expression levels (Kito et al., 2002). Cre recombinase is an enzyme that catalyzes the integration of a DNA sequence within a 34 base pair recognition site known as a loxP site. This reaction is reversible since after integration, two loxP sites are present and act as substrates for the enzyme to catalyze excision of the DNA sequence between these sites. Despite this disadvantage, the Cre/loxP system has a recombination specificity of about 80% (Raymond and Soriano, 2007). A vector was constructed that contained the genes for green fluorescent protein (GFP) and *dhfr* downstream of a loxP recognition site. After transfection into DHFR-deficient CHO cells, clones with the highest fluorescence were selected. Clones were then screened for a further increase in fluorescence after MTX amplification. The resulting clones could be deemed both transcriptionally active and gene amplifiable. In

the next step a targeting vector containing the antibody light and heavy chain genes fused to a hygromycin resistant marker and a loxP site was co-transfected along with Cre recombinase to catalyze the site-specific recombination. The result was the integration of the gene into a locus that is transcriptionally active and gene amplifiable. This resulted in clones producing 160 mg/L of human monoclonal antibody in 7 days in a spinner flask (Kito et al., 2002), a substantial improvement compared to the previously reported value of 40 mg/L obtained under similar conditions (Colcher et al., 1989). This procedure shows a significant improvement in production and decreased development time since gene integration is occurring at a location that has been pre-selected for its ability to amplify. This method may decrease the time-consuming amplification process by requiring less amplification in order to reach an acceptable production level. This method only requires an initial fluorescent screen to determine the optimum cell line. Once isolated, a gene coding for any therapeutic protein can be targeted to that location to exhibit the favorable expression and amplification characteristics. This gene targeting procedure also requires the use of Cre recombinase to perform the Cre/loxP recombination reaction which may increase development costs.

Flp/FRT recombination

A similar study incorporated the use of an analogous gene targeting system known as the Flp/FRT recombination system. In this system the enzyme Flp recombinase catalyzes the recombination of gene sequences that are tagged with a FRT sequence. This reaction is also reversible, similar to the Cre/loxP system. Unlike

Cre/loxP, the Flp/FRT system has a poor recombination efficiency (<10%) (Raymond and Soriano, 2007). Researchers (Huang et al., 2007) screened different gene integration sites using a vector with two weakened markers (β -galactosidase and *dhfr*). This vector contained a FRT sequence for subsequent recombination into the transcriptionally active sites. After selection of 20 candidate clones with amplifiable expression sites, three were successfully used as hosts for Flp recombination of antibody genes. The highest producing cell line was capable of producing 200 mg/L of full-length anti-CD20 antibody after 6 days of culture in a spinner flask. Clones were also isolated that were high producers of human anti-SARS antibody and scFv-Fc fusion protein, indicating that this technique can be applied to a broad range of antibodies (Huang et al., 2007). Zhou *et. al.* performed similar work using the Flp/FRT recombination system with a target vector tagged with FRT and a neomycin resistance marker. They isolated high producing clones of tissue plasminogen activator (tPA). The most productive cell line was capable of producing $17.1 \mu\text{g}/10^6 \text{cells/day}$ (Zhou et al., 2007). This method is very similar to the Cre/lox recombination method and has the same advantages and drawbacks.

Φ C31 integrase recombination

Another recombination system known as the Φ C31 system has an advantage over the two aforementioned systems in that the recombination reaction is irreversible. The Φ C31 integrase catalyzes recombination between an attP and attB site which are different sequences. The resulting recombination creates two hybrid sites that are no longer

substrates for the integrase, thus making the process irreversible. A second advantage is that several genomic pseudo-attP sites with high sequence similarity to the correct attP sequence can act as substrates for the enzyme, thus catalyzing integration into these sites. A disadvantage of Φ C31 when compared to Cre/LoxP is that Φ C31 integrase recombination specificity is less than 10%, similar to Flp recombinase. This issue has been solved through protein engineering efforts that created a mouse codon-optimized mutant of Φ C31 integrase known as Φ C31o (Raymond and Soriano, 2007). This mutant enzyme is shown to have a recombination specificity almost identical to Cre recombinase. Integration into mammalian pseudo-attP sites can be advantageous. CHO cells were co-transfected with a plasmid transcribing the Φ C31 enzyme and a plasmid containing an attB site for integration into pseudo-attP sites and the gene for luciferase. Luciferase expression was found to be 60-fold higher using this recombination system compared to random transfection (Thyagarajan and Calos, 2005).

Artificial Chromosomes

An Artificial Chromosome Expression (ACE) system utilizes murine artificial chromosomes (MACs) engineered to contain >50 site specific recombination sites (attP) specific to lambda integrase. These engineered chromosomes are generated *de novo* in a host cell line and then purified away from the natural chromosomes. They are then introduced into a recipient cell line suitable for biopharmaceutical production (Lindenbaum et al., 2004). After stable incorporation of the chromosome into this new host cell line, a target vector carrying an attB site and containing the gene of interest and

a vector coding for lambda integrase are co-transfected into the cells. Upon successful recombination, a drug resistant marker gene such as hygromycin becomes joined to an SV40 promoter conferring drug resistance. A second loading can be performed using a different drug marker such as zeomycin to further increase expression (Lindenbaum et al., 2004). This second loading required about an extra 2 months to generate clonal candidates; however it increased production titers by about 50% (Kennard et al., 2007). CHO cell lines have been used as a host for the ACE system to produce a human monoclonal IgG1 antibody. Candidate CHO clones were isolated in under 6 months and achieved titers of over 1g/L under non-optimized conditions making the ACE system competitive (Kennard et al., 2009). Use of such a technology may increase development costs and could save time as gene amplification was not used.

Each of the methods mentioned above aimed at increasing therapeutic protein production while minimizing development time and costs. They are tabulated in Table 1 based on several criteria.

Table 1: Summary of cell and vector development methods

<i>Strategy</i>	<i>Additional resources involved</i>	<i>Gene amplification required</i>	<i>Disadvantage</i>	<i>Advantage</i>
Dual marker amplification	ADA minigene; deoxycoformycin	+	Time-consuming second amplification	Tailored levels of two products
Weakened <i>dhfr</i> expression	ARE, PEST and splicing sequences	+	Only moderate improvement without amplification	Greater amplification potential
Vector sequences that enhance expression	Optimized promoters; S/MARS; CpG-rich sequences	-	Amplification still needed for high-level production	Less or no amplification needed
Gene targeting	LoxP,FRT,attB sequences; Cre, Flp, ΦC31, lambda recombinase genes	-	Amplification still needed for high-level production	Less or no amplification needed

Summary

Efficient isolation of high-producing CHO clones for biopharmaceutical products continues to be an industrial challenge. Standard methods are very time consuming and need improvement. While utilizing technology such as FACS or automated colony pickers based on fluorescence allows high-throughput selection, the procedures are often expensive and needful of constant modification to adapt to different cell lines and protein

products. The recent work involving vector engineering and site-specific recombination has shown promising results in isolating very high-producing clones. Implementation of such strategies has the potential to increase overall production titers of biopharmaceutical drugs and reduce the time required to isolate useful amounts of protein for pre-clinical evaluation.

A dual marker amplification strategy increases the maximum production of high-producing clones but increases development time. Utilizing vectors that weaken the *dhfr* selection marker still require time-consuming MTX amplification for maximum product yield. Site-specific recombination techniques that identify locations in the genome with favorable expression characteristics can decrease the time required to isolate high-producing clones. Although these strategies require both initial screening for amplifiable sites, this process only needs to be performed once to produce a cell that can be utilized to produce a range of desired gene products. However, the use of a site-specific recombinase may increase development costs. Use of an optimized promoter within the vector containing the gene for the biopharmaceutical of choice has potential to decrease development time but has not been sufficiently tested in CHO cells or other mammalian production platforms. Utilization of other vector enhancers that make genes coding for the drug of choice accessible to transcription machinery has been explored in CHO cells and does not result in production increases comparable to MTX amplification. Nevertheless, high producers can be isolated in a short period of time which obviates the need for a lengthy amplification step, and these techniques could be combined with an abbreviated amplification regimen. Overall, progress toward decreasing development time of high-producing clones has been made. The methods reviewed also show progress

toward increasing maximum therapeutic protein productivity per cell. These enhancements appear largely independent of the therapeutic protein being produced and can be expected to become easier to use with additional experience.

Acknowledgements

We thank Mauricio Arias, Ron Gejman, Shengdong Ke, Shulian Shang, and Dennis Weiss for useful discussions. JJC was supported by a NSF GK-12 Graduate STEM Fellowship while this manuscript was being written.

References

- Allison DS, Brandenstein M, Davis L, Running Deer J, Shah N, Ziegler K. Rapid development of CHO cell lines for high-level production of recombinant antibodies. *BioProcessing Journal*, 2003; 2, 33-40.
- Alper H, Fischer C, Nevoigt E, Stephanopoulos G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci*, 2005; 102(36), 12678-83.
- Antoniou M, Harland L, Mustoe T, Williams S, Holdstock J, Yague E, Mulcahy T, Griffiths M, Edwards S, Ioannou PA, Mountain A, Crombie R. Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. *Genomics*, 2003; 82, 269-79.
- Arden N, Nivtchanyong T, Betenbaugh MJ. Cell Engineering blocks stress and improves biotherapeutic production. *Bioprocessing*, 2004; 3, 23-28.
- Backliwal G, Hildinger M, Chenuet S, Wulhfard S, De Jesus M, Wurm FM. Rational vector design and multi-pathway modulation of HEK 293E cells yield recombinant antibody titers exceeding 1 g/l by transient transfection under serum-free conditions. *Nucleic Acids Res*, 2008a; 36(15), e96.
- Backliwal G, Hildinger M, Kuettel I, Delegrange F, Hacker DL, Wurm FM. Valproic acid: a viable alternative to sodium butyrate for enhancing protein expression in mammalian cell cultures. *Biotechnol Bioeng*, 2008b; 101(1), 182-9.
- Bebbington CR, Renner G, Thomson S, King D, Abrams D, Yarranton GT. High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. *Biotechnology (N Y)*, 1992; 10(2), 169-75.
- Benton T, Chen T, McEntee M, Fox B, King D, Crombie R, Thomas TC, Bebbington C. The use of UCOE vectors in combination with a preadapted serum free, suspension cell line allows for rapid production of large quantities of protein. *Cytotechnology*, 2002; 38(1-3), 43-6.
- Biamonti G, Ruggiu M, Saccone S, Della Valle G, Riva S. Two homologous genes, originated by duplication, encode the human hnRNP proteins A1 and A2. *Nucleic Acids Res*, 1994; 22, 1996-2002.
- Bode J, Benham C, Knopp A, Mielke C. Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Crit Rev Eukaryot Gene Expr*, 2000; 10, 73-90.

- Bohm E, Voglauer R, Steinfeldner W, Kunert R, Borth N, Katinger H. Screening for improved cell performance: selection of subclones with altered production kinetics or improved stability by cell sorting. *Biotechnol Bioeng*, 2004; 88(6), 699-706.
- Bonifer C, Jäggle U, Huber MC. The chicken lysozyme locus as a paradigm for the complex developmental regulation of eukaryotic gene loci. *J Biol Chem*, 1997; 272, 26075-78.
- Borth N, Zeyda M, Kunert R, Katinger H. Efficient selection of high-producing subclones during gene amplification of recombinant Chinese hamster ovary cells by flow cytometry and cell sorting. *Biotechnol Bioeng*, 2000; 71(4), 266-73.
- Brezinsky SC, Chiang GG, Szilvasi A, Mohan S, Shapiro RI, MacLean A, Sisk W, Thill G. A simple method for enriching populations of transfected CHO cells for cells of higher specific productivity. *J Immunol Methods*, 2003; 277(1-2), 141-55.
- Brown ME, Renner G, Field RP, Hassell T. Process development for the production of recombinant antibodies using the glutamine synthetase (GS) system. *Cytotechnology*, 1992; 9(1-3), 231-6.
- Browne SM, Al-Rubeai M. Selection methods for high-producing mammalian cell lines. *Trends in Biotechnology*, 2007; 25(9), 425-32.
- Chen C, Chasin LA. Cointegration of DNA molecules introduced into mammalian cells by electroporation. *Somat Cell Mol Genet*, 1998; 24(4), 249-56.
- Cockett MI, Bebbington CR, Yarranton GT. High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification. *Biotechnology (N Y)*, 1990; 8(7), 662-7.
- Colcher D, Milenic D, Roselli M, Raubitschek A, Yarranton G, King D, Adair J, Whittle N, Bodmer M, Schlom J. Characterization and biodistribution of recombinant and recombinant/chimeric constructs of monoclonal antibody B72.3. *Cancer Res*, 1989; 49, 1738-45.
- Cresswell JR, Lin N, Richardson GA, Kayser KJ. Cell XPress Applications in Development and Characterization of Biopharmaceutical Recombinant Protein Producing Cell Lines. In: M. Al-Rubeai, editor. *Cell Line Development*, Vol.6. Springer Netherlands, 2009. p.109-26.
- Davis J. *Basic Cell Culture-A Practical Approach*. Oxford University Press, 2002.
- de Boer L, Gray PP, Sunstrom NA. Enhanced productivity of G1 phase Chinese hamster ovary cells using the GADD153 promoter. *Biotechnol Lett*, 2004; 26(1), 61-5.

- de Poorter JJ, Lipinski KS, Nelissen RG, Huizinga TW, Hoeben RC. Optimization of short-term transgene expression by sodium butyrate and ubiquitous chromatin opening elements (UCOEs). *J Gene Med*, 2007; 9(8), 639-48.
- DeMaria CT, Cairns V, Schwarz C, Zhang J, Guerin M, Zuena E, Estes S, Karey KP. Accelerated clone selection for recombinant CHO CELLS using a FACS-based high-throughput screen. *Biotechnol Prog*, 2007; 23(2), 465-72.
- Fouser LA, Swanberg SL, Lin BY, Benedict M, Kelleher K, Cumming DA, Riedel GE. High level expression on a chimeric anti-ganglioside GD2 antibody: genomic kappa sequences improve expression in COS and CHO cells. *Biotechnology (N Y)*, 1992; 10(10), 1121-7.
- Gerber MA, Arunakumari A, Huang D, Goldstein J, S. M, Drew B, Caple MV, J.M. H, Wathen M, Kayser KJ. Integrated Strategies for Clone and Media Formulation Selection. *BioProcess International*, 2008; 6, 58-63.
- Girod PA, Mermod N. Use of scaffold/matrix attachment regions for protein production. *Gene Transfer and Expression in Mammalian Cells*, 2003; 38, 359-79.
- Gross G, Hauser H. Heterologous expression as a tool for gene identification and analysis. *J Biotechnol*, 1995; 41, 91-110.
- Guhaniyogi J, Brewer G. Regulation of mRNA stability in mammalian cells. *Gene*, 2001; 265(1-2), 11-23.
- Hanania EG. Automated in situ measurement of cell-specific antibody secretion and laser-mediated purification for rapid cloning of highly-secreting producers. *Biotechnol Bioeng*, 2005; 91, 872-76.
- Holmes P, Al-Rubeai M. Improved cell line development by a high throughput affinity capture surface display technique to select for high secretors. *Journal of Immunological Methods*, 1999; 230, 141-47.
- Huang Y, Li Y, Wang YG, Gu X, Wang Y, Shen BF. An efficient and targeted gene integration system for high-level antibody expression. *Journal of Immunological Methods*, 2007; 322, 28-39.
- Imanaka T, Aiba S. A perspective on the application of genetic engineering: stability of recombinant plasmid. *Ann N Y Acad Sci*, 1981; 369, 1-14.
- Imbert G, Trottier Y, Beckmann J, Mandel JL. The gene for the TATA binding protein TBP that contains a highly polymorphic protein coding CAG repeat maps to 6q27. *Genomics*, 1994; 21, 667-68.

- Jayapal KP, Wlaschin KF, Hu WS, Yap MGS. Recombinant Protein Therapeutics from CHO Cells-20 Years and Counting. *Chemical Engineering Progress*, 2007; 103, 40-52.
- Jost JP, Oakeley EJ, Zhu B, Benjamin D, Thiry S, Siegmann M, Jost YC. 5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation. *Nucleic Acids Res*, 2001; 29, 4452-61.
- Jun SC, Kim MS, Baik JY, Hwang SO, Lee GM. Selection strategies for the establishment of recombinant Chinese hamster ovary cell line with dihydrofolate reductase-mediated gene amplification. *Appl Microbiol Biotechnol*, 2005; 69(2), 162-9.
- Jun SC, Kim MS, Hong HJ, Lee GM. Limitations to the development of humanized antibody producing Chinese hamster ovary cells using glutamine synthetase-mediated gene amplification. *Biotechnol Prog*, 2006; 22(3), 770-80.
- Kaufman RJ, Murtha P, Ingolia DE, Yeung CY, Kellems RE. Selection and amplification of heterologous genes encoding adenosine deaminase in mammalian cells. *Proc Natl Acad Sci*, 1986; 83(10), 3136-40.
- Kennard ML, Goosney DL, Ledebur HC. Generating stable, high-expressing cell lines for recombinant protein manufacture. *BioPharm International*, 2007; 20(3), 52-59.
- Kennard ML, Goosney DL, Monteith D, Zhang L, Moffat M, Fischer D, Mott J. The generation of stable, high MAb expressing CHO cell lines based on the artificial chromosome expression (ACE) technology. *Biotechnol Bioeng*, 2009; 104(3), 540-53.
- Kim SJ, Kim NS, Ryu CJ, Hong HJ, Lee GM. Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnol Bioeng*, 1998; 58(1), 73-84.
- Kito M, Itami S, Fukano Y, Yamana K, Shibui T. Construction of engineered CHO strains for high-level production of recombinant proteins. *Appl Microbiol Biotechnol*, 2002; 60, 442-48.
- Koller M. High-throughput laser-mediated in situ cell purification with high purity and yield. *Cytometry*, 2004; 61, 153-61.
- Kozu T, Henrich B, Schafer KP. Structure and expression of the gene (HNRPA2B1) encoding the human hnRNP protein A2/B1. *Genomics*, 1995; 25, 365-71.

- Lee C. High-throughput screening of cell lines expressing monoclonal antibodies. *Bioproc Int*, 2006; 4, 32-35.
- Lindenbaum M, Perkins E, Csonka E, Fleming E, Garcia L, Greene A, Gung L, Hadlaczky G, Lee E, Leung J, MacDonald N, Maxwell A, Mills K, Monteith D, Perez CF, Shellard J, Stewart S, Stodola T, Vandenborre D, Vanderbyl S, Ledebur HC, Jr. A mammalian artificial chromosome engineering system (ACE System) applicable to biopharmaceutical protein production, transgenesis and gene-based cell therapy. *Nucleic Acids Res*, 2004; 32(21), e172.
- Lipscomb ML, Mowry MC, Kompala DS. Production of a secreted glycoprotein from an inducible promoter system in a perfusion bioreactor. *Biotechnol Prog*, 2004; 20(5), 1402-7.
- Manz R, Assenmacher M, Pfluger E, Miltenyi S, Radbruch A. Analysis and sorting of live cells according to secreted molecules, relocated to a cell-surface affinity matrix. *Proc Natl Acad Sci U S A*, 1995; 92(6), 1921-5.
- Mirkovitch J, Mirault ME, Laemmli UK. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*, 1984; 39, 223-32.
- Morrow JK. Optimizing Transient Gene Expression: Applications Expected to Move Beyond Discovery and the Preclinical to Clinical Realm. *Genetic Engineering & Biotechnology News*, 2008; 28(5), 54-59.
- Ng SK, Wang DI, Yap MG. Application of destabilizing sequences on selection marker for improved recombinant protein productivity in CHO-DG44. *Metabolic Engineering*, 2007; 9, 304-16.
- Nunberg JH. Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate-resistant Chinese hamster ovary cell line. *Proc Natl Acad Sci*, 1978; 75(11), 5553-56.
- Page MJ, Sydenham MA. High level expression of the humanized monoclonal antibody Campath-1H in Chinese hamster ovary cells. *Biotechnology (N Y)*, 1991; 9(1), 64-8.
- Pearson S. Producing Protein Therapeutics By Mammalian Cell Culture. *BioProcess International*, 2007; 5, 30-37.
- Peroni CN, Soares CR, Gimbo E, Morganti L, Ribela MT, Bartolini P. High-level expression of human thyroid-stimulating hormone in Chinese hamster ovary cells by co-transfection of dicistronic expression vectors followed by a dual-marker amplification strategy. *Biotechnol Appl Biochem*, 2002; 35, 19-26.

- Pichler J, Hesse F, Wieser M, Kunert R, Galosy SS, Mott JE, Borth N. A study on the temperature dependency and time course of the cold capture antibody secretion assay. *J Biotechnol*, 2009; 141(1-2), 80-3.
- Puck TT, Marcus PI. A rapid method for viable cell titration and clone production with HeLa cells in tissue culture: the use of X-irradiated cells to supply conditioning factors. *Proc Natl Acad Sci*, 1955; 41, 432-37.
- Raymond CS, Soriano P. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One*, 2007; 2(1), e162.
- Rogers S, Wells, R., Rechsteiner, M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*, 1986; 234, 364-68.
- Sleiman RJ. Accelerated Cell Line Development Using Two-Color Fluorescence Activated Cell Sorting to Select Highly Expressing Antibody-Producing Clones. *Biotechnol Bioeng*, 2007; 99(3), 578-86.
- Stief A, Winter, D.M., Strätling, W.H., Sippel, A.E. A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature*, 1989; 341, 343-45.
- Thyagarajan B, Calos MP. Site-specific integration for high-level protein production in mammalian cells. *Methods Mol Biol*, 2005; 308, 99-106.
- Trachtulec Z, Hamvas RM, Forejt J, Lehrach HR, Vincek V, Klein J. Linkage of TATA-binding protein and proteasome subunit C5 genes in mice and humans reveals synteny conserved between mammals and invertebrates. *Genomics*, 1997; 44, 1-7.
- Trill JJ, Shatzman AR, Ganguly S. Production of monoclonal antibodies in COS and CHO cells. *Curr Opin Biotechnol*, 1995; 6(5), 553-60.
- Umana P, Jean-Mairet J, Moudry R, Amstutz H, Bailey JE. Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. *Nat Biotechnol*, 1999; 17(2), 176-80.
- Urlaub G, Chasin LA. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc Natl Acad Sci*, 1980; 77(7), 4216-20.
- Urlaub G, Käs E, Carothers AM, Chasin LA. Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell*, 1983; 33(2), 405-12.
- Wiebe ME. A multifaceted approach to assure that recombinant tPA is free of adventitious virus. *Advances in animal cell biology and technology for bioprocesses Op cit*, 1989, 68-71.

- Wurm F, Bernard A. Large-scale transient expression in mammalian cells for recombinant protein production. *Curr Opin Biotechnol*, 1999; 10(2), 156-9.
- Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol*, 2004; 22(11), 1393-8.
- Xiong KH, Liang QC, Xiong H, Zou CX, Gao GD, Zhao ZW, Zhang H. Expression of chimeric antibody in mammalian cells using dicistronic expression vector. *Biotechnology Letters*, 2005; 27, 1713-17.
- Yoshikawa T. Flow cytometry: an improved method for the selection of highly productive gene-amplified CHO cells using flow cytometry. *Biotechnol Bioeng*, 2001; 74, 435-42.
- Zahn-Zabal M, Kobr M, Girod PA, Imhof M, Chatellard P, de Jesus M, Wurm F, Mermoud N. Development of stable cell lines for production or regulated expression using matrix attachment regions. *J Biotechnol*, 2001; 87, 29-42.
- Zhou H, Liu ZG, Sun ZW, Yu WY. Development of site-specific integration system to high-level expression recombinant proteins in CHO cells. *Chinese Journal of Biotechnology*, 2007; 23(4), 756-62.
- Zhu B, Benjamin D, Zheng Y, Angliker H, Thiry S, Siegmann M, Jost JP. Overexpression of 5-methylcytosine DNA glycosylase in human embryonic kidney cells EcR293 demethylates the promoter of a hormone-regulated reporter gene. *Proc Natl Acad Sci*, 2001; 98, 5031-36.

Chapter 3

Recombinant protein production in CHO cells from a site selected for a high amplification rate

(Being revised for acceptance in the Journal of Biotechnology)

Jonathan J. Cacciatore^a, Edward F. Leonard^a, Lawrence A. Chasin^{b,*}

^a Department of Chemical Engineering, Columbia University, New York, NY 10027
United States

^b Department of Biological Sciences, Columbia University, New York, NY 10027 United
States

Corresponding Author: * 912 Fairchild Center, Mail Code 2433, Columbia University,
New York, NY 10027, United States Tel.:+ 1 212 854 4645 Fax: +1 212 531 0425
Email: lac2@columbia.edu (Lawrence Chasin)

Jonathan J. Cacciatore: Performed experiments, writing, and creation of figures

Edward F. Leonard: Assisted in writing and editing

Lawrence A. Chasin: Assisted with experimental design, writing, and creation of figures

Abstract

Co-amplification of transgenes using the dihydrofolate reductase/methotrexate (DHFR/MTX) system is a widely used method for the isolation of Chinese hamster ovary (CHO) cell lines that secrete high levels of therapeutic and diagnostic proteins. A bottleneck in this process is the stepwise selection for MTX-resistant populations; which can be slow, tedious and erratic. We sought to speed up and regularize this process by isolating *dhfr*⁻ CHO cell lines capable of integrating a transgene of interest into a defined chromosomal location that supports a high rate of gene amplification. We isolated 100 independent transfectants carrying a minigene for human adenosine deaminase (*ada*) linked to a ϕ C31 attP site and a portion of the dihydrofolate reductase (*dhfr*) gene. Measurement of the *ada* amplification rate in each transfectant using Luria-Delbruck fluctuation analysis revealed a wide clonal variation; sub-cloning showed these rates to be heritable. Site directed recombination was used to insert a transgene carrying a reporter minigene for secreted embryonic alkaline phosphatase (SEAP) into the attP site at this location in a high amplification rate clone (DG44-HA4). Subsequent selection for gene amplification of the reconstructed *dhfr* gene yielded reproducible rates of *seap* gene amplification and concomitant increased levels of SEAP secretion. This cell line as well as this method of screening for high amplification rates may prove helpful for the reliable amplification of recombinant genes for therapeutically or diagnostically useful proteins.

Keywords: fluctuation analysis, gene amplification, recombinant protein, CHO cells, DG44, *dhfr*

Abbreviations

CHO, Chinese hamster ovary; DHFR, dihydrofolate reductase; MTX, methotrexate; ADA, adenosine deaminase; dCF, deoxycoformycin; SEAP, secreted embryonic alkaline phosphatase; PTF, purine and thymidine free.

1. Introduction

The phenomenon of gene amplification is often exploited to produce therapeutic recombinant proteins, monoclonal antibodies being the prime example. A gene or genes coding for the therapeutic protein(s) of choice (i.e. monoclonal antibody heavy and light chain) are transfected into a host cell along with a selectable drug-resistance marker gene. After an initial isolation of transfectants exhibiting a minimal level of drug resistance, selection for transfectants that have acquired a modestly higher level of resistance by virtue of amplification of the marker gene is carried out. Many iterations of this selection process can result in cell clones with a high level of resistance, a high level of marker gene expression, and a high number of copies of the marker gene. Since the size of the amplified region is much larger than a single gene (Hamlin, 1992) and since transfected genes usually co-integrate in the host genome, even when co-transfected on separate plasmids (Chen and Chasin, 1998), the gene of interest is also co-amplified and its protein produced at a high level (Wigler et al., 1980).

A popular system for this approach uses a minigene specifying the enzyme dihydrofolate reductase (DHFR) as the selective marker (Ringold et al., 1981), methotrexate (MTX) as the drug and a Chinese hamster ovary (CHO) cell line deficient in this ubiquitous enzyme (Urlaub and Chasin, 1980; Urlaub et al., 1983) as a host. CHO cells are capable of gene amplification, a trait not shared by normal cells (Livingstone et al., 1992; Wahl et al.,

1984a; Yin et al., 1992). In this case the target of the drug is DHFR, and resistance is gained by overproducing it. Transfected cells are cultured in a medium lacking a source of purine and thymidine nucleotides. Since DHFR-deficient host cells are unable to synthesize these metabolites only the transfectants can grow. MTX is a specific and tight-binding inhibitor of DHFR, but cells that have undergone *dhfr* amplification (“amplificants”) can overcome a judiciously chosen concentration of the drug.

Although the *dhfr*-MTX gene amplification method can result in as much as a 500-fold increase in gene copy number (Hamlin, 1992), it suffers from long and variable development times. Each amplification step brings about only a modest increase in gene copy number, thus severely limiting the concentration of MTX that can be applied at each step. It is common to take six months or longer to isolate a cell line with the desired recombinant protein production level. This time bottleneck inhibits the rapid testing of multiple new candidates for pre-clinical evaluation and ultimately limits how quickly a new drug candidate gets to market (Trill et al., 1995).

Although the factors influencing gene amplification are not well understood, it has been shown that clones within a CHO transfectant population exhibit a wide distribution in the yield of amplificants (Kito et al., 2002; Wahl et al., 1984b). The simplest explanation for this variability is a position effect within the CHO genome. We reasoned that the gene amplification regimen could be both shortened and made more reliable by inserting the gene of interest along with the *dhfr* minigene at a defined locus in the CHO genome chosen for supporting a high amplification rate. We identified such sites by screening 100

transfectants for their rate of amplification (as opposed to the frequency of amplified cells) using Luria-Delbruck fluctuation analysis. A chosen site could subsequently be used to insert genes of interest via an included site-specific recombination sequence. A winning clone was shown to amplify a *dhfr* minigene along with a gene of interest at a rate higher than randomly integrated transfectants and to secrete the protein of interest at increased levels. This clone or similarly isolated clones may prove advantageous for those using gene amplification in CHO cells to drive high level production of therapeutically or diagnostically useful proteins.

2. Materials and methods

2.1 Plasmid construction

We constructed a vector (pKAPD1) to be transfected into host DHFR-deficient CHO-DG44 cells (Urlaub et al., 1983) with a neomycin resistance gene and an amplifiable human adenosine deaminase (*ada*) gene. A portion of the *dhfr* gene and a ϕ C31 attP site were included for subsequent recombination and amplification steps. A fragment containing the *dhfr* promoter, exon 1 and part of intron 1 was isolated by PCR using primers appended with AscI restriction site ends (forward: ACGTAGGCGCGCCGGGGCCTCTGATGTTCAAT; reverse: ACGTAGGCGCGCCCCCTCATGACTGTCCCTAA), digested with AscI, and ligated into an AscI site sequence that had been inserted into a unique BglII site on plasmid pEGFP-C3 so as to retain one BglII site (ClonTech). A synthetic 39 bp attP site for ϕ C31

recombinase (Thyagarajan et al., 2001) was then ligated into the unique BglII site on this plasmid (pKPD1) downstream of the *dhfr* insert. A human adenosine deaminase gene (*ada*) driven by a CMV promoter was isolated by ApaLI digestion from plasmid pCMV*ada* (American Type Culture Collection). The insert was ligated into the unique ApaLI site in pKPD1 to create vector pKAPD1 containing the *dhfr* promoter and exon 1, the attP site, a kanamycin/neomycin resistance gene, and the *ada* gene in that order (Fig. 1). Next, we constructed a plasmid (pBE26) to be used for recombination into the aforementioned attP site containing an attB site, the remaining 3' portion of the *dhfr* gene, and a *seap* gene coding for secreted embryonic alkaline phosphatase (SEAP) for future protein assays. A plasmid was first created by removal of the *dhfr* promoter, exon 1 and part of intron 1 from pDCH1P11 by digestion at a unique SmaI site upstream of the promoter and a unique PstI site within intron 1. The plasmid pDCH1P11 was previously constructed (Noe et al., 1999) by cloning a *dhfr* minigene, driven by the *dhfr* promoter and containing only intron 1 and one natural polyadenylation signal, between unique SmaI and HindIII restriction sites in pSP72 (Promega). A synthetic 34 bp ϕ C31 recombinase attB site (Thyagarajan et al., 2001) was ligated between these restriction sites so as to lie within intron 1, upstream of the contiguous *dhfr* exons 2 through 6. To create a vector (pSEAPBE26, Fig. 3) directing the synthesis of SEAP, a region of pBE26 comprised of the attB site, *dhfr* exons 2 through 6 and a *dhfr* polyadenylation site was isolated by PCR using primers with NotI restriction site ends (forward:

ACGTAGCGGCCGCGCCGATTCATTAATGCAGGT ; reverse:

ACGTAGCGGCCGCTGCTCTCAGGGGCTCTATGT) and ligated into the unique NotI

site in the plasmid pCMVSEAP (Addgene #24595) downstream of the *seap* gene driven by a CMV promoter.

2.2 *Stable transfection*

Vector pKAPD1 was linearized by digestion at a unique AgeI restriction site and transfected into host CHO DG44 cells by electroporation with low input DNA levels to promote single copy integration. Stable transfectants were selected with 800 µg/ml G418 for 2 to 3 weeks with medium renewals about every five days. One hundred colonies were isolated with cloning cylinders, expanded and frozen.

2.3 *Fluctuation tests*

Fluctuation analysis was used to measure the rate of amplification of the adenosine deaminase (*ada*) gene (Kaufman et al., 1986). High levels of this enzyme afford resistance to a combination of high adenosine and deoxycoformycin (dCF), an ADA inhibitor, as described in Results. A pilot study performed with several transfectant clones showed that the inclusion of 0.5 µM deoxycoformycin reduced survival in the presence of 1.1 mM adenosine to an average of about 10 colonies per 10⁶ treated cells and that dCF resistance in DG44 transfectants arises by amplification of the transgene, as described previously (Kaufman et al., 1986). Each of the 100 stable clones was seeded into 12 wells of a 96-well dish at a density of approximately 10 cells per well so that inclusion of a preexisting amplificant was highly unlikely. Cells in each well were grown

to confluence and then expanded into 12 wells in 6-well dishes. At about 75% confluence the cells in each well were challenged with 0.5 μM dCF plus 1.1 mM adenosine for 2 to 3 weeks. Surviving colonies were fixed with 3.7% formaldehyde and stained with crystal violet. Colony counts from the 12 wells were used to calculate an amplification rate (events/cell/generation) using the P_0 and mean methods (Luria and Delbruck, 1943) and the median method (Lea and Coulson, 1949). A given method did not always allow a calculation of an amplification rate; in these cases we were still able to calculate a maximum or minimum rate, as follows. Sometimes all 12 cultures yielded dCF-resistant colonies, undermining the calculation of P_0 , the proportion of cultures yielding no colonies. In these cases we estimated a minimum amplification rate by assuming that the thirteenth culture would have yielded no colonies, i.e., a P_0 of 1/13. In the opposite situation, when none of the 12 cultures yielded resistant colonies we assumed the 13th culture would have yielded colonies, for a P_0 of 12/13, so that the amplification rate was a maximum estimate. In other cases some wells contained colonies too numerous to count; here we assigned a value of 40 to the number of colonies; and so the estimate of the amplification rate using the mean or median methods was a minimum. Candidate clones were subcloned using limiting dilution and the measurements of amplification rates were repeated.

2.4 Dhfr and seap co-amplification and expression

To target the site specific recombination site, cells were co-transfected with pSEAPBE26 and pPGKPhiC31obpA (Addgene #13795, ref. (Raymond and Soriano, 2007)) for 6

hours using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions and selected for a functional *dhfr* gene by growth in purine and thymidine free medium ("PTF" medium: Alpha MEM without nucleosides and deoxynucleosides, Thermo Fisher) for 2 to 3 weeks. To generate transfectants that had undergone random integrations, cells were co-transfected with the vector pCMVSEAP for SEAP production and pDCH1P11 containing a complete *dhfr* minigene for amplification. MTX amplification was performed on the pools of stable transfectants that survived in each case. The concentrations of MTX used can be seen in Fig. 4; 2 to 3 weeks were allowed to gauge survival at each step. After isolation of DNA from these pools, copy numbers of *seap* were determined by triplicate real time PCR measurements using Sybr Green staining in an Applied Biosystems 7000 instrument. To determine absolute *seap* copy numbers, we set the value found for female human genomic DNA (Promega) equal to four, as the primers used (forward: GGCCTGACTGAGACGATCA; reverse: GAGAAGACGTGGGAGTGGTC) are able to amplify both the placental and the placental-like *seap* genes. The amount of secreted SEAP from each pool of cells was measured after a 24 hour accumulation period in serum-free medium using the Sensolyte pNPP Secreted Alkaline Phosphatase Reporter Gene Assay Kit (Colorimetric), from AnaSpec (Cat # 72144) with p-nitrophenyl phosphate as a substrate. Absorbance readings were taken at 405 nm after 30 minutes and related to a standard curve using purified SEAP provided in the Sensolyte kit assayed in the presence of serum free medium. Background levels using medium from untransfected cells were subtracted.

3. Results

3.1 Strategy

Our goal was to isolate a DG44 cell with a site-specific recombination site at a location where gene amplification occurs at a relatively high and reproducible rate. Our plan was to transfect DG44 cells with a plasmid carrying a site-specific recombination site linked to a minigene that can act as a selective marker for gene amplification. One hundred independent transfectant clones, each presumed to carry this construct at a different location, would be isolated. The gene amplification rate of each transfectant clone would be measured and a winner chosen for further study.

We used a plasmid carrying the attP from ϕ C31 (Thyagarajan et al., 2001) for site specific recombination and a human adenosine deaminase (*ada*) minigene for measuring gene amplification. High concentrations of adenosine inhibit the growth of most mammalian cells, but this inhibition can be reversed by high cellular levels of ADA, which detoxifies adenosine by catalyzing its deamination to inosine. Deoxycoformycin (dCF) is a potent ADA inhibitor that counteracts this detoxification. Cells in which the *ada* gene has been amplified can overcome the inhibition caused by a marginal amount of dCF (Ingolia et al., 1985). In particular, the use of an iterated selection for resistance to dCF in the presence of high adenosine has been successfully used for amplification of an *ada* minigene in CHO *dhfr*⁻ cells (Kaufman et al., 1986).

It is important to note that amplification rates cannot be determined by simply comparing the frequency of dCF-resistant colonies arising from each transfectant clone, as this frequency is dependent not only on the rate at which amplifiants are generated, but also on their subsequent growth in each culture. For instance, an amplification event that occurred at an early time in the growth of a culture will yield a high number of colonies even if such events occur at a modest rate. Fluctuation analysis was developed by Luria and Delbruck to separate the event rate from the products of those events (Luria and Delbruck, 1943). As applied here, the frequency of appearance of dCF-resistant colonies in a series of 12 parallel cultures is measured; assuming a Poisson distribution of amplification events, the rate of amplification can be calculated from the proportion of cultures with no resistant colonies (Po method), or from the variance or the median of the number of such colonies per culture. This type of analysis has been applied to both mutation rates (Chasin, 1973) and gene amplification rates (Tlsty et al., 1989) in mammalian cells. One hundred transfectant clones would then be individually screened for the rate of *ada* gene amplification on the basis of their resistance to a combination of high adenosine and dCF.

More specifically, we transfected DG44 cells with a plasmid carrying the attP site (Thyagarajan et al., 2001), an *ada* minigene, the 5' part of a *dhfr* minigene, and a neomycin resistance marker (Fig. 1). One hundred colonies resistant to G418 were cloned and their *ada* amplification rates were measured as described above.



Figure 1. DNA for measuring *ada* amplification rates. The linearized DNA construct transfected into DG44 cells for integration into random locations in the genome.

3.2 *Ada* amplification rates

Amplification rates (events per cell per generation) were measured for each of the 100 transfected *ada* genes, each presumably integrated at a different random location. Three methods were used: P_0 , mean, and median (see Methods). Amplification rates for individual transfectants ranged widely, from 0 (no amplification evident) to about 5×10^{-6} amplification events per cell per generation. An exact rate could not always be calculated (e.g., no cultures with colonies or colonies too numerous to count); in these cases, a maximum or minimum rate, respectively, could be estimated. A complete list of the data is presented in Supplementary Data Table S1. The results for all 100 clones are shown in ranked order in Fig. 2A and as a distribution in Fig. 2B. Almost all of the clones with rates above 7 per million cells per generation included at least one estimated rate and so were considered less reliable. Three transfectant clones showing high amplification (HA) rates numbered 4, 35, and 63 (see arrows in Fig. 2A) were chosen for further study (DG44-HA4, DG44-HA35, and DG44-HA63). Clones DG44-HA4 and DG44-HA35 yielded the highest amplification rates that were measurable by all three methods. Clone DG44-HA63 had an even higher amplification rate but its rate could only be measured with two of the methods (Supplemental Table S1). Subclones were isolated from each of these three clones to assess the heritability of the amplification rates. The subclones produced amplifiants at rates in the same range as their progenitors (open symbols in Fig. 2A).

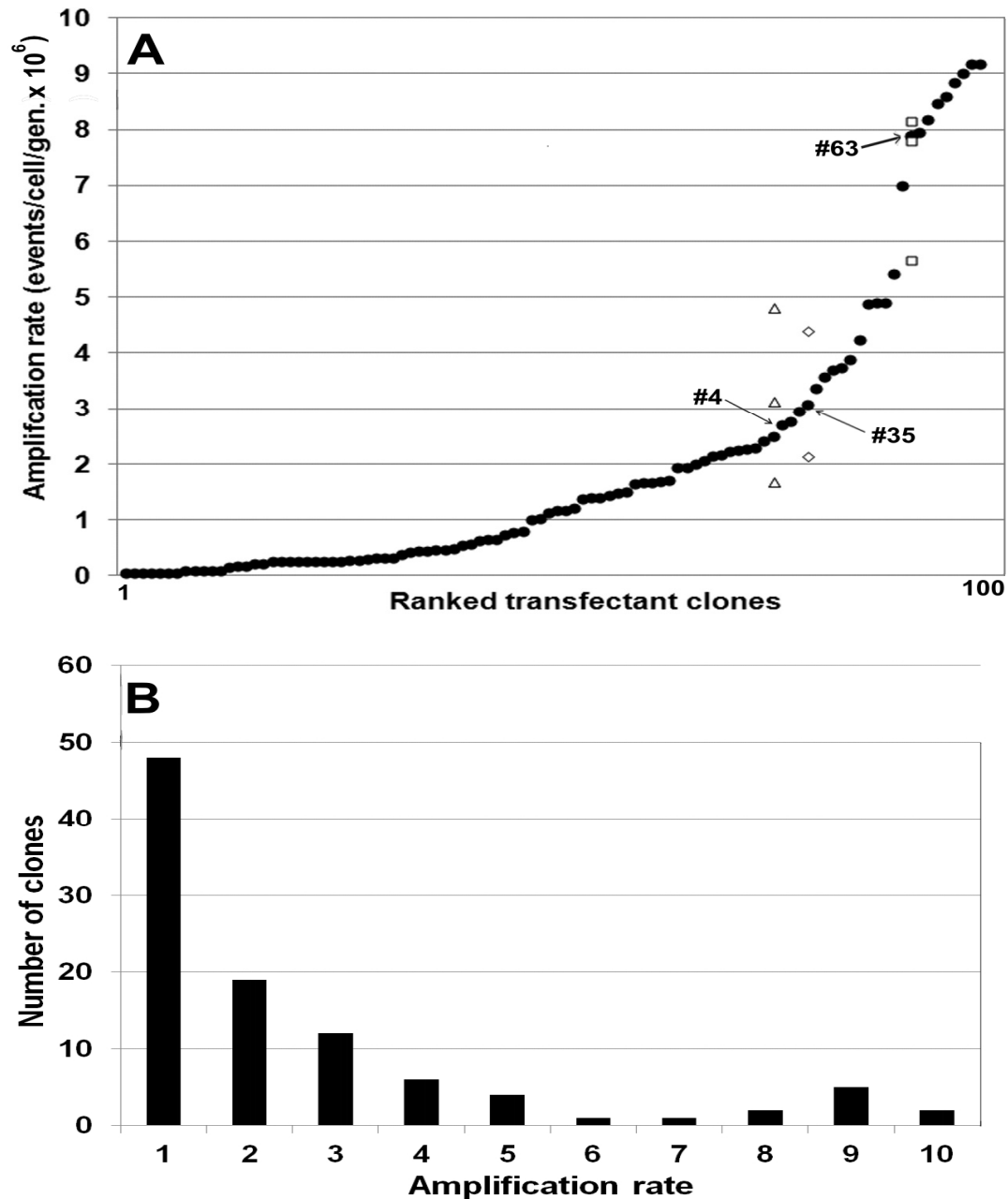


Figure 2. Amplification rates of transfectant clones. A. Clones are shown ranked; points represent the average of the Po, mean and median methods. Open symbols show results for subclones of the clone in the same column. These clones (4, 35 and 63, left to right) were chosen for further testing. B. Distribution of amplification rates; each bin number represents rates at or below that value and higher than the preceding value. Rates are defined as in A.

3.3 Gene amplification in high *ada* amplification rate clones

We next sought to determine whether a transfectant clone exhibiting a high rate of amplification for *ada* would be capable of amplifying additional transgenes at a comparable rate if inserted into the ϕ C31 attP site upstream of the *ada* minigene. For this experiment we used a plasmid carrying two informative sequences in addition to a ϕ C31 attB site for site-specific recombination. The first sequence was the downstream portion of the *dhfr* minigene comprised of part of intron 1 plus the fused exons 2 through 6; recombination between the attP and attB sites reconstitutes an active *dhfr* minigene after the two portions are joined within intron 1 (Fig. 3). This reconstituted *dhfr* gene was used to select for gene amplification by resistance to the DHFR inhibitor methotrexate (Nunberg et al., 1978). The second sequence was a cDNA-based minigene specifying human secreted embryonic alkaline phosphatase (SEAP). This enzyme is easily quantified and was used to gauge the usefulness of the transfectant clones for the production of secreted proteins of interest (Zhou et al., 2010). We chose clone DG44-HA4 for site specific recombination.

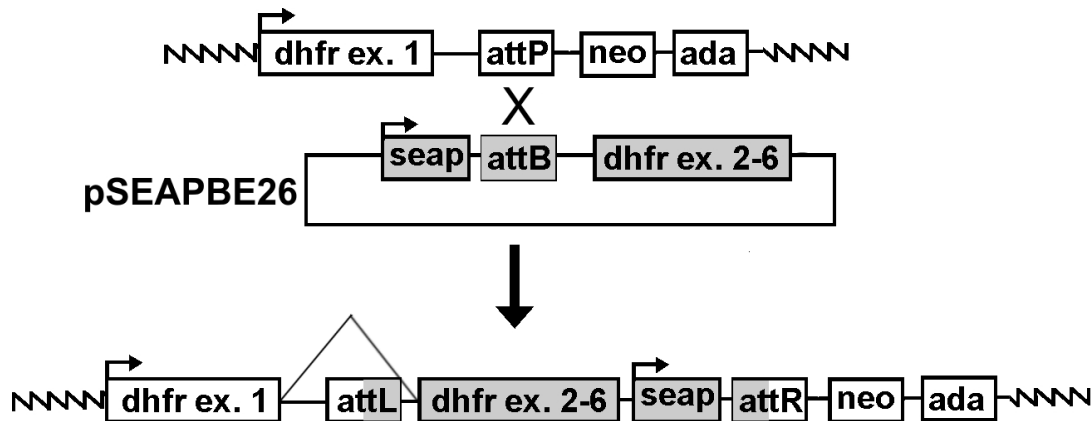


Figure 3. Site specific recombination scheme. Exons 2 through 6 of *dhfr* and a full *seap* gene were integrated at the location of the *attP* site in transfectant clone DG44-HA4 via site specific recombination with an *attB* site on the incoming plasmid pSEAPBE26. ϕ C31 recombinase was provided by co-transfection. The result was a reconstituted functional *dhfr* minigene together with a functional *seap* gene integrated into the same locus that supported a high amplification rate for *ada*. Successful site specific recombination was selected for by growth in purine-free and thymidine-free medium afforded by the reconstituted *dhfr* gene. Jagged lines: chromosomal DNA; lower line joining *dhfr* exons: pre-mRNA splicing pattern.

Clone DG44-HA4 was co-transfected with pSEAPBE26 and pPGKPhiC31obpA (recombinase) in three separate experiments, selecting for DHFR-positive populations by growth in purine and thymidine free (PTF) medium. Pooled survivors from each of these three experiments (4RA, 4RB and 4RC) were then tested for amplification of the reconstituted *dhfr* gene by selecting for step-wise resistance to MTX (0, 0.01, 0.02, 0.1, 0.5, 5, and 20 μ M). Resistance to 20, 20 and 5 μ M MTX was ultimately achieved for 4RA, 4RB and 4RC cells, respectively. In the case of 4RA, quantitative real time PCR (QPCR) was used to measure *seap* copy numbers at each stage of the selection. To determine the absolute number of *seap* genes at these stages we normalized the QPCR measurements to the value found with diploid human DNA (set to 4 due to the presence of two *seap* isozyme genes). As seen in Fig. 4, an overall amplification to about 600

copies of the *seap* gene was reached after 172 days. The original 4RA transfectants carried an average of 3 copies of the *seap* gene, but only one should be linked to the single functional reconstructed *dhfr* minigene so the fold amplification was 600. To verify that a *dhfr* minigene had been recombined into the ϕ C31 attP site in clone 4RA, we PCR amplified the region spanning the predicted attL recombination joint. As expected, the recombined region appeared after transfection and this region was amplified after MTX selection (Fig. 5).

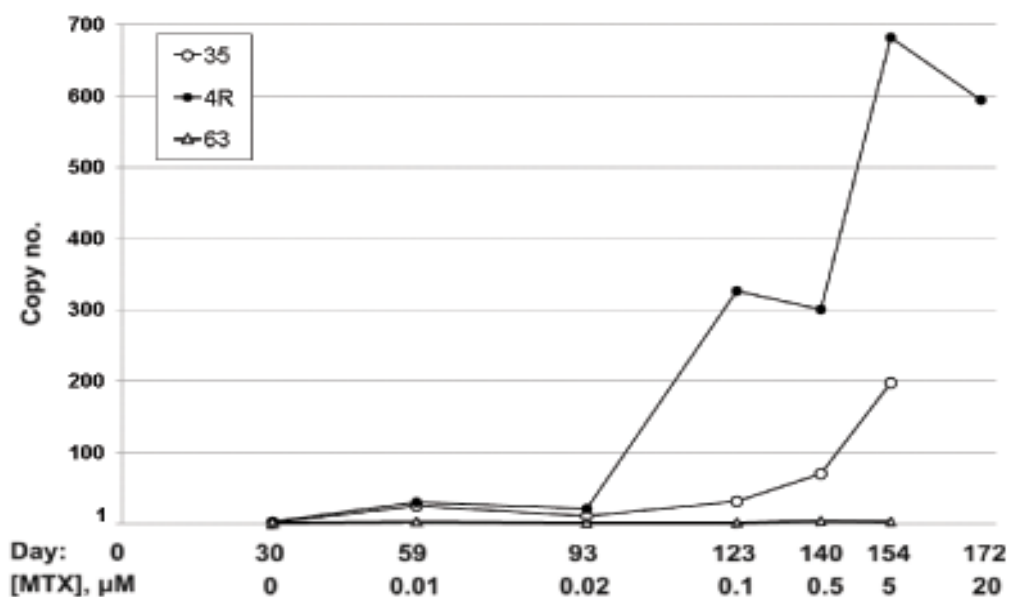


Figure 4. Gene amplification of transgenes in recombinant clones. *Seap* gene copy number was measured by QPCR after selection for *dhfr* gene amplification by methotrexate resistance as indicated. Closed circles: clone 4RA (site specific recombinant); open circles: clone 35A (random integration); open triangles: clone 63A (random integration). MTX, methotrexate.

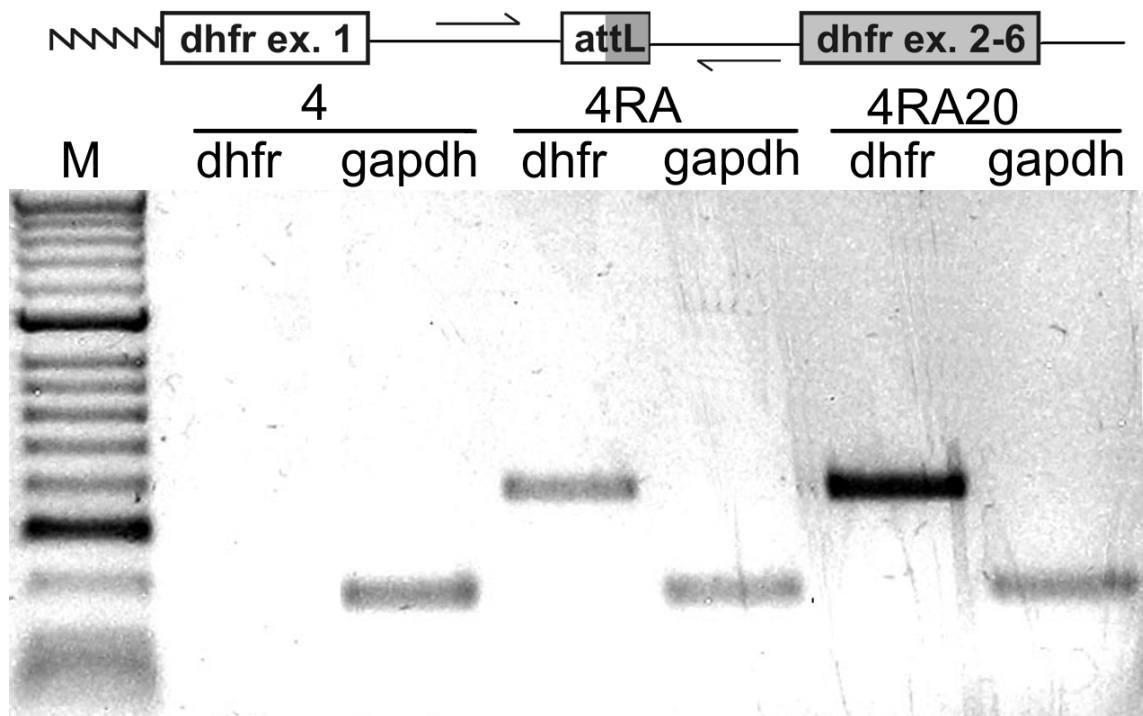


Figure 5. Recombined *dhfr* becomes amplified in clone 4. The region spanning the site specific recombination site (*attL*) was PCR amplified using primers depicted at the top. This joint appeared after transfection (compare 4RA vs. 4) and became amplified after MTX selection (compare 4RA20 vs. 4RA). *Gapdh*, a region of the glyceraldehyde-3-phosphate dehydrogenase gene as a control; 4RA, clone DG44-HA4 transfectant selected for DHFR activity; 4RA20, clone 4RA selected for resistance to 20 μ M MTX; M, HyperladderTM II (Bioline) molecular size markers.

For comparison, we transfected a plasmid carrying a complete *dhfr* mini-gene (as well as a *seap* gene) so as to confer a DHFR-positive phenotype without site specific recombination. This plasmid was transfected into two other clones that exhibited high *ada* amplification rates: clones DG44-HA35 and DG44-HA63 (Fig. 2). Two independent transfections were carried out for each clone; no ϕ C31 recombinase gene was co-transfected in these experiments. Selection in PTF medium yielded transfectants (35A, 35B, 63A, and 63B) that had integrated the complete *dhfr* gene at random sites (i.e., not requiring reconstitution). These cells were then subjected to MTX selection for *dhfr* gene amplification as above. For 35A and 63A resistance to 5 μ M MTX was achieved, but

35B and 63B only reached resistance to 0.5 μ M MTX. That is to say, no resistant colonies appeared at the next level of MTX challenge. *Seap* copy number measurements in the case of 35A showed gene amplification, but at a slower rate and to a lesser extent than the recombined 4RA cells. Surprisingly, 63A cells exhibited no evidence of gene amplification despite their resistance to high levels of MTX. We attempted to perform the recombination reaction via co-transfection similar to clone 4, however we observed no colony formation upon selection in PTF medium. This may be due to the inefficiency of the recombination reaction at the attP site in these clones. Also, recombination reactions were not performed in higher amplification rate clones (above 4) since rate data for such clones was not obtained yet.

The higher rate of amplification in the case of the plasmid recombined into the attP site (4RA) compared to random insertions (35A) is consistent with the idea that the higher rate of amplification is due to the chromosomal location of the former.

3.4 Recombinant protein secretion in transfectant clones

Along with gene copy number, we measured the secretion of SEAP in the cell cultures selected at each stage of the *dhfr* amplification process. SEAP enzymatic activity in the medium was assayed after a 24 hour accumulation at each selection step in all seven transfection/amplification experiments. As can be seen in Fig. 6A, SEAP secretion increased at a similar rate in the three experiments using the site specific recombinant DG44-HA4 ($R^2=0.89$ using all points from the three experiments). Two independent full

dhfr minigene transfectants of clone DG44-HA35 exhibited increased SEAP secretion at a rate similar to that of DG44-HA4, although there was somewhat more variation ($R^2=0.67$). The same experiment with clone DG44-HA63 yielded a low rate of increase in SEAP secretion during the MTX selection, in agreement with the gene copy number results showing a lack of gene amplification in this transfectant (Fig. 4).

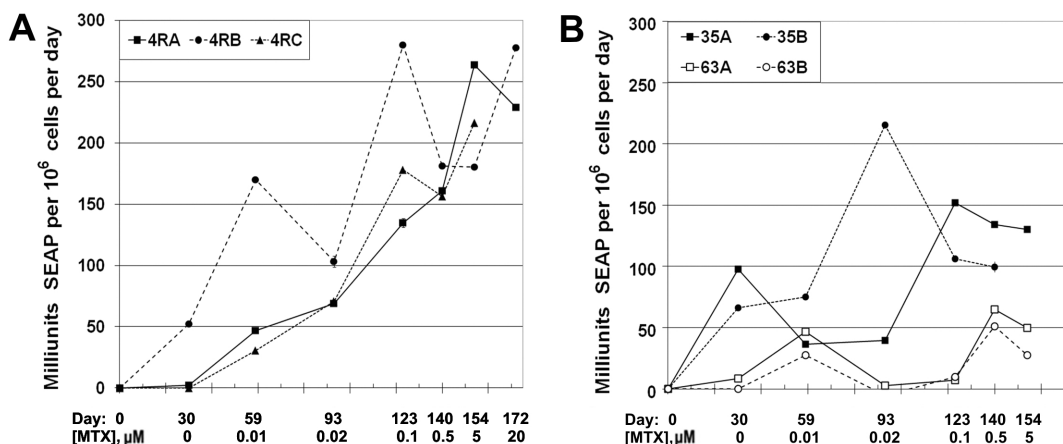


Figure 6. SEAP secretion during gene amplification. A. Transfection of a partial *dhfr* minigene into the ϕC31 attP site of DG44-HA4 to reconstitute a functional gene. The results of three independent amplification regimens with 3 independent transfectant pools are shown. Day zero is the day of transfection. Clone 4RC cells did not survive a challenge in 20 μM MTX. B. Transfection of the full *dhfr* minigene into random sites in DG44-HA35 and DG44-HA63. The results of four independent amplification regimens with 2 independent transfectant pools are shown. Neither of these clones survived in 20 μM MTX, and 63B cells did not survive a challenge in 5 μM MTX.

4. Discussion

Our intention was to isolate a clone with a site specific recombination site located at a chromosomal position that yields a high rate of gene amplification. Such a clone could prove advantageous for the production of recombinant proteins by inserting the corresponding gene into that location. Previous studies (Gajduskova et al., 2007; Kito et al., 2002; Wahl et al., 1984b) showed that only a minority of cell transfectants (27 of 82,

2 of 16, and 1 of 4 respectively) are capable of efficient transgene amplification.

Although high *frequencies* of gene amplification were documented in these and similar studies, there was no attempt to measure the amplification *rate*. Thus the high frequencies seen with individual transfectants could be due to the stochastic nature of the amplification process, such that an early event in any one experiment could be mistakenly interpreted to represent a high amplification rate. Here we have used fluctuation analysis to estimate amplification rates of 100 independent CHO cell transfectants. We saw a wide range of amplification rates among the transfectants: most transfectants exhibited low rates; two-thirds had rates that were less than the mean of 2×10^{-6} . A test of 3 clones chosen from the top quartile showed that their amplification rate was heritable, ruling out a simple statistical origin and as predicted if the rate depends on chromosomal location.

To see whether additional genes inserted in a recombination site would also be amplified at a high rate we transfected clone DG44-HA4 with a plasmid targeted to the recombination site. Successful recombination was selected for by the reconstruction of a functional *dhfr* gene and confirmed by PCR analysis. This plasmid also carried a transcription unit for SEAP. Three independently transfected populations derived from clone DG44-HA4 were subjected to a regimen selecting for increasing resistance to MTX. All three independently generated transfectant populations of clone DG44-HA4 responded to MTX selection by secreting increased levels of SEAP. The rate of increase was remarkably reproducible, yielding an r^2 value of 0.89 for the three transfection experiments combined (Fig. 6A). Stepwise amplification of the *seap* gene was confirmed by QPCR in one of the transfections, which ultimately generated approximately 600

copies of the *seap* gene. The *seap* gene here is standing in for a gene of interest; the reproducible increase in the production of the enzyme is consistent with the expectation that a targeted chromosomal location can confer reliability for recombinant protein production. In contrast, a pool of transfectants made up of random integrants of the full *dhfr* gene showed a lower amplification rate and more variability.

Clone DG44-HA63 acquired drug resistance by a means other than *dhfr* gene amplification. Alternative mechanisms of MTX resistance include altered folate permeation (Flintoff et al., 1976) or amplification of the endogenous P-glycoprotein gene leading to increased drug efflux (Assaraf et al., 1989). Perhaps in the initial screen the dCF-resistant colonies arose by one of these mechanisms rather than by *ada* gene amplification. For example clone DG44-HA63 may have started out with one allele of a dCF permease gene knocked out or an earlier amplification of the P-glycoprotein gene.

DG44 and DXB11, CHO *dhfr*⁻ cell lines have been widely used for gene amplification in the service of recombinant protein production (Jayapal et al., 2007). As described, clone DG44-HA4 could represent an improvement over these cell lines for providing more reliable amplification. That is not to say that clone DG44-HA4 is the optimum clone for this purpose, even among the 100 integrants screened here. Clone DG44-HA4 was chosen conservatively as exhibiting the highest average amplification rate of the three methods used among those with no estimated values and with a standard error of less than 10% of the average. It could be that one of the clones with a higher estimated amplification rate (e.g. $\sim 8 \times 10^{-6} \text{ cell}^{-1} \text{ gen}^{-1}$) would be a better choice.

A more general reason that clone DG44-HA4 may not always be the best general choice for gene amplification is that DG44 derivatives have been improved through genetic manipulation. As one example, considerable effort was invested in the isolation of a DG44 mutant lacking fucosyl transferase activity since unfucosylated immunoglobulins perform better in antibody dependent cellular cytotoxicity (Yamane-Ohnuki et al., 2004). In this and analogous cases, the method used here could be used to screen for an attP integration site supporting a high amplification rate in a special cell line.

In the experiments reported here no effort was made to screen for a high expression. Rather, the focus was solely on amplification rate. The reasoning behind this strategy was that gene expression can be increased by genetic manipulation of the incoming vector, through for example the addition of powerful promoter/enhancers (Allison et al., 2003), ubiquitous chromatin opening elements (de Poorter et al., 2007) and/or insulator elements (Zahn-Zabal et al., 2001). The incoming *dhfr* gene could also be weakened by including mutations that decrease the efficiency of the splice site preceding exon 2 (Chen and Chasin, 1993) to provide a greater range of amplification by starting at a lower MTX concentration.

In conclusion, we have described a method that can be used to select for mammalian cell clones with high and reliable rates of transgene amplification. The resulting clones could be used to facilitate the isolation of mammalian cell lines that produce recombinant proteins at high levels.

5. Acknowledgments

We would like to thank Mauricio Arias, Shengdong Ke, Vincent Anquetil, Laurens Moore van Tienen, Mrinalini Gururaj, Dennis Weiss, Ron Gejman, Ashira Lubkin and Ye Jung Ferraboli for insightful advice pertaining to experiments performed throughout this work. This work was funded by ImClone Systems, a wholly owned subsidiary of Eli Lilly and Company.

**Amplification rate
(events per cell per generation)**

Clone #	P₀	Mean	Median	Total Average
1	1.10	1.28	1.70	1.36
2	0.69	0.42	0.52	0.54
3	0.69	2.11	0.52	1.11
4	2.48	3.11	3.61	3.07
5	0.08 ^b	0.08	0.06	0.07
6	0.18	0.24	0.06	0.16
7	1.10	3.38	6.56	3.68
8	0.69	1.75	0.52	0.99
9	0.88	3.34	4.06	2.76
10	0.29	0.28	0.06	0.21
11	0.41	0.28	0.06	0.25
12	0.41	0.28	0.06	0.25
13	0.29	0.57	0.06	0.30
14	0.08 ^b	0.08	0.06	0.07
15	0.41	0.28	0.06	0.25
16	0.09	1.19	0.06	0.44
17	0.08 ^b	0.08	0.06	0.07
18	0.29	0.28	0.06	0.21
19	1.79	1.33	1.96	1.69
20	0.41	0.28	0.06	0.25
21	0.18	0.20	0.06	0.15
22	0.69	0.73	0.52	0.65
23	0.29	0.38	0.06	0.24
24	0.41	0.28	0.06	0.25
25	0.41	0.28	0.06	0.25
26	1.10	1.35	1.70	1.38
27	1.79	1.87	3.16	2.28
28	1.10	1.10	0.85	1.02
29	0.41	0.28	0.06	0.25
30	0.69	2.23	1.96	1.63
31	0.41	0.28	0.06	0.25
32	0.88	0.57	0.85	0.77
33	0.88	0.65	0.85	0.79
34	0.88	1.56	1.15	1.20
35	1.39	3.62	2.45	2.49
36	1.39	2.25	3.16	2.27
37	2.56 ^a	1.89	2.21	2.22
38	0.08 ^b	0.08	0.06	0.07
39	1.79	1.46	2.69	1.98
40	1.39	2.09	2.69	2.06
41	2.56 ^a	3.55	6.56	4.23
42	2.56 ^a	6.71	14.56	7.94
43	2.56 ^a	3.09	5.54	3.73
44	1.79	2.01	3.39	2.40
45	2.56 ^a	1.46	2.69	2.24

46	0.08 ^b	0.08	0.06	0.07
47	0.29	0.96	0.06	0.43
48	2.56 ^a	2.59	4.92	3.36
49	2.56 ^a	8.10 ^c	16.31 ^d	8.99
50	2.56 ^a	1.26	1.96	1.93
51	2.56 ^a	1.05	1.43	1.68
52	2.48	1.33	1.96	1.92
53	2.56 ^a	7.64 ^c	16.31 ^d	8.84
54	0.08 ^b	0.08	0.06	0.05
55	0.69	4.91 ^c	9.10 ^d	4.90
56	0.88	1.43	1.15	1.15
57	0.08 ^b	0.08	0.06	0.05
58	0.41	0.83	0.06	0.43
59	0.18	1.46 ^c	0.06	0.57
60	0.69	1.37	1.43	1.16
61	1.39	5.00 ^c	9.85 ^d	5.41
62	0.08 ^b	0.08	0.06	0.05
63	2.48	6.61^c	14.56^d	7.89
64	1.39	4.59 ^c	4.71	3.56
65	0.09	0.31	0.06	0.15
66	1.39	5.57 ^c	14.03 ^d	7.00
67	0.41	1.69 ^c	0.06	0.72
68	2.56 ^a	8.62 ^c	16.31 ^d	9.17
69	0.29	0.48	0.06	0.27
70	0.88	2.57 ^c	0.85	1.43
71	2.56 ^a	4.78 ^c	7.36	4.90
72	1.39	2.96 ^c	4.49	2.95
73	1.39	5.07 ^c	8.14	4.86
74	2.48	6.61 ^c	16.31 ^d	8.47
75	2.56 ^a	8.62 ^c	16.31 ^d	9.17
76	0.69	3.48 ^c	2.21	2.13
77	0.09	0.62	0.06	0.26
78	0.18	1.17	0.06	0.47
79	0.08 ^b	0.08	0.06	0.05
80	0.54	0.51	0.06	0.37
81	0.41	0.48	0.06	0.31
82	0.29	0.59	0.06	0.31
83	0.69	0.68	0.52	0.63
84	0.54	1.24	0.06	0.61
85	1.79	6.41 ^c	16.31 ^d	8.17
86	0.41	0.78	0.06	0.41
87	0.88	1.58	1.96	1.47
88	2.56 ^a	6.89 ^c	16.31 ^d	8.59
89	1.39	1.89 ^c	1.70	1.66
90	0.69	4.31 ^c	1.43	2.14
91	1.39	1.37	1.70	1.49
92	0.69	2.59 ^c	0.85	1.38
93	0.88	2.67 ^c	1.43	1.66
94	2.56 ^a	4.38 ^c	4.71	3.88
95	2.56 ^a	2.84 ^c	2.69	2.70

96	0.29	0.54	0.06	0.29
97	0.08 ^b	0.08	0.06	0.05
98	0.41	0.88	0.06	0.45
99	0.08 ^b	0.08	0.06	0.05
100	0.08 ^b	0.08	0.06	0.05

Supplemental Data Table 1. Calculated amplification rates using the Po, mean and median methods. High amplification (HA) rate clones chosen for further evaluation are shown in bold.

^a None of the 12 wells had zero colonies; a minimum estimate was made based on the assumption that the thirteenth well would have zero colonies.

^b All of the 12 wells had zero colonies; a maximum estimate was made based on the assumption that the thirteenth well would have more than zero colonies.

^c At least one well had colonies too numerous to count accurately; these wells were assigned value of 40 colonies; thus the estimate is a minimum.

^d At least 6 wells had colonies too numerous to count accurately; these wells were assigned a value of 40 colonies; thus the estimate is a minimum.

Clones in bold were chosen for further study.

References

- Allison, D.S., Brandenstein, M., Davis, L., Running Deer, J., Shah, N., Ziegler, K., 2003. Rapid development of CHO cell lines for high-level production of recombinant antibodies. *BioProcessing Journal*. 2, 33-40.
- Assaraf, Y.G., Molina, A., Schimke, R.T., 1989. Sequential amplification of dihydrofolate reductase and multidrug resistance genes in Chinese hamster ovary cells selected for stepwise resistance to the lipid-soluble antifolate trimetrexate. *J. Biol. Chem.* 264, 18326-18334.
- Chasin, L.A., 1973. The effect of ploidy on chemical mutagenesis in cultured Chinese hamster cells. *J. Cell Physiol.* 82, 299-307.
- Chen, C., Chasin, L.A., 1998. Cointegration of DNA molecules introduced into mammalian cells by electroporation. *Somat. Cell Mol. Genet.* 24, 249-256.
- Chen, I.T., Chasin, L.A., 1993. Direct selection for mutations affecting specific splice sites in a hamster dihydrofolate reductase minigene. *Mol. Cell Biol.* 13, 289-300.
- de Poorter, J.J., Lipinski, K.S., Nelissen, R.G., Huizinga, T.W., Hoeben, R.C., 2007. Optimization of short-term transgene expression by sodium butyrate and ubiquitous chromatin opening elements (UCOEs). *J. Gene Med.* 9, 639-648.
- Flintoff, W.F., Davidson, S.V., Siminovitch, L., 1976. Isolation and partial characterization of three methotrexate-resistant phenotypes from Chinese hamster ovary cells. *Somat. Cell Genet.* 2, 245-261.
- Gajduskova, P., Snijders, A.M., Kwek, S., Roydasgupta, R., Fridlyand, J., Tokuyasu, T., Pinkel, D., Albertson, D.G., 2007. Genome position and gene amplification. *Genome Biol.* 8, R120.
- Hamlin, J.L., 1992. Amplification of the dihydrofolate reductase gene in methotrexate-resistant Chinese hamster cells. *Mutat. Res.* 276, 179-187.
- Ingolia, D.E., Yeung, C.Y., Orengo, I.F., Harrison, M.L., Frayne, E.G., Rudolph, F.B., Kellems, R.E., 1985. Purification and characterization of adenosine deaminase from a genetically enriched mouse cell line. *J. Biol. Chem.* 260, 13261-13267.
- Jayapal, K.P., Wlaschin, K.F., Hu, W.S., Yap, M.G.S., 2007. Recombinant Protein Therapeutics from CHO Cells-20 Years and Counting. *Chemical Engineering Progress*. 103, 40-52.
- Kaufman, R.J., Murtha, P., Ingolia, D.E., Yeung, C.Y., Kellems, R.E., 1986. Selection and amplification of heterologous genes encoding adenosine deaminase in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3136-3140.

Kito, M., Itami, S., Fukano, Y., Yamana, K., Shibui, T., 2002. Construction of engineered CHO strains for high-level production of recombinant proteins. *Appl. Microbiol. Biotechnol.* 60, 442-448.

Lea, D.E., Coulson, C.A., 1949. The distribution of the numbers of mutants in bacterial populations. *J. Genetics.* 49, 264-285.

Livingstone, L.R., White, A., Sprouse, J., Livanos, E., Jacks, T., Tlsty, T.D., 1992. Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. *Cell.* 70, 923-935.

Luria, S.E., Delbruck, M., 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics.* 28, 491-511.

Noe, V., Ciudad, C.J., Chasin, L.A., 1999. Effect of differential polyadenylation and cell growth phase on dihydrofolate reductase mRNA stability. *J. Biol. Chem.* 274, 27807-27814.

Nunberg, J.H., Kaufman, R.J., Schimke, R.T., Urlaub, G., Chasin, L.A., 1978. Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate-resistant Chinese hamster ovary cell line. *Proc. Natl. Acad. Sci. U.S.A.* 75, 5553-5556.

Raymond, C.S., Soriano, P., 2007. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One.* 2, e162.

Ringold, G., Dieckmann, B., Lee, F., 1981. Co-expression and amplification of dihydrofolate reductase cDNA and the *Escherichia coli* XGPRT gene in Chinese hamster ovary cells. *J. Mol. Appl. Genet.* 1, 165-175.

Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., Calos, M.P., 2001. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell Biol.* 21, 3926-3934.

Tlsty, T.D., Margolin, B.H., Lum, K., 1989. Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell lines as measured by Luria-Delbruck fluctuation analysis. *Proc. Natl. Acad. Sci. U.S.A.* 86, 9441-9445.

Trill, J.J., Shatzman, A.R., Ganguly, S., 1995. Production of monoclonal antibodies in COS and CHO cells. *Curr. Opin. Biotechnol.* 6, 553-560.

Urlaub, G., Chasin, L.A., 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc. Natl. Acad. Sci. U.S.A.* 77, 4216-4220.

Urlaub, G., Kas, E., Carothers, A.M., Chasin, L.A., 1983. Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell.* 33, 405-412.

- Wahl, G.M., Allen, V., Delbruck, S., Eckhart, W., Meinkoth, J., de Saint Vincent, B.R., Vitto, L., 1984a. Analysis of CAD gene amplification using a combined approach of molecular genetics and cytogenetics. *Adv. Exp. Med. Biol.* 172, 319-345.
- Wahl, G.M., Robert de Saint Vincent, B., DeRose, M.L., 1984b. Effect of chromosomal position on amplification of transfected genes in animal cells. *Nature.* 307, 516-520.
- Wigler, M., Perucho, M., Kurtz, D., Dana, S., Pellicer, A., Axel, R., Silverstein, S., 1980. Transformation of mammalian cells with an amplifiable dominant-acting gene. *Proc. Natl. Acad. Sci. U.S.A.* 77, 3567-3570.
- Yamane-Ohnuki, N., Kinoshita, S., Inoue-Urakubo, M., Kusunoki, M., Iida, S., Nakano, R., Wakitani, M., Niwa, R., Sakurada, M., Uchida, K., Shitara, K., Satoh, M., 2004. Establishment of FUT8 knockout Chinese hamster ovary cells: an ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnol. Bioeng.* 87, 614-622.
- Yin, Y., Tainsky, M.A., Bischoff, F.Z., Strong, L.C., Wahl, G.M., 1992. Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. *Cell.* 70, 937-948.
- Zahn-Zabal, M., Kobr, M., Girod, P.A., Imhof, M., Chatellard, P., de Jesus, M., Wurm, F., Mermod, N., 2001. Development of stable cell lines for production or regulated expression using matrix attachment regions. *J. Biotechnol.* 87, 29-42.
- Zhou, H., Liu, Z.G., Sun, Z.W., Huang, Y., Yu, W.Y., 2010. Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. *J. Biotechnol.* 147, 122-129.

Chapter 4

Fluctuation Analysis Theory

Abstract

Fluctuation analysis is a method developed by Luria and Delbruck to calculate an inherent mutation rate. Its implementation allows for differentiating between the rate at which a mutation occurs and the frequency of its occurrence. This method was initially applied to studying bacterial mutation rate to confer resistance to phage virus. We have adapted this method to calculate gene amplification rates in Chinese Hamster Ovary (CHO) cells. We specifically utilize the amplifiable marker adenosine deaminase (ADA) and the addition of its specific inhibitor deoxycoformycin (dCF) in the presence of excess adenosine to select for amplified cells. We use the number of surviving colonies to calculate amplification rates using three methods: the P_0 method, mean method, and median method. An explanation of the equations used to calculate this rate is described. Analysis of amplification rate data for 100 separate CHO clones shows that there is a high fluctuation in the frequency of amplification events which occur in these cells. Using this analysis to measure amplification rate may be useful in isolating cell lines which can quickly produce high levels of therapeutic proteins.

Introduction

The demand for therapeutic proteins has been increasing rapidly over the past decade. This class of drugs is produced using mammalian cell lines. A majority of these proteins are currently produced using Chinese hamster ovary (CHO) cells for many reasons. One of the most important reasons is due to the fact that this cell line is known to have an unstable genome relative to other candidate cell lines considered for production. This inherent instability allows for the manipulation of their genome for high expression of a transgene, making them ideal candidates for biopharmaceutical production (Cacciatore et al., 2010). This inherent instability allows for the isolation of CHO cells with many copies of a desired transgene through a phenomenon known as gene amplification. Gene amplification is a term describing the stochastic duplication of a portion of the genome during cell mitosis, which can be propagated in subsequent cell divisions. The ability of the CHO cell to undergo this process is taken advantage of to isolate a high producing CHO clone with many copies of a desired transgene (Yin et al., 1992).

A gene or genes coding for the therapeutic protein(s) of choice (i.e. monoclonal antibody heavy and light chain) are transfected into the CHO host along with a selectable drug-resistance marker gene. After an initial isolation of transfectants exhibiting a minimal level of drug resistance, selection for transfectants that have acquired a modestly higher level of drug resistance, selection for transfectants that have acquired a modestly higher level of resistance by virtue of amplification of the marker gene is carried out. Many iterations of this selection process can result in cell clones with a high level of resistance, a high level of marker gene expression, and a high number of copies of the marker gene which in turn increases the specific productivity of each cell (Wigler et al., 1980). The

popular DHFR/MTX system used for this approach is explained in detail in Chapter 2 and Chapter 3.

Gene amplified cells are then utilized for production of the therapeutic product because it is the most efficient scheme in a fixed size reactor since each cell has its gene copy number maximized. The time needed to isolate these amplified cells is dependent upon the rate at which the transfected therapeutic gene amplifies or the mutation rate. The mutation in this case is the DNA duplication of the region containing the marker gene (*dhfr*) and the gene of interest. The calculation of mutation rate for a specific transfectant clone would allow for the estimation of time needed to isolate offspring of that clone containing a certain number of copies of the transgene which correlates to higher productivities. In short, a higher mutation rate would allow for quicker isolation of such a high producer. This overall process is depicted in Figure 1.

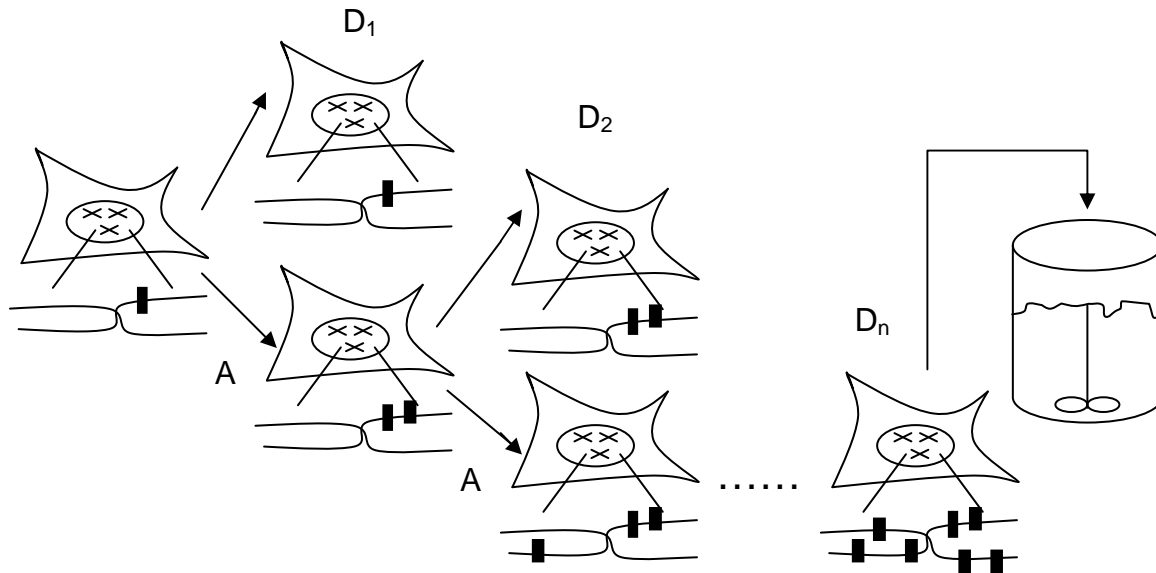


Figure 1-Schematic of CHO gene amplification process. A CHO cell containing one copy of a transgene (black rectangle) integrated into one of its chromosomes has an inherent amplification rate (A) which is the probability that it will duplicate in the genome in one cell division. One daughter cell (bottom) is shown to have two copies of the transgene after cell division 1 (D₁). Subsequently, the original gene copy amplifies a second time so that one daughter cell (bottom) has a total of three gene copies after cell division 2 (D₂) originating from the 2 gene copy number parent. After many cell divisions (D_n) a cell is isolated with many copies of the transgene and is ultimately used to seed a bioreactor to maximize protein productivity expressed by these many transgenes.

When selecting cells of a particular phenotype (eg. colony formation in selective medium containing a particular inhibitory drug) one may calculate a mutation rate through the observation of the survival phenotype. Several underlying assumptions for making this calculation are that there are no preexisting mutants, no lag in expression of the desired mutation being selected for, and no heterogeneity in mutant colonies (Andersson et al., 2011). One assumption that is not valid for this analysis is that mutant frequency is a perfect reflection of mutation rate. Despite the fact that the probability of this mutation occurring in a given cell division is a small fixed quantity; the time at which it occurs varies. Correcting for this while calculating mutation rates, or for our purposes

amplification rates (λ) is assessed using the fluctuation analysis of Luria and Delbrouck (L&D). This analysis and its application to CHO cells are discussed below.

Theory

The L&D analyses were based on the discovery that bacterial mutations can cause phenotypical changes such as resistance to certain viruses. L&D showed that such mutations occur independent of the presence of the virus. Each bacterium has an inherent probability that it will mutate from the sensitive phenotype to the resistant phenotype. Since this mutation may occur at any time, cultures of resistant bacteria of various sizes will exist depending upon what point in time such a mutation has occurred. Attempts to determine the proportion of resistant bacteria to the bacteriophage virus found very large variations. These large fluctuations are a consequence of the varying time in the culture where a bacterium mutated to become resistant to the virus and produced daughter cells which also exhibit resistance (Luria and Delbruck, 1943). L&D have formulated two different methods for determining mutation rate of a bacterium which is the chance of mutation per cell per time unit. These methods were derived on the assumption that there is a small fixed chance per time unit for each bacterium to undergo a mutation. These methods to determine mutation rate can be applied to other mutations as well as to eukaryotic cells. A requirement for such an analysis is that the mutation must be an observable property such as resistance to a specific agent. Similar to the bacterial mutation conferring bacteriophage resistance, this probability is captured experimentally if the potential gene undergoing mutation were a drug resistance gene. A random, low

probability gene duplication would result in cell survival when subject to an empirically determined concentration of the drug that the gene codes resistance for.

It has been previously hypothesized that the rate of gene amplification is dependent upon the location of the gene in the genome (Kito et al., 2002). In the case of producing biotherapeutics, genes for these proteins are randomly transfected into various locations in the genome and amplified. Presumably a large fraction of these transfectants have integrated the gene of interest and *dhfr* into a genomic location with an average or below average ability to amplify. However, statistically a small fraction will have integrated into genomic locations that are inherently prone to very fast amplification. In this work we aim to develop a method of identifying transfectants with the property of having an integrated transgene in a location with a high gene amplification rate. Determining which locations in the genome have high amplification rates and targeting a therapeutic gene to that location would prove to be valuable in shortening development time of biopharmaceuticals.

Development of an assay to determine amplification rates of various transfectants is not trivial. Such an assay needs to quantify a phenotype brought about through gene amplification that can be converted to a rate. We have developed an assay which converts the number of surviving colonies in culture to an amplification rate. This rate is the probability that a transfectant with a gene of interest at a specific genomic location will duplicate during a given cell division in an arbitrary time period. Directly correlating the number of surviving colonies to a gene amplification rate requires the

derivation of a mathematical model to distinguish colonies arising due to two possibilities:

1. An early gene amplification event and many daughter colonies forming from that early event
2. Many gene amplification events occurring later (after many more cell divisions have occurred)

Colonies arising due to the latter possibility would represent a high amplification rate while colony survival due to the first possibility would not. The equations used to calculate amplification probability per cell division which is directly related to the rate have been adapted from L&D.

L&D formulated several methods for determining mutation rate of a bacterium which is the chance of mutation per cell per time unit. In our specific case, the mutation is defined as a gene amplification event. It is known that the probability of this event occurring in a given time unit is very small. Despite this, similar to the bacterial mutation conferring bacteriophage resistance, this probability can be captured experimentally since the gene is a drug resistance gene. A low probability duplication event results in cell survival when subject to a low concentration of the drug that the gene codes resistance for. Our antagonist in this case is the drug deoxycoformycin (dCF) as opposed to phage in Luria and Delbruck's case. Our mitigator is a gene amplification event of the gene for adenosine deaminase which upon amplification will confer survival in CHO cells in the

presence of dCF and excess adenosine (Kaufman et al., 1986). The ADA/dCF amplification protocol is explained in detail by Kaufman et al. This event is analogous to a mutation conferring phage resistance in the case of L&D. A schematic of each case is shown in Figure 2.

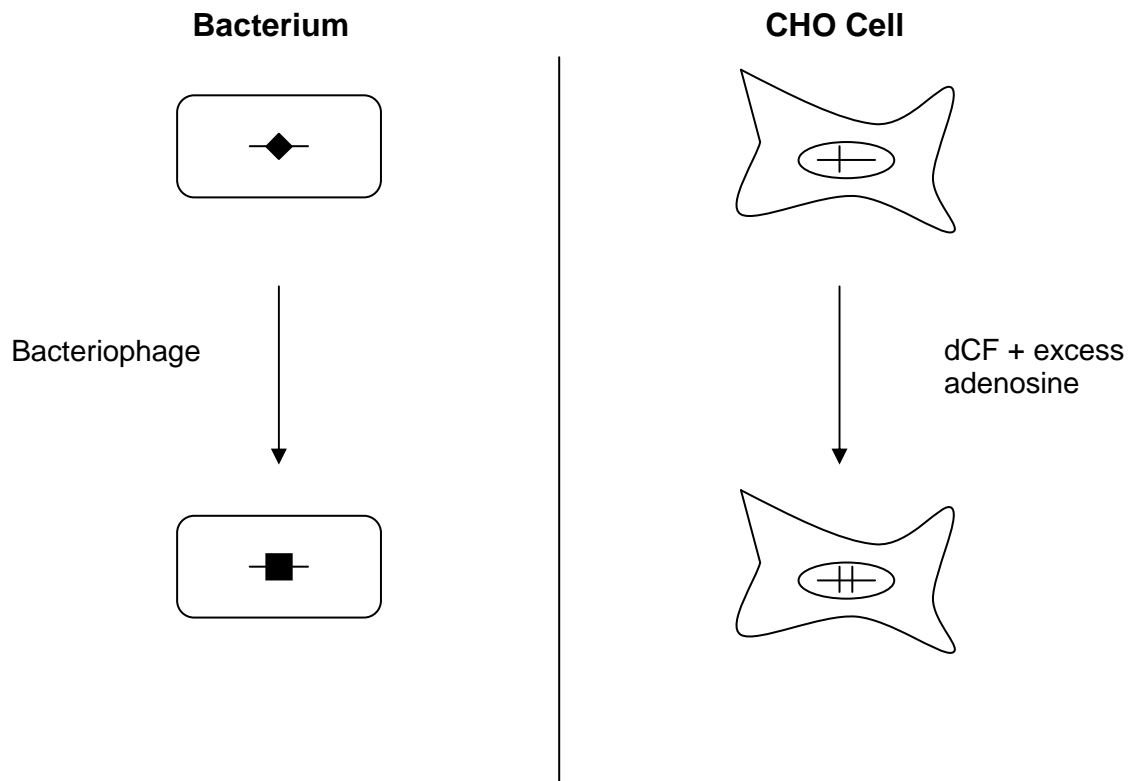


Figure 2 – Comparison of L&D mutation study vs. CHO gene amplification study. Left: If a bacterium undergoes a genetic mutation from the diamond to the square it will confer resistance to bacteriophage. Right: If the *ada* gene duplicates it will confer resistance to a low concentration of dCF and excess adenosine.

Two sets of equations have been derived in order to determine amplification rate of a given CHO clone (Luria and Delbruck, 1943). One is known as the p_0 method and one is known as the mean method. Both derivations are based on a fluctuation analysis, that is, the number of surviving colonies will fluctuate based on how early an event occurred. The assay is based on seeding transfectants in a specified number of parallel cultures

starting with a very low number of cells from which cell division will originate. The number of surviving colonies is counted and the amplification rate is calculated using either the fraction of cultures with no survivors (p_0) or the mean of the number of surviving colonies in each culture. The equations for each method applied to our system are outlined below.

P₀ Method

Equation 1: $dN_t/dt = N_t$

Equation 2: $N_t = N_0e^t$

Equation 3: $dm = adtN_t$

Equation 4: $m = a(N_t - N_0)$

Equation 5: $p_0 = e^{-m}$

Mean Method

Equation 6: $r = (t-t_0) aN_t$

Equation 7: $1 = aC(N_{t_0} - N_0)$

Equation 8: $N_{t_0} = N_t e^{-(t-t_0)}$

Equation 9: $t - t_0 = \ln(N_tCa)$

Equation 10: $r = aN_t \ln(N_tCa)$

m = number of gene amplification events

a = probability of gene amplification per unit time

N_t = number of cells in total culture at time t

N_0 = number of cells in total culture at $t = 0$

p_0 = fraction of cultures with no surviving colonies

r = average number of resistant cells

C = number of parallel similar cultures

t_0 = critical time at which the first amplification event occurs in culture

L&D assumed the time interval to be in units of division cycles of bacteria ($t/\ln 2$) which enables the use of equation 1 above. The integrated form of this result is equation 2. The p_0 method is formulated assuming that the number of amplification events in a culture (m) that occur during a time interval dt is equal to the chance of such an event occurring (a) multiplied by the number of cells in that culture, which gives rise to Equation 3.

The integrated form of this equation over a given time interval from 0 to t is shown in Equation 4. The number of amplification events in a culture will be small and it is assumed will follow a Poisson distribution ($P(x) = e^{-\lambda} \lambda^x / x!$). Equation 5 shows that P_0 , the fraction of cultures in which no amplification has occurred (no surviving colonies), is a function of the number of amplification events (m). This holds true since P_0 represents the probability of 0 amplification occurrences across all cultures. Knowing this value which is an observable quantity, we calculate the number of mutations using Equation 5 (the first term of a Poisson distribution; $p(0) = e^{-\lambda}$) and then use this value of m to determine the amplification rate (a) in Equation 4. This development allows us to calculate amplification rate as a function of the fraction of cultures with no surviving colonies.

The mean method is formulated using Equation 6. The average (r) number of cells that have undergone ADA amplification over a given time interval can be determined by calculating the average number of surviving colonies in a culture. Equation 7 shows that the average number of amplification events will equal 1 at a given time set equal to t_0 . Since the growth rate of cells (dN_t/dt) is proportional to the number of cells present, the number of cells present when the first amplification event occurs can be determined by

Equation 8. Substituting for N_{t0} in Equation 7 and knowing that the initial number of seeded cells (N_0) is negligible, one arrives at Equation 9. Combination of Equation 6 and Equation 9 results in Equation 10 which relates average number of amplification events across all cultures (r) to amplification rate (a). Equation 10 is the final equation utilized. This equation allows us to calculate amplification rate as a function of the observed mean of surviving colonies in parallel cultures.

A third equation (Equation 11) has also been implemented in calculating amplification rate that was developed by Lea and Coulson (L&C) known as the median method. This equation is solved for m to determine the number of mutations as a function of the median number of surviving colonies across all cultures (Lea and Coulson, 1949). After solving for the number of mutations per culture (Equation 11), one can use Equation 4 above to solve for amplification rate (a).

Median Method

$$\text{Equation 11: } (R_0/m) - \log m = 1.24$$

R_0 = median value of resistant colonies across all cultures
 m = number of mutations per culture

Specifically, our assay measures gene amplification by utilizing the drug marker adenosine deaminase. Adenosine deaminase is an enzyme which deaminates adenosine in the cell. Excess adenosine is toxic to the cell and the action of this enzyme prevents cell death by reducing intracellular concentration of adenosine. Deoxycoformycin (dCF) is a specific inhibitor of the ADA enzyme. Cells that have undergone gene amplification at the site of the ADA gene (thus increasing its gene copy number) will be able to survive in the presence of excess adenosine and dCF (Kaufman et al., 1986). In comparison to

L&D, CHO cells are analogous to bacterial cells, the presence of 0.5 μM dCF and 1.1 mM excess adenosine are analogous to the presence of bacteriophage virus, and an amplification of the ADA gene is analogous to a genetic mutation causing resistance to bacteriophage. In both cases, the observable phenotype is cell survival. We utilized the methods employed by L&D and L&C to calculate amplification rates on 100 different clones who have presumably integrated a copy of ADA at different genomic loci. The observable frequency (number of colonies) in 12 separate cultures was used to calculate rates from their p_0 , mean, and median.

Results

The detailed methods used for challenging each CHO clone and observing mutant or “amplificant” frequency is explained in our previous work from Chapter 3. Upon mimicking a fluctuation analysis we made several observations about the resulting data which can be seen in Table 1. Rates were calculated using all three of the aforementioned methods and the total average was taken. The calculated amplification rates are reported as duplication events per cell generation $\times 10^6$. The rates had a wide variation ranging from 0.05 to 9.17. In the case of clone 97, each of the 12 cultures failed to yield any colonies (amplificant frequency of zero). This represents the minimum extreme in which the clone has zero chance of undergoing a gene amplification event. In the case of clone 75, all cultures had too many colonies to accurately count, so the value was estimated to be 40 (the largest number that can be comfortably counted given the size of the culture dishes). This represents the maximum extreme in which the clone

presumably has the ability to amplify that particular region of DNA at the highest rate among the 100 that we tested.

A more detailed look at the data is shown where the raw numbers are listed for clone 4 out of 100 in Table 1. The number of counted amplificant colonies is represented for each culture well (C1 through C12). The value of 0.88 for the standard deviation divided by the mean confirms that there are indeed high fluctuations of amplificant frequency for this clone. A high deviation among cultures also exists among most of the 100 clones tested. This example further illustrates the need to apply these mathematical equations to correct for mutant frequency in determining rates. Another observation for this clone shows that there is good agreement between the determined A using each of the three methods. This tends to be true for lower calculated A values, but there tends to be poor agreement amongst the three methods for those with high calculated A values. This can be attributed to the inherent differences in the 3 methods since p_0 frequency is totally independent of mean and median, and mean and median can also be quite different for the cases where there are a few outliers within the 12 cultures.

Clone 4	
C1	5
C2	1
C3	4
C4	2
C5	24
C6	0
C7	5
C8	23
C9	8
C10	19
C11	26
C12	18
p0	2.48
std	9.92
std/mean	0.88
mean	3.11
median	3.61
Totavg	3.07

Table 1 – Raw colony count data for clone 4 used to calculate amplification rate. Calculations are based on 12 parallel cultures (C1-C12). Rates calculated using the p_0 , mean, and median method are bolded. The average of each of these values is shown as total average (Totavg). Standard deviation (std) and standard deviation relative to the mean (std/mean) are also tabulated.

Discussion

In this work we sought to apply fluctuation analysis which has been derived and successfully applied to bacterial mutations in previous works. We apply this analysis to a scenario in which identifying high mutation or amplification rates could be beneficial to the more efficient production of important therapeutic or diagnostic proteins. We developed a specific assay analogous to L&D's analysis of bacterial mutation rate against bacteriophage to assess how quickly a marker gene in a given genomic location will amplify over the period of several cell divisions. Isolation of a CHO clone with a high

amplification rate would presumably be an advantage over a clone with a low or average rate since this theoretically would result in higher protein production in a shorter amount of time.

After applying this analysis to our system using a marker gene transfected into 100 CHO host cells at various genomic locations and challenging with a specific ADA inhibitor, we found a wide range of amplification rates using 3 different derived mathematical methods. Since presumably the chromosomal location at which this marker gene integrated is the only variable in this process, this wide range agrees with previous results which infer that chromosomal location is a major determinant of amplification rate. This may be due to varying instability throughout various regions of the genome due to chromatin structure, surrounding genes, promoters and regulatory sequences, or proximity to centromeres. This “position effect” may be taken advantage of through the development of a cell line with the ability to integrate a transgene at that particular location for quick amplification and subsequent protein production. A scheme involving site-specific recombination to pursue this is discussed in previous work in Chapter 3.

The variability among specific methods (p_0 , mean, and median) is attributed to the amplificant frequency distributions among each clone. For example, a clone with all 12 parallel cultures containing amplifiants but each containing a low number of these amplifiants will correspond to a high rate using the p_0 method but a low rate using the median or mean method. In another example, a clone with most cultures containing zero amplifiants but a low fraction of the cultures with many amplifiants will correspond to

a low rate using the p_0 or median method but a high rate using the mean method. There is no direct answer as to which method is the best for calculating rates as each is mathematically derived in a logical fashion. However, it may be more appropriate to use one method over the other depending on the distribution of amplificant frequency observed among parallel cultures.

In conclusion, we show that fluctuation analysis can be applied to more sophisticated biological systems involving CHO cells undergoing transgene amplification and that this method has the potential to engineer cell lines for faster production of therapeutic proteins. Furthermore, such a method may be applied to other sophisticated systems in which the rate of a biologically important mutation is desired as opposed to simply the frequency of that mutation.

References

- Andersson, D.I., Hughes, D., Roth, J.R., 2011. *The Origin of Mutants under Selection: Interactions of Mutation, Growth, and Selection*. ASM, Washington, DC.
- Cacciatore, J.J., Chasin, L.A., Leonard, E.F., 2010. Gene amplification and vector engineering to achieve rapid and high-level therapeutic protein production using the Dhfr-based CHO cell selection system. *Biotechnol. Adv.* 28, 673-681.
- Hamlin, J.L., 1992. Amplification of the dihydrofolate reductase gene in methotrexate-resistant Chinese hamster cells. *Mutat. Res.* 276, 179-187.
- Kaufman, R.J., Murtha, P., Ingolia, D.E., Yeung, C.Y., Kellems, R.E., 1986. Selection and amplification of heterologous genes encoding adenosine deaminase in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3136-3140.
- Kito, M., Itami, S., Fukano, Y., Yamana, K., Shibui, T., 2002. Construction of engineered CHO strains for high-level production of recombinant proteins. *Appl. Microbiol. Biotechnol.* 60, 442-448.
- Lea, D.E., Coulson, C.A., 1949. The distribution of the numbers of mutants in bacterial populations. *J. Genetics.* 49, 264-285.
- Luria, S.E., Delbruck, M., 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics.* 28, 491-511.
- Ringold, G., Dieckmann, B., Lee, F., 1981. Co-expression and amplification of dihydrofolate reductase cDNA and the *Escherichia coli* XGPRT gene in Chinese hamster ovary cells. *J. Mol. Appl. Genet.* 1, 165-175.
- Urlaub, G., Chasin, L.A., 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc. Natl. Acad. Sci. U.S.A.* 77, 4216-4220.
- Wigler, M., Perucho, M., Kurtz, D., Dana, S., Pellicer, A., Axel, R., Silverstein, S., 1980. Transformation of mammalian cells with an amplifiable dominant-acting gene. *Proc. Natl. Acad. Sci. U.S.A.* 77, 3567-3570.
- Yin, Y., Tainsky, M.A., Bischoff, F.Z., Strong, L.C., Wahl, G.M., 1992. Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. *Cell.* 70, 937-948.

Chapter 5

Achieving immediate high transgene copy numbers using pre-amplified site-specific recombination sites

Abstract

Site-specific recombination (SSR) is a technology that can be utilized to integrate a DNA sequence into a specified location in the genome. We have developed a cell line which contains several thousand recognition sequences (attP sites) for the enzyme ϕ C31 integrase. Our goal is integrate many copies of a transgene into these locations. Despite the generation of many genomic attP sites, subsequent integration of a plasmid containing the gene for dihydrofolate reductase (*dhfr*) and an attB recognition sequence was inefficient. Using calcium phosphate-DNA co-precipitation to co-transfect this plasmid and a plasmid coding for ϕ C31 recombinase we integrated about 25 copies of the gene. We attempted to improve upon this modest recombination efficiency by increasing the concentration of the attB plasmid through polyoma virus large T-antigen replication but saw no improvement. We found that the addition of a specific 5 base pair flank on either end of both the attP and attB site improved recombination efficiency by over 20 fold as determined by an *in vitro* recombination assay. We plan to select for DNA sequence flanks that promote more efficient recombination and incorporate these sequences into cell lines with many recombination sites. This may allow for massive integration of a transgene for a therapeutic protein and decrease the time needed to isolate a highly productive cell line.

Introduction

The first step for producing biopharmaceutical drug candidates such as monoclonal antibodies is the permanent transfection of its DNA sequence into a host cell line such as Chinese hamster ovary (CHO) cells. After isolation of transfectants, a lengthy process known as gene amplification is utilized to increase the gene copy number, therefore increasing overall expression and protein secretion per cell. This is required since random transfection by various methods (calcium-phosphate co-precipitation, electroporation, liposomes) result in transfectants with 1-5 copies of the gene integrated, some of which are in poor expression locations (Zhou et al., 2010). In previous work we utilized site-specific recombination using ϕ C31 integrase to target a transgene to one specific genomic location which was deemed to be highly amplifiable to decrease the time necessary to isolate a high producing cell line. In this work we aim to extend the use of site-specific recombination technology to test whether many copies of a transgene can be integrated into the genome in one single recombination reaction, therefore abrogating the need for lengthy MTX amplification.

Site-specific recombination (SSR) is a novel method that has been investigated for the precise integration of therapeutic transgenes into host cell lines for production. This process is catalyzed by enzymes known as site-specific recombinases (Calos, 2006). This method allows one to integrate a transgene into the genome at a specific location which contains a recognition sequence for recombination into that site. This also relies on the transgene being coupled to another recognition sequence and the presence of a recombinase to bind each substrate and catalyze the recombination of the transgene into the genome. This can be thought of as a three bodied intermolecular reaction. One

molecule is the incoming DNA containing one SSR recognition sequence and the therapeutic transgene. The second molecule is the chromosome containing the other SSR recognition sequence. The third is the SSR enzyme that catalyzes recombination between the two sequences. In this work we aim to maximize all three components. First we aim to increase the number of genomic recombination sequences. Once we have maximized this component we switch to maximizing the number of incoming plasmids containing the other recombination site and our transgene. Finally, we focus on improving the efficacy of the site-specific recombinase. Our overall goal is to drive the reaction forward so that a high number of transgenes can be integrated into the genome in one process.

One recombination system we used was the cre/lox system which uses cre recombinase as the enzyme and DNA recognition sequences known as loxP sites (Hoess and Abremski, 1985). This system was initially chosen among other recombinase systems such as the flp/FRT system because it has been previously used in other studies and shows that its recombination efficiency is better than other recombinases (Raymond and Soriano, 2007). Initially, we sought to use recombinase mediated cassette exchange (RMCE) for the purpose of integrating a transgene between two loxP sequences at many genomic sites which we amplify using MTX amplification. This scheme first involved permanently transfecting a construct containing the endogenous *dhfr* gene with exon 2 flanked by two loxP sites. Next, we aimed to amplify this construct using MTX amplification to many copies so that targeting to these many locations can occur with a subsequent co-transfection of cre recombinase vector and an incoming plasmid with the same loxP sites

flanking a hygromycin resistance gene. Despite high efficiencies associated with the cre recombinase system, recognition sequences that are present after integration can act as substrates to promote gene excision in the presence of cre recombinase. Due to this fact, we also used another recombination system known as the ϕ C31 system since transgene integration cannot be reversibly excised (Thyagarajan et al., 2001a).

Unlike the cre recombinase system, upon site-specific integration of the transgene the recombined substrates (attL and attR sites) are not substrates for excision. Similarly to the cre system mentioned above, ϕ C31 recognition sequences (attP sites) are transfected into the CHO genome and amplified to many copies for subsequent targeting of a transgene to those many sites. A schematic of each process can be seen in Figure 1 and Figure 2.

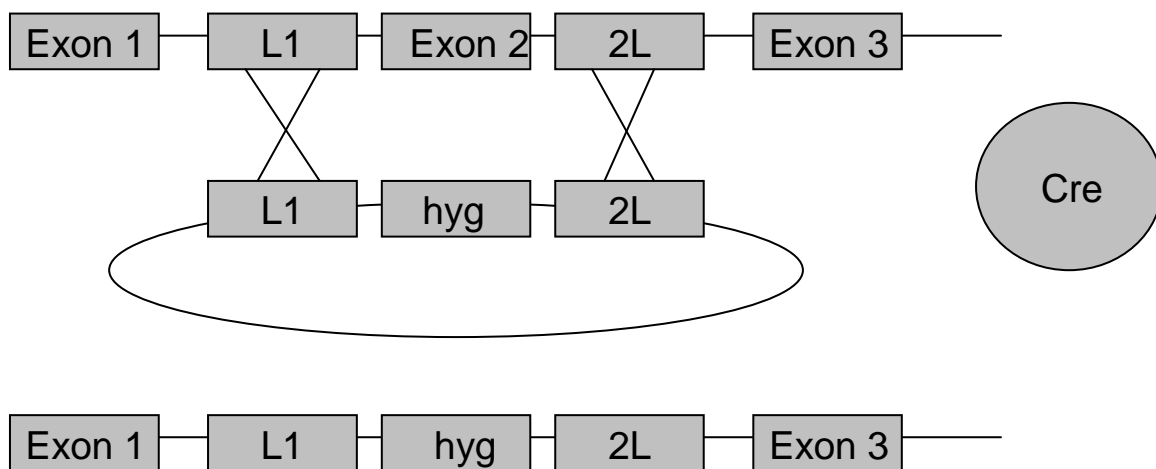


Figure 1 – Schematic of cre/lox recombination system. A construct containing the full *dhfr* gene (only first three exons shown) with lox sites L1 and 2L flanking exon 2. In the presence of cre recombinase the L1 and 2L sites will interact with the incoming L1 and 2L sites respectively which flank the *hyg* gene. The result of this recombinase mediated cassette exchange (RMCE) is the *hyg* gene integrated into the genome at the previous location of exon 2.

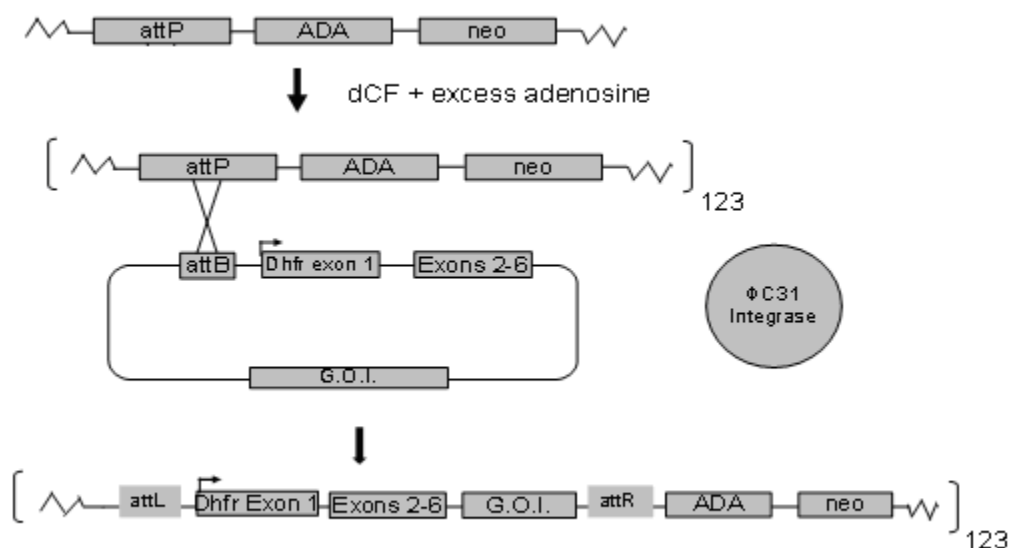


Figure 2 - Schematic of ϕ C31 integrase system. An *attP* site in the genome interacts with an *attB* site on the incoming *dhfr*/G.O.I. (*dhfr* marker gene with a gene of interest) plasmid in the presence of ϕ C31 integrase. 123 copies of this construct exist after amplification using dCF (deoxycoformycin) and excess adenosine. This recombination results in *dhfr* and the G.O.I. being integrated into the genome and the creation of hybrid *attL* and *attR* sites. An *attL* site is a hybrid between the upstream portion of *attP* and the downstream portion of *attB* since the DNA is cleaved in the center of this sequence. An *attR* site is a hybrid between the upstream portion of *attB* and the downstream portion of *attP*. *ADA* is an adenosine deaminase amplifiable marker and *neo* is a neomycin resistance gene used for selection. Ideally recombination takes place at all 123 sites to yield 123 copies of *dhfr* and the G.O.I.

We ultimately moved forward using the ϕ C31 integrase system due to our observations that *cre* recombinase was causing unwanted excision of our integrated transgenes to occur. The ϕ C31 enzyme, like many recombinases, is inefficient in performing recombination; however we describe methods below aimed at improving this recombination. In moving forward with the ϕ C31 integrase system to integrate a high number of copies into many locations, there are 3 parameters we focused on to optimize this process. First it is obvious that maximizing the number of *attP* sites for integration should increase the probability of targeting many transgenes into the genome. Alternatively, increasing

both the amount of attB sites and ϕ C31 integrase efficacy could drive the reaction *in vivo*. Here we describe how we increase each of these parameters which in turn may prove to drive this recombination into many attP sites.

Increasing attP concentration

Since an attP site is a specific 39 base pair sequence (Zhou et al., 2010), one cannot expect it to be naturally occurring in the genome. Previous studies have shown that pseudo attP sites exist in the CHO cell genome and can act as a substrate for ϕ C31 integrase with as little as 44% sequence similarity to the wild type attP site (Ginsburg and Calos, 2005). Despite this finding, site-specific recombination may not be as efficient in these sites compared to the wild type. An engineered cell line containing many wild type attP recognition sites may prove to be useful to maximizing SSR of a certain therapeutic gene. We describe two methods of increasing the amount of substrate attP sites in the CHO genome. The first method of increasing genomic attP substrate for ϕ C31 recombinase is pre-amplifying the cell using an amplifiable marker linked to attP sites. Adenosine deaminase also known as ADA was chosen for this purpose. Adenosine deaminase is an enzyme which deaminates adenosine in the cell. Excess adenosine is toxic to the cell and the action of this enzyme prevents cell death by reducing intracellular concentration of adenosine. Deoxycoformycin (dCF) is a specific inhibitor of the ADA enzyme. Cells that have undergone gene amplification at the site of the ADA gene (thus increasing its gene copy number) will be able to survive in the presence of excess adenosine and dCF (Kaufman et al., 1986). We have linked an attP site to ADA and amplified cells to contain 123 copies of attP sites (Figure 2).

A second method used to increase attP substrate was to create a concatemer of attP sites on one single vector. This was performed through subsequent digestion and ligation of attP sequence to create a concatemer of 64 attP sites (see Methods). This effectively increases the attP immediately following transfection of the construct into cells. Furthermore, by coupling this method with dCF amplification one can increase the number of attP recognition sites significantly. An overview of this process is shown in Figure 3. Ultimately, the copy number of genomic attP sites was maximized by combining the two aforementioned methods.

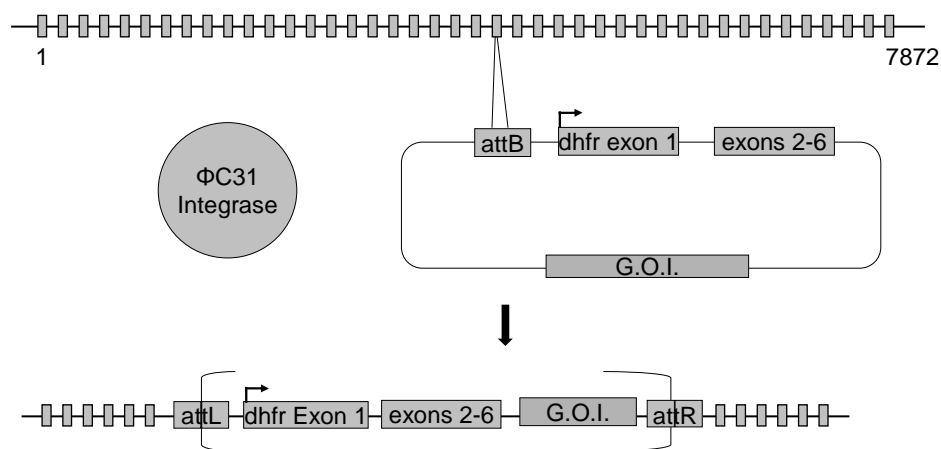


Figure 3 – Schematic of amplified attP genomic DNA. The small rectangles represent a concatemer of 64 attP sites. One of these sites will interact with the incoming attB site on the *dhfr*/G.O.I. plasmid to integrate this sequence into the genome in the presence of ϕ C31 integrase. The remaining attP sites remain intact for additional recombination reactions with other *dhfr*/G.O.I. plasmids

Increasing attB concentration

Another way to maximize recombination into these many attP sites is to increase the amount of incoming plasmid containing attB sites and the transgene of interest. Unlike the case for the attP site, we cannot create concatemers of attB sites on each incoming plasmid since upon recombination, original attP and attB sites will be within close proximity of each other. This may cause intrachromosomal DNA deletions between these sequences and excise the transgene of interest in a similar way as the cre/lox system. Two other methods were employed for this purpose.

The first method is to improve the delivery of this plasmid into host cells for recombination to occur. When co-transfecting the attB/G.O.I. plasmid with the plasmid coding for ϕ C31 integrase, any standard method of DNA delivery may be used. In an attempt to maximize the amount of attB we conducted experiments using both Lipofectamine 2000 and calcium phosphate DNA co-precipitation. Presumably, an effective delivery method will increase the intracellular concentration of the ϕ C31 expression plasmid also.

A more novel method that we employed is the use of a viral replication mechanism to exponentially increase the target attB/G.O.I. plasmid. This method has been successfully used in CHO cells expressing a polyoma large T-antigen which binds a polyoma origin of replication inserted into the attB/G.O.I. plasmid. This was shown to cause a high degree of replication (La Bella and Ozer, 1985).

Increasing ϕ C31 Integration Efficiency

As mentioned earlier, the amount of intracellular ϕ C31 expression vector may be increased by optimizing the DNA delivery method. It may also be possible to modify the

ϕ C31 integrase to maximize recombination efficiency into many genomic sites. There are two methods we utilize in an attempt to improve the recombination efficiency of the enzyme. The first is linking zinc finger proteins onto both the amino and carboxyl termini of the integrase and in parallel adding zinc finger binding sites (ZBS) flanking the attP and attB DNA. This strategy of using different zinc finger chimeras to improve protein binding has been studied extensively (Akopian et al., 2003). Another method we explored was to create a mutant library of ϕ C31 integrase enzymes and assess which performed recombination the best. In performing these experiments, an *in vitro* assay was developed to more quickly assess recombination efficiency.

Methods

Cre/Lox Experiments

Plasmids pDS1, pHygLox, and pCre were obtained from Shulian Shang. Plasmid pDS1 developed by Shulian Shang contains a *dhfr* minigene with two inverted and heterologous loxP sequences flanking exon 2 (L1 and 2L). L1 is the wild type recognition sequence and 2L is the inverted sequence of mutant lox site lox72 (Albert et al., 1995). The sequence of these 34 bp recognition sequences are as follows: L1: 5'-ATAACTTCGTATAGCATAACATTATACGAAGTTAT-3'; 2L: 5'-TACCGTTCGTATAATGTATGCTATACGAACGGTA-3'. Plasmid pHygLox contained a hygromycin resistance gene (*hyg*) flanked by these same two inverted loxP sequences. Plasmid pCre coded for the enzyme cre recombinase which catalyzed site-specific recombination. Linearized plasmid pDS1 was transfected into host DG44 cells by electroporation. The stable cell line containing this endogenous *dhfr* was subject to

MTX amplification to achieve high copy number of the transgene using the following stepwise concentrations: 10nM, 20nM, 50nM, 200nM, 1 μ M, 5 μ M, 20 μ M, 100 μ M, 200 μ M. Plasmids pHygLox and pCre were co-transfected using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. Cells that successfully received the *hyg* gene were selected with 200 μ g/mL hygromycin. Relative copy number of *hyg* and *dhfr* were measured using real-time PCR on an Applied BioSystems 700 thermocycler. The endogenous gene for adenine phosphoribosyltransferase (APRT) was used as an internal standard. The following primers were used: *Dhfr* exon 2: Left: 5'-CCATGTTAACGCAGTGTTTCTC-3'; Right: 5'-CCCACGGGAGACTTCGCACT-3'; *Hyg*: Left: 5'-TACATCAATGGGCGTGGATA-3'; Right: 5'-GATGTTGGCGACCTCGTATT-3'; *aprt*: Left: 5'-CCTGGAGGCAGGACTGTAAG-3'; Right: 5'-GGCAGCCTGGTAAGTGAAAG-3'.

Plasmid Construction (ϕ C31 experiments)

The adenosine deaminase gene driven by a CMV promoter was isolated by ApaLI digestion from plasmid pCMVADA (American Type Culture Collection). The insert was ligated into a unique ApaLI site in the vector pEGFP-C3 (Clontech: GenBank Accession # U57607) to create vector pADAGFP containing kanamycin/neomycin resistance and the adenosine deaminase gene. A synthetic attP double stranded oligomer was inserted in between unique XhoI and SalI restriction sites to create the vector pKAP1 containing a single attP site. The components of this plasmid are shown as the top linear construct depicted in Figure 2. Adjacent attP sites were added to the plasmid in a geometric fashion. Vector pKAP1 was digested with SalI and BamHI located downstream from SalI. The single attP site was digested from vector pKAP1 between XhoI and BamHI

sites and inserted into the previously plasmid digested with SalI and BamHI. This ligation was permitted due to XhoI and SalI having matching overhangs. The result was a plasmid containing two attP sites (pKAP2). The following process was repeated since all three restriction sites were reconstituted. The final plasmid contains 64 attP sites (pKAP64). This method is outlined in Figure 4.

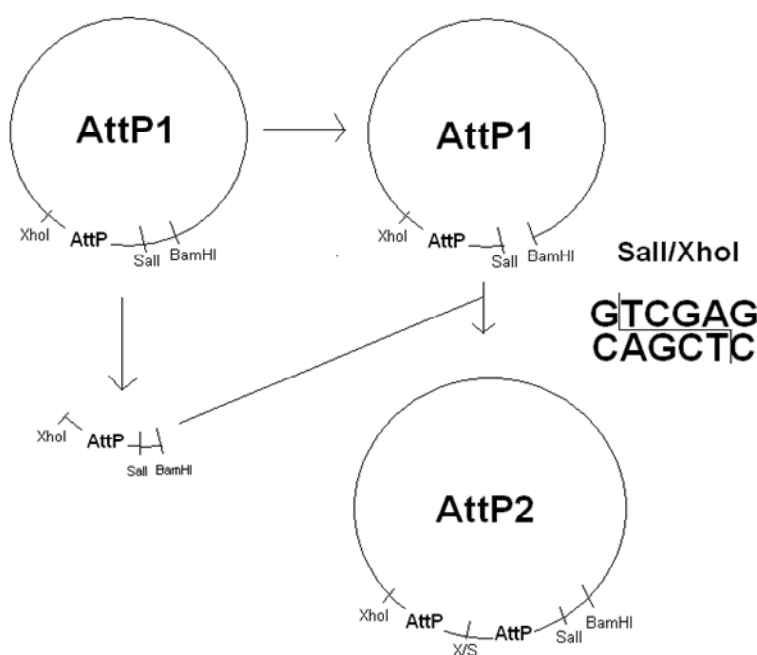


Figure 4 – Method for creating attP concatemers. A plasmid containing one attP site flanked by XhoI and SalI restriction sites and a BamHI restriction site downstream of SalI was used as the starting DNA. This plasmid was digested in two separate reactions. One reaction digested the plasmid with XhoI and BamHI to isolate the attP site. The other reaction digested the plasmid with SalI and BamHI. The attP insert was then ligated into the previously digested plasmid at the SalI and BamHI sticky ends. This is possible since SalI and XhoI share the same 4 base pair overhang upon digestion. The result is a plasmid containing two attP inserts with each of the three original restriction sites reconstituted for further digestion. This process is repeated to add attP sites in a geometric fashion (2 to 4 to 8, etc.). The final plasmid contained a concatemer of 64 attP sites and was used for subsequent transfection into host DG44 cells.

Plasmid pPGKPhiC31obpA was obtained from Addgene (#13795) (Raymond and Soriano, 2007) and contains the mouse-codon optimized ϕ C31 integrase gene driven by the PGK promoter. The plasmid pdchip11 contains a full minigene for dihydrofolate reductase (*dhfr*). A synthetic attB site with XmaI sticky ends was ligated into a unique XmaI restriction site upstream of the *dhfr* gene creating the plasmid pBE16. The sequences of the 39 bp attP site and 34 bp attB site used were 5'-CCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG-3' and 5'-GTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCG-3' respectively.

Transfection/Gene Amplification

DG44 cells were transfected with pKAP64 and selected for 3 weeks in 800 μ g/mL G418 (Geneticin®). Surviving colonies were pooled and subject to dCF amplification utilizing the ADA marker. Stepwise dCF amplification was performed by adding incremental amounts of dCF and selecting at each stage for two weeks. These final drug concentrations were 0.5, 1, 5, 10, 20, 50, and 100 μ M. Recombination reactions were performed on 1 million cells by co-transfection with pPGKPhiC31obpA and pBE16 using either Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions or calcium phosphate DNA co-precipitation using standard protocols from Murray (Murray, 1991). Cells were selected in purine and thymidine free medium ("PTF" medium: Alpha MEM without nucleosides and deoxynucleosides, Thermo Fisher) for 3 weeks. Colonies were fixed with 3.7% formaldehyde and stained with crystal violet.

φC31 Recombination Reactions

Recombination reactions were performed by co-transfecting plasmid pBE16 and plasmid pPhiC31obP in a 1:4 molar ratio using Lipofectamine 2000 or calcium phosphate DNA co-precipitation. After 48 hours cells were transferred to PTF medium at a dilution of 1:10. Cells were selected for three weeks for stable *dhfr* integration. Surviving colonies were pooled and grown to confluency. Genomic DNA was isolated from the resulting cells using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich).

Quantitative Real-Time PCR

Real-time quantitative PCR was performed using the Applied Biosystems Thermocycler 7300 and SYBR green as a reporter. The gene for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as an internal standard to determine copy number of genes in amplified cell lines using the $\Delta\Delta C_t$ method. To determine the number of attP sites in the genome the following primers were used:

Left: 5'-CGACAACCACTACCTGAGCA - 3'; Right: 5'-
CCAGTTGGGGCTCGAGAT-3'.

To determine the number of subsequently integrated *dhfr* minigenes into the genome by both random integration and ϕ C31 recombination the following primers were used: Left: 5' – CTGGCCAATGCTCAGGTACT – 3'; Right: 5' - GGACTTTGCTCCCAGCTAGA – 3'. The primers used for the *gapdh* internal standard were: Left: 5' – ACCCAGAAGACTGTGGATGG – 3' Right; 5' – CACATTGGGGGTAGGAACAC – 3'.

Self-Replicating Plasmids

The details of the protocols associated with modifying the attB/G.O.I. plasmid and current CHO host for self replication using polyoma Large T-antigen can be seen in the Master's thesis of Mrinalini Gururaj.

Improving ϕ C31 Integrase Efficiency

The details of the protocols associated with improving the ϕ C31 integrase using both zinc-finger chimeras and directed evolution are discussed in detail in the Master's thesis of Laurens Moore van Tienen.

Results

Cre/Lox site specific recombination

A cell line was successfully isolated containing a *dhfr* minigene with exon 2 flanked by inverted heterologous lox sites (L1 and 2L) by selection in PTF medium (from Shulian Shang). This cell line was taken through several rounds of MTX amplification up to 200 μ M and *dhfr* copy number was measured to be 115 with respect to the *aprt* standard (2 copies in DG44) and was named JC200. Next we sought to integrate the *hyg* resistance gene into many of the exon 2 locations via cre/lox mediated RMCE (Figure 1). After performing the recombination reaction (co-transfection of a cre recombinase expressing plasmid and the *hyg* plasmid flanked by lox sites) in cell line JC200 and selecting for hygromycin resistance, we isolated genomic DNA from the transfectant pool and measured *hyg* copy number along with exon 2 copy number to assess hygromycin gain against exon 2 copy loss. A large gain in *hyg* coupled with a large loss of exon 2 in

the genomic DNA would be indicative of successful integration of the *hyg* transgene into many loci via RMCE.

We observed that during the first few days of co-transfection of pCre (cre recombinase expressing plasmid) and pHygLox (plasmid containing the hygromycin gene flanked by lox sites) there was considerable cell death which is uncommon at this stage during transfection. After isolation of genomic DNA from transfectants we measured a very large decrease in exon 2 copy number coupled with only a very modest number of integrated *hyg* transgenes. Control transfections into JC200 cells were performed using only the pCre plasmid or only the pHygLox plasmid. We observed similar cell death during the first few days post-transfection using only pCre. We also measured similar *hyg* gene levels in transfectants only transfected with pHygLox which randomly integrated into the genome. These results were indicative of a massive excision of exon 2 and genomic DNA between amplified identical lox sites without the expected coupled massive integration of *hyg* into the previous locations of exon 2 via RMCE. This massive excision was seen from a sharp decrease in the real time PCR signal of the cells that have been subject to the co-transfection compared to the original cells. The cycle threshold (C_t) for detecting PCR product of exon 2 went from 16 in the original cells to 23 in the recombined cells. This corresponds to a 128 fold decrease in exon 2 presence which implies that almost all exon 2 sequences were excised.

In order to avoid the excision reaction of exon 2 we flanked that sequence with inverted loxP sites which prevents cre from catalyzing excision. Despite this, we were implementing this reaction in a cell containing 115 copies of this pair of loxP sites. Our

result was most likely due to intrachromosomal excision between alike L1 sites or alike 2L sites. Since identical sites (L1 or 2L) were initially amplified in the genome of these cells, there should be on the order of 10^5 base pairs between recognition sites (Hamlin, 1992). Hamlin found this base pair distance to be the size of a typical amplicon while studying gene amplification of the *dhfr* gene in CHO cells. We had assumed that this distance would be too large for cre recombinase to act in a *cis* manner for the excision of the sequence between two identical sites. Based on this result, this assumption was incorrect: this intrachromosomal recombination reaction apparently dominated over the *trans* integration reaction via RMCE. This is most likely due to identical sites being in close proximity because of the compact structure of DNA in chromatin.

φC31 Recombination in an Amplified attP Cell Line (DG44-PP)

After our results using the cre/lox recombination system indicated cre-mediated massive excision, we utilized the irreversible φC31 integrase system. This was an improvement over the cre/lox system since we used only attP sites to populate the host genome and excision cannot occur between two of these sites (Thyagarajan et al., 2001b).

Furthermore, upon integration using an attB bearing plasmid, hybrid attL and attR sites are formed which are not substrates for subsequent excision in the presence of φC31 integrase. AttB sites were not chosen to be the sequence present in many copies in the genome since it has been shown that pseudo attP sites already exist in the CHO genome, whereas pseudo attB do not (Chalberg et al., 2006). This study shows that as low as a 44% sequence similarity to the wild-type attP sequence can result in recombination with an attB site in the presence of φC31 integrase. After transfection of plasmid pKAP64

(see Methods) into DG44 cells we isolated the host cell line DG44-PP which contains at least one stably integrated concatemer of 64 adjacent attP sites. This cell line was then subject to dCF amplification for resistance up to 200 μ M and the genomic DNA from this pool of cells was isolated to measure the copy number of this construct. The average copy number of the pKAP64 construct from this pool was found to be 123. The calculated copy number of attP was then multiplied by 64 since each transfected cassette contained 64 attP sites yielding a total of 7872 attP sites available for subsequent targeted integration of a transgene comprised of the attB/G.O.I. plasmid. Once this cell line was established we performed co-transfections of both the attB/G.O.I. plasmid (pBE16) and the plasmid coding for ϕ C31 integrase (pPGKphiC31obpA).

First we observed the number of surviving colonies in purine and thymidine free (PTF) medium for both this co-transfection and a control transfection with plasmid pBE16 containing an attB site and the full *dhfr* minigene. Both are expected to result in colony formation since both random integration and ϕ C31 directed integration into one of many attP sites will both yield *dhfr* expression and survival in this medium. After transfection of 5×10^5 cells and splitting to a 100mm dish and selection in -PTF medium for 3 weeks the resulting stained colonies are shown in Figure 5 (A). We see a much higher density of colonies in the co-transfection with ϕ C31 indicative of more cells stably integrating and expressing *dhfr*. Based on this result we repeated these transfections and began selection immediately in PTF medium supplemented with 10nM and 20nM MTX mimicking the early levels of MTX amplification. We hypothesize that since more cells integrated *dhfr*, a fraction of them may have integrated multiple copies. Figure 5 (B and C) show clearly that many more colonies survived when ϕ C31 was included in the

transfection when challenged at these MTX levels. At the 20nM MTX level, there were no survivors at all for the control transfection experiment (only pBE16). MTX concentrations of 50nM and higher were tested similarly, however there were no apparent survivors in either case (not shown). Gene copy numbers at these levels of resistance usually correspond to only modest numbers (less than 10). Colony formation at much higher concentrations post-transfection would have been indicative of much higher copy numbers.

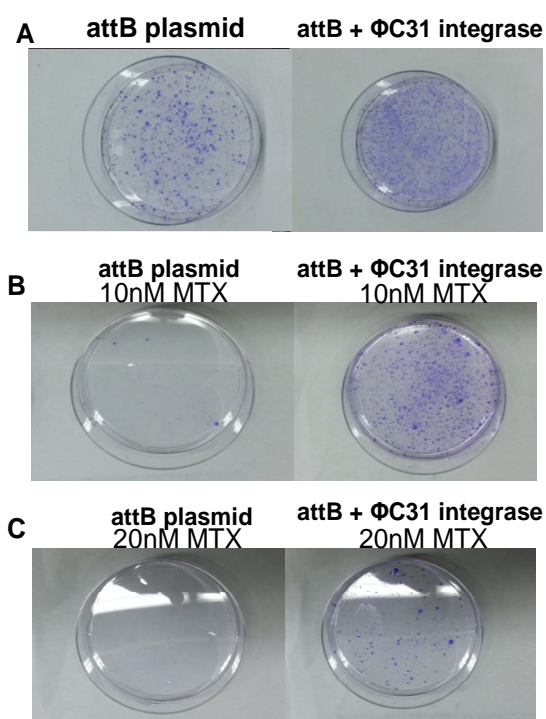


Figure 5 – Stained Transfectants with and without ϕ C31 integrase vector. A) DG44-PP cells transfected with either the incoming *dhfr*/G.O.I. plasmid named the attB plasmid for short or the attB plasmid with the vector coding for ϕ C31 integrase. Transfectants were selected for in PTF medium. B) DG44-PP cells were transfected in the same way as in A. Transfectants were selected for in PTF medium supplemented with 10nM MTX. C) DG44-PP cells were transfected in the same way as in A. Transfectants were selected for in PTF medium supplemented with 20nM MTX. In all cases cells were selected for 3 weeks and colonies fixed with 3.7% formaldehyde and stained with crystal violet.

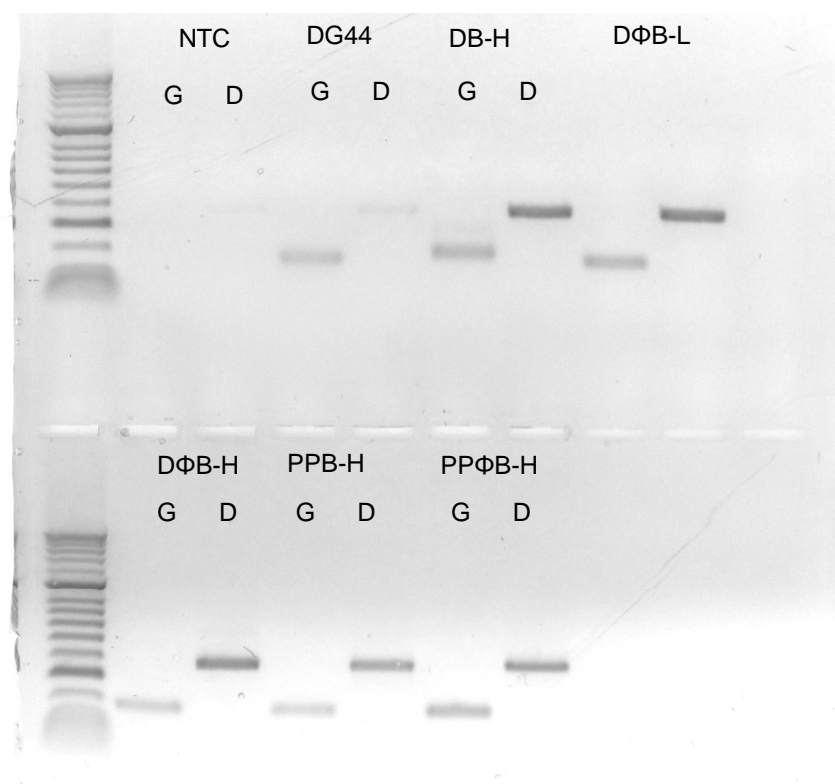


Figure 6 – Endpoint PCR of Genomic DNA from Transfectant Pools. PCR products for GAPDH (G) and *dhfr* (D) are shown using genomic DNA from different pooled transfectants as template. A no template control (NTC) is run in the first two lanes. DG44 is used as a negative control since it is deficient in both copies of *dhfr* and is run in the next two lanes. The next two lanes are from DG44 cells transfected with 4 μ g (high amount) of attB plasmid (DB-H). The next lanes are DG44 cells transfected with 1 μ g (low amount) of attB plasmid with ϕ C31 plasmid (D Φ B-L). The first two lanes in the bottom half of the gel are from DG44 cells transfected with 4 μ g of attB plasmid with ϕ C31 plasmid (D Φ B-H). The next two lanes are from DG44-PP cells transfected with 4 μ g of attB plasmid (PPB-H). The last two lanes are from DG44-PP cells transfected with 4 μ g attB plasmid with ϕ C31. The ladder used is HyperLadder II from BioLine.

We followed this experiment up with real time PCR using GAPDH as a standard to determine relative copy number. The values for each transfection are relative to DG44 cells transfected with 4 μ g pBE16 which mimics a standard method to randomly integrate a transgene. These results are shown in Figure 7.

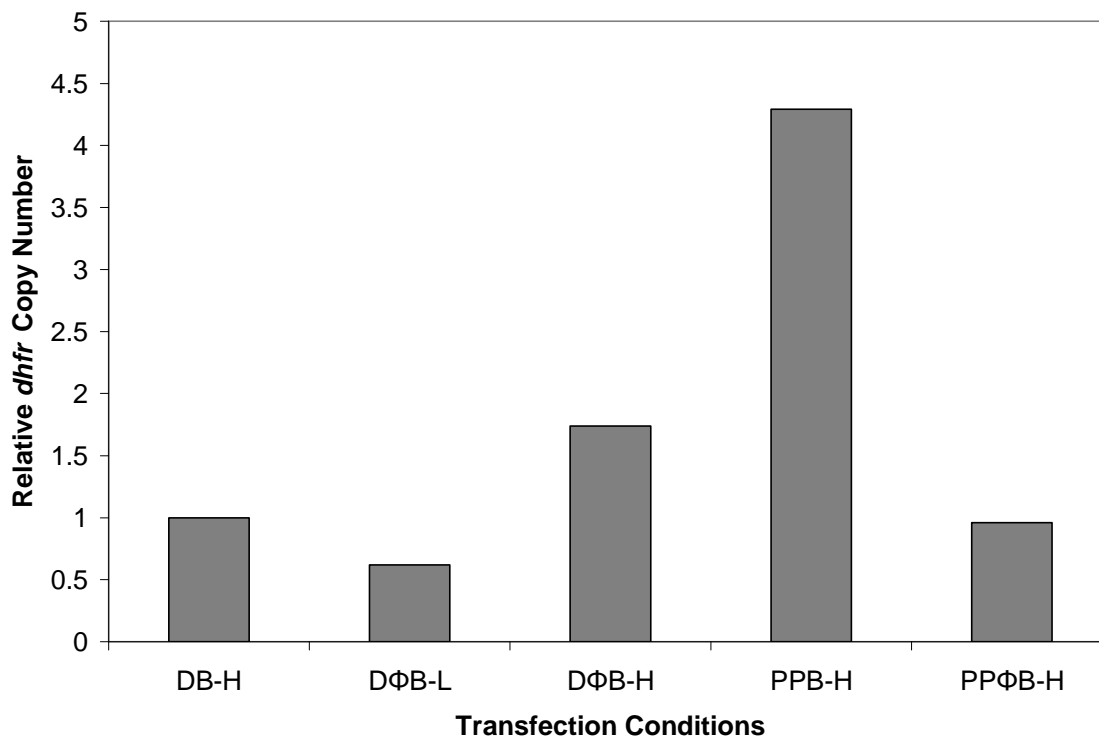


Figure 7 – Relative *dhfr* copy number measured using real time PCR. Copy numbers for integrated *dhfr* are shown relative to transfecting DG44 cells with the attB plasmid only (DB-H) shown as the first bar. Values were calculated using GAPDH as an internal standard. Abbreviations are explained in the legend to Figure 6.

Consistent with the gel, we observe similar copy numbers of *dhfr* relative to the control, and all levels are very modest. Despite more survivors with modest increases in *dhfr* copy number as seen in Figure 6, these pools of cells showed no improvement with respect to the controls. We surmised two ways to improve this system. The first was to change transfection methods since a different way of delivering DNA may be more amenable to site-specific recombination into many sites even if it hasn't proven to be superior for random integration. The second modification we made was to isolate individual clones instead of measuring the pool of cells to see if there was a high variation in *dhfr* copy number. Perhaps we would be able to find one outlying clone containing many integrated *dhfr* copies.

We repeated the experiments using calcium phosphate DNA co-precipitation. This method has already been proven to be able to successfully deliver large amounts of DNA into a variety of cell hosts (Murray, 1991). We transfected pBE16 into DG44-PP cells both with (site-specific integration) and without (random integration) the ϕ C31 integrase gene. We once again measured integrated *dhfr* copy number using GAPDH as an internal standard relative to UA41 cells which have been previously isolated and contain only 1 endogenous copy of *dhfr*. The results are shown in Table 1. The most important result is that calcium phosphate DNA co-precipitation yields about a 5 fold greater efficiency of site-specific recombination than Lipofectamine 2000. This occurs despite the result that Lipofectamine 2000 is more efficient for delivering DNA that is randomly integrated into the genome.

Recombination	Reagent	Copy no.
Site-specific	Lipofectamine	5.2
Site-specific	Ca ₃ (PO ₄) ₂	25.4
Random	Lipofectamine	19.4
Random	Ca ₃ (PO ₄) ₂	1.1

Table 1 – *Dhfr* copy number across different transfection conditions. Copy number was measured using real-time PCR and GAPDH as an internal standard. Copy numbers are absolute relative to the signal provided by UA41 cells which contain one allele of *dhfr*. The highest value obtained using calcium phosphate DNA co-precipitation (Ca₃(PO₄)₂) and transfecting with ϕ C31 integrase vector (site-specific) is in bold.

After taking measurements on transfectant pools resulting from the calcium phosphate method, we isolated individual colonies that survived selection in PTF medium supplemented with 20nM MTX. This ensured that we were isolating only those that were

able to overcome the MTX by virtue of an increased copy number of *dhfr*. We measured *dhfr* copy number for 6 individual clones and two pools of transfectants relative to the control transfection of our cell line without ϕ C31 integrase. These values are shown in Figure 8. Clone 1 had the highest relative copy number of about 5 which was an improvement over the average copy number of the two transfectant pools. Although this is an improvement over random transfection, it is still very low considering the number of available attP sites in the genome of DG44-PP cells.

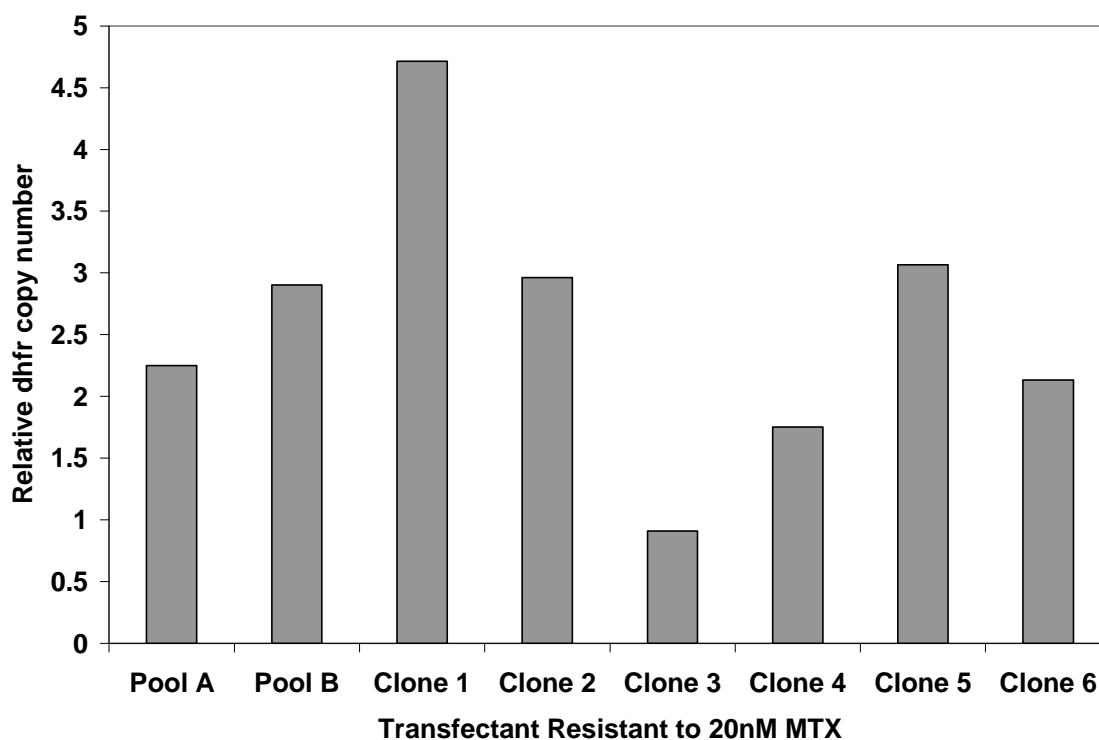


Figure 8 – Relative *dhfr* copy number using calcium phosphate DNA co-precipitation. Measurements using real time PCR on two different pools and 6 clones of transfectants initially resistant to 20nM MTX are shown. Values are relative to a standard transfection of DG44 cells with only the attB plasmid. GAPDH was used as an internal standard.

Polyoma Large T-Antigen Mediated attB Plasmid Replication

Our previous results show that by increasing the number of genomic substrates (attP sites) to a very high level we were unable to achieve a large improvement over randomly transfecting transgenes into host CHO cells in terms of the number of transgenes integrated. Therefore, using the system as currently developed, one would still need to perform MTX amplification to further amplify the *dhfr* transgene along with the G.O.I. Since the DG44-PP cell line has many attP copies in the genome, it is possible that the limiting factor in promoting many integration events is the number of available attB sites. To increase the number of attB substrates we added an origin of replication for polyoma virus to the pBE16 plasmid which yielded plasmid pBE16ori. Concurrently, we developed a cell line that stably expresses the large-T antigen from polyoma virus driven by a CMV promoter (DG44-PPTA). Only polyoma large-T antigen needs to be expressed to allow replication of polyoma DNA in mammalian cells. Previous studies have shown that this system has successfully replicated an incoming plasmid about 10000 fold relative to plasmid levels without the large-T antigen in another CHO cell host (La Bella and Ozer, 1985). A schematic of our experiments is shown in Figure 9.

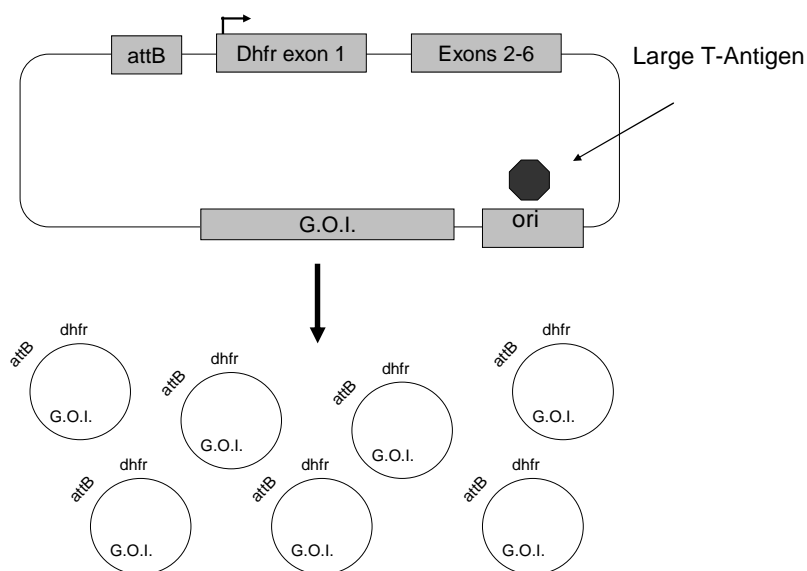


Figure 9 – Schematic of Replicating Plasmid Experiments. The *dhfr*/G.O.I. bearing plasmid now contains a polyoma origin of replication (*ori*). This DNA is transfected into DG44-PPTA cells which contain many *attP* sites and polyoma large T-antigen driven by a CMV promoter. The large T-antigen (black octagon) will bind the *ori* causing replication of the *dhfr*/G.O.I. bearing plasmid which increases the *attB* substrate available for recombination into genomic *attP* sites.

We found that this cell line has successfully integrated the large T-antigen gene using PCR and confirmation of the correct sized band on an agarose gel. We also found that it is expressing its mRNA by performing RT-PCR and confirmation of the correct sized band compared to a negative control where the reverse transcription step was not performed. We measured the amount of *attB* plasmid 48 hours after transfection into the DG44-PPTA cell line compared to the same transfection in our original DG44-PP cells that does not express large T-antigen. Using a Hirt extraction (Ziegler et al., 2004) which isolates only circular plasmid DNA from cells we measured the amount of *attB* plasmid and its degree of replication using both a *DpnI* assay and real-time PCR. The *DpnI* restriction endonuclease only cleaves methylated recognition sites, which are synthesized

during growth in bacteria, therefore the newly replicated plasmid will remain undigested. The result from this assay was inconclusive since there was a high background of genomic DNA when performing the Hirt extraction. This made it difficult to visualize the digested and undigested plasmid on an agarose gel. Using real-time PCR we determined there was twice as much pBE16ori 48 hours post transfection than the pBE16 plasmid in a parallel transfection. This difference is not likely to substantially increase recombination of a transgene into the many genomic attP sites. It is possible that the cellular environment of our cells differs from the cell line used previously to obtain such a high degree of plasmid replication using this system. Detailed figures pertaining to these experiments can be found in the Master's thesis of Mrinalini Gururaj.

Engineered Integrase

The major bottleneck in trying to achieve many integration events may be the efficacy of the ϕ C31 enzyme. Although we are currently using a vector which strongly expresses the enzyme in the cell, its binding and subsequent catalysis of recombination may be improved through generating mutant enzymes. The first method we employed was creating a protein chimera of ϕ C31 integrase attached to 2 zinc finger binding domains (ZFBDs) via various protein linkers. Concurrently we developed vectors containing attP and attB sites flanked by zinc finger binding sequences with various size sequence spacers between the attP or attB site and the zinc finger binding sequence. The goal was to rationally find the best combination of ϕ C31 chimera and attP/attB sequences which promotes maximum recombination efficiency. A schematic of this process from the Master's thesis of Laurens Moore van Tienen is shown in Figure 10.

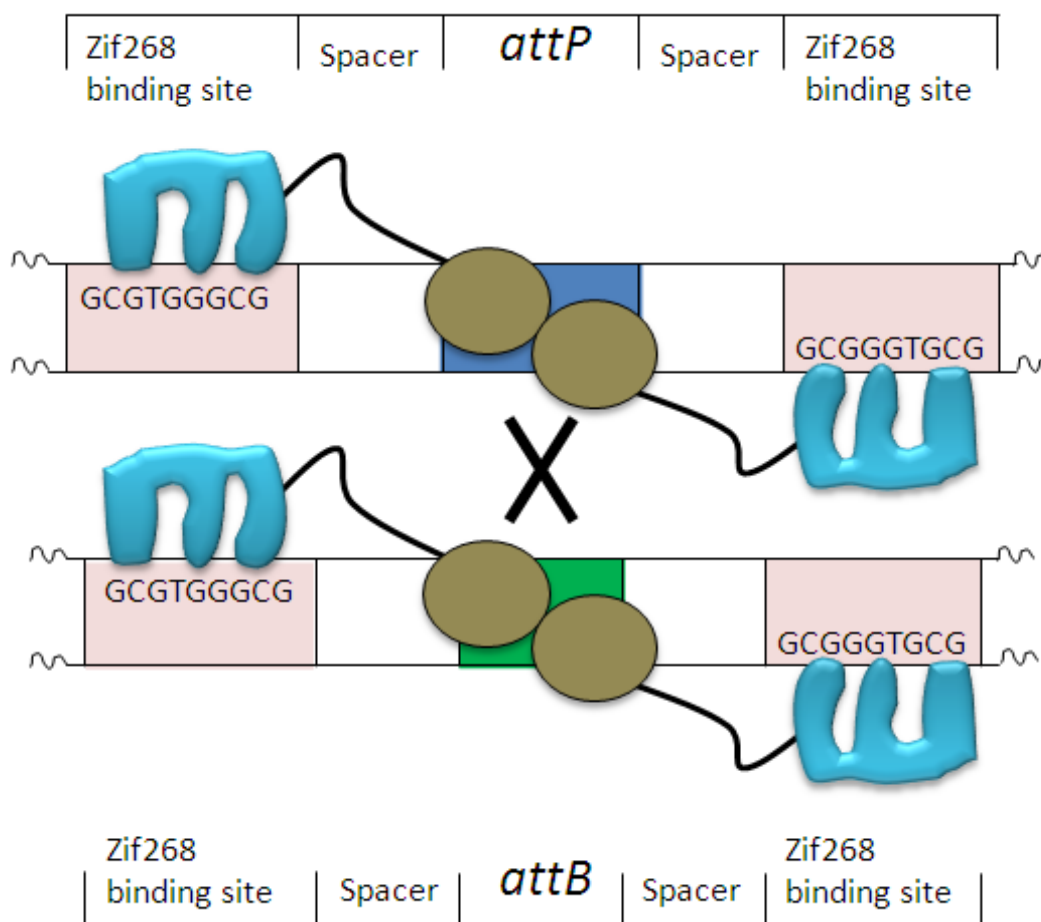


Figure 10 – Schematic of ϕ C31 integrase/zinc finger chimera. AttP (dark blue) and attB (green) sites are flanked by variable spacer sequences and zinc finger binding sites. The ϕ C31 integrase dimer (brown) is attached to a protein linker (black line) and a three finger zinc finger known as Zif268 (light blue) on either side. Figure is from the Master's thesis of Laurens Moore van Tienen.

Testing these constructs *in vivo* was very time consuming due to the necessity to permanently transfect attP constructs, followed by the transfection of ϕ C31 chimeras and attB constructs to assess colony formation by selecting for DHFR activity. Due to this constraint we developed an *in vitro* assay which relies on transcription/translation (TnT) of the enzyme *in vitro*. This involves transcribing and translating ϕ C31 in solution and adding an attP and an attB construct to allow recombination to occur. The degree of

recombination is then assessed by PCR using specific primers that only prime recombination products. An overview of this process is shown in Figure 11.

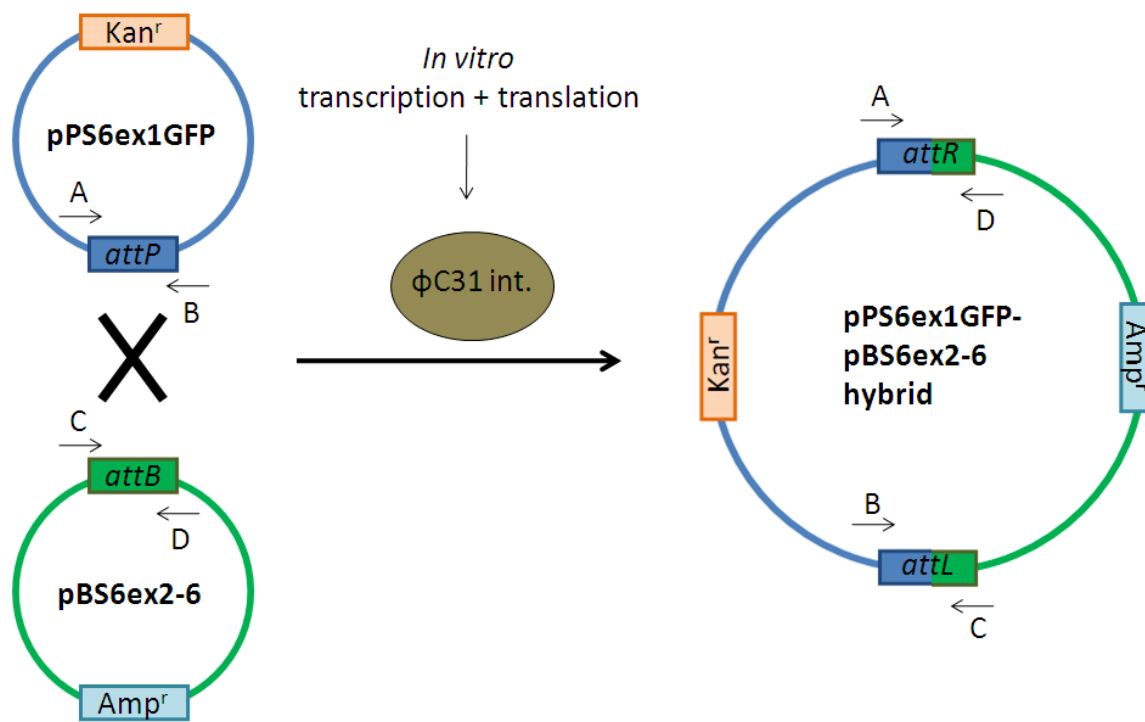


Figure 11 – *In vitro* Recombination Assay. A kanamycin resistant plasmid (pPS6ex1GFP, blue) containing an attP site is mixed with an ampicillin resistant plasmid (pBS6ex2-6, green) containing an attB site. The ϕ C31 vector is also introduced to the mixture and transcribed and translated using an *in vitro* transcription/translation kit (TnT) to produce the enzyme (brown). Successful recombination will result in the creation of a larger hybrid plasmid (pPS6ex1GFP-pBS6ex2-6 hybrid, blue and green). Primer pairs A and D or B and C are used to create PCR products that will only be synthesized if recombination occurs. Using primer A with B or C with D will result in a product with both original plasmids present. The newly synthesized plasmid will confer resistance on agar plates supplemented with both kanamycin and ampicillin. Figure is from the Master's thesis of Laurens Moore van Tienen.

Five different spacer lengths were tested ranging from 5 to 8 base pairs. Nine different protein linkers were also tested of varying lengths and amino acid compositions. Pure glycine linkers of lengths 4, 7, 11, 13, and 16 were tested along with 10 amino acid linkers of alternating serine and glycine, or serine and arginine. A 10 amino acid linker of

only arginine was also tested along with a linker known as L6 which was previously shown to improve site-specific recombination. All of these perturbations were tested to ensure that a wide range of lengths and linker flexibilities were assessed. Of all the combinations of spacer lengths and linkers tested, only one gave a significant improvement in recombination efficiency. The spacer length of 5 base pairs showed the most improvement in recombination regardless of whether the wild-type integrase or an integrase linked to a ZFBD was used. The major result from this part of the work was that despite the presence of different ZFBDs and zinc finger binding sequences with various linkers and spacers, one spacer sequence on both sides of the attP and attB was the best option. This spacer sequence improved ϕ C31 recombination 23 fold over the wild-type attP and attB sequences we used in previous experiments using the *in vitro* assay. Notably, this result was also seen using an *in vivo* assay which showed this improvement even with wild-type ϕ C31 without ZFBDs. This assay involved recombining an attB bearing plasmid with this specific spacer sequence flanking either end into a similar genomic attP site with the same flanks in the genome of CHO cells. Successful recombination was assayed by counting surviving colonies in PTF medium since the *dhfr* gene was reconstituted upon recombination. This is indicative that sequence context alone surrounding the attP and attB sites can greatly affect recombination efficiency both *in vivo* and *in vitro*. Interestingly, this 5 base pair long spacer contained a G on the 5' side of the attP and attB site and a C on the 3' side. This corresponds to what naturally flanks the original attP site in phage phiC31. The bacterial attB site has a G as the 5' flank as well. Perhaps the minimal 34bp and 39bp attB and

attP sequences previously determined are far from optimal for recombination (Groth et al., 2000)

A second method for improving ϕ C31 integrase is using directed evolution. We developed a method that allows for the generation of a large mutant ϕ C31 library and several iterations of selecting improved mutants by linking genotype to phenotype. Each mutant is kept separate from the other by using *in vitro* compartmentalization in a water/oil emulsion (Ghadessy and Holliger, 2004). The details of this method are outlined in the Master's thesis of Laurens Moore van Tienen. Figure 12 shows a schematic of this process. This process has not been yet implemented for our purposes but our future plans include using this method for isolating mutant ϕ C31 integrases with very high recombination efficiencies.

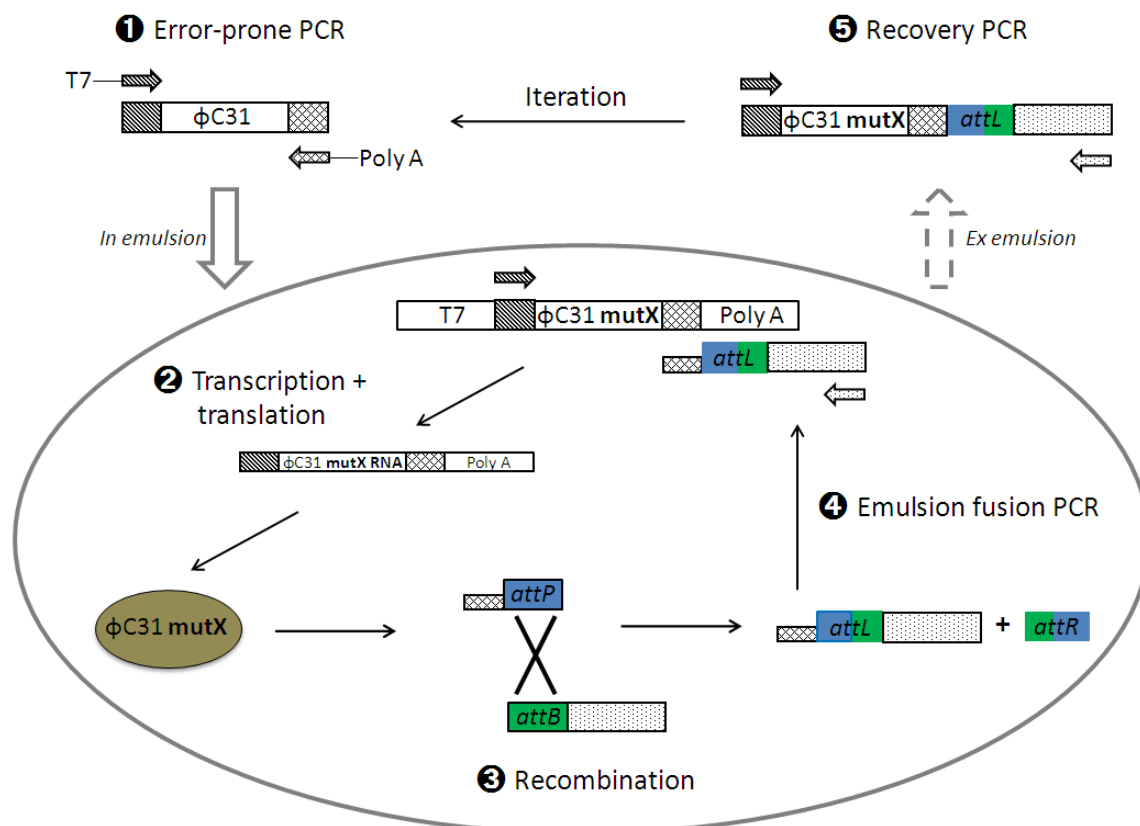


Figure 12 – *In vitro* compartmentalization (IVC) Process for Isolating Efficient ϕ C31 enzymes. Step 1: Error-prone PCR is used to generate many mutants of ϕ C31. Each mutant is introduced into its own micro-droplet created from an oil/water emulsion at a limiting amount. Step 2: Each mutant is transcribed and translated into ϕ C31 enzyme. Step 3: Each mutant enzyme performs recombination on the introduced *attP* and *attB* substrates to form products containing *attL* and *attR*. Step 4: PCR is performed in the emulsion and is successful to the degree that *attL* product is formed since this product is used as a primer. Step 5: The emulsion is broken and successful mutants will have had their phenotype of efficient recombination linked back to its genotype to be recovered by PCR and passed onto another iteration of the process. Figure is from the Master's thesis of Laurens Moore van Tienen.

Discussion

Our overall goal was integrating a transgene into our CHO host into as many locations as possible in one single transfection in order to abrogate the need for MTX amplification.

The methods we employed to achieve this goal utilized site-specific recombination to target transgenes to many pre-established locations in the host genome. The ϕ C31

recombination system proved to be superior for this purpose due to its irreversibility for the integration reaction. The three parameters we increased were increasing the number of attP sites in the genome through dCF gene amplification/ligation of attP concatemers, increasing the number of attB/G.O.I. incoming plasmids through viral replication, and improving the action of the ϕ C31 integrase through modifying both the protein and the attP/attB binding sequences.

Despite successfully isolating cell line DG44-PP containing over 7000 attP sites for potential targeting of a transgene, we were only able to achieve a 5 fold improvement over using random transfection of the same plasmid (pBE16) under the same conditions for an approximate copy number of 25. This may speed up development of high copy number cell lines by allowing one to skip the first 2 steps (10nM and 20nM) of MTX amplification since cells will grow with up to 25 copies after a single transfection step. This is still a very modest increase and several more rounds of MTX amplification would be needed. Increasing attP sites alone increases the probability of isolating clones resistant to 20nM, however it does not allow for isolation of a very high copy number (several hundred) clone resistant to the high levels of MTX. It is possible that by performing a large enough transfection, one may isolate such a clone.

Based on this result we concluded that at this point another bottleneck had been created either in the amount of available incoming *dhfr*/G.O.I. plasmid or the efficacy of the ϕ C31 integrase.

When we attempted to mimic previous work (La Bella and Ozer, 1985) that used viral replication to increase the amount of incoming plasmid 10000 fold we saw a negligible increase in our system. As mentioned earlier, perhaps there are proteins in our host cell

that inhibit the action of the large T-antigen for binding to its origin of replication on our plasmid. Another possible reason for this negligible increase is that other proteins are competing for this binding sequence on the plasmid hindering its replication. It is also possible that the expressed large T-antigen is degraded before it is able to perform its function. Adding mRNA stabilizing elements to its DNA sequence may show improvement.

Our results shown in Table 1 indicate that calcium phosphate DNA co-precipitation is a superior transfection method over Lipofectamine 2000. This may be due to its ability to physically deliver more *dhfr*/G.O.I. plasmid into the nucleus for recombination, or because it protects the ϕ C31 DNA from degradation allowing it to be expressed. Future work may be done to modify the timing of the delivery of both the ϕ C31 plasmid and the *dhfr*/G.O.I. plasmid into host cells. For example, it is possible that ϕ C31 needs more time to be expressed before introduction of the *dhfr*/G.O.I. plasmid. Perhaps transfection of ϕ C31 first, followed by introduction of the *dhfr*/G.O.I. at later various times may improve recombination into more genomic attP sites.

The ϕ C31 enzyme has been shown to be inefficient in various other cell lines (Keravala et al., 2006; Thyagarajan et al., 2001b). At this point it became clear we should focus our attention to modifying the enzyme itself to improve recombination efficiency. This is the most likely bottleneck at this point in integrating transgenes into a large number of genomic recombination sites. Work done by Laurens Moore van Tienen (Master's student) showed that including a specific sequence flanking the attP and attB site could drastically improve recombination. This leads us to continue investigating this phenomenon by using a high-throughput method of isolating attP/attB flanking sequences

that promote very efficient recombination. We plan to develop a high throughput *in vitro* recombination assay for this purpose. One plan is to test all random 5-mers flanking the attP and attB sequences for efficient recombination. This process is described in more detail in Chapter 6. Other options include generating a mutant library of 5' and 3' flanks of the original attP and attB site described by Calos et. al. on two separate DNA molecules. We can then use high-throughput sequencing to identify mutant flanks that promote recombination. Another option is to use SELEX (Systematic Evolution of Ligands by Exponential Enrichment). This would use PCR to reproduce efficiently recombined sequences to enrich the subsequent population for further testing. This may be iterated for several rounds before we isolate molecules that promote high levels of recombinations. These experiments may ultimately lead to a small number of flank sequences that greatly improve recombination efficiency.

Our work has modestly alleviated the time bottleneck associated with isolating high gene copy number cell lines coding for a specific therapeutically or diagnostically important protein. We achieved a modest improvement in gene copy number isolated from a single transfection through the use of the ϕ C31 recombination system and increasing its recognition substrates (attP and attB). This result achieved led us to pursue modification of the ϕ C31 enzyme which ultimately resulted in the discovery that attP and attB sequence context is important for recombination efficiency. This work has paved the way for our future endeavor to isolate attP and attB flanking sequences that greatly improve this process. This may ultimately allow a very high number of transgenes to be integrated into our host cells in one single transfection protocol, thus further alleviating this developmental time bottleneck.

References

- Akopian, A., He, J., Boocock, M.R., Stark, W.M., 2003. Chimeric recombinases with designed DNA sequence recognition. *Proceedings of the National Academy of Sciences of the United States of America*. 100, 8688-8691.
- Albert, H., Dale, E.C., Lee, E., Ow, D.W., 1995. Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *The Plant journal : for cell and molecular biology*. 7, 649-659.
- Calos, M.P., 2006. The phiC31 integrase system for gene therapy. *Current gene therapy*. 6, 633-645.
- Chalberg, T.W., Portlock, J.L., Olivares, E.C., Thyagarajan, B., Kirby, P.J., Hillman, R.T., Hoelters, J., Calos, M.P., 2006. Integration specificity of phage phiC31 integrase in the human genome. *Journal of molecular biology*. 357, 28-48.
- Ghadessy, F.J., Holliger, P., 2004. A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein engineering, design & selection : PEDS*. 17, 201-204.
- Ginsburg, D.S., Calos, M.P., 2005. Site-specific integration with phiC31 integrase for prolonged expression of therapeutic genes. *Advances in genetics*. 54, 179-187.
- Groth, A.C., Olivares, E.C., Thyagarajan, B., Calos, M.P., 2000. A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Academy of Sciences of the United States of America*. 97, 5995-6000.
- Hamlin, J.L., 1992. Amplification of the dihydrofolate reductase gene in methotrexate-resistant Chinese hamster cells. *Mutat. Res*. 276, 179-187.
- Hoess, R.H., Abremski, K., 1985. Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *Journal of molecular biology*. 181, 351-362.
- Kaufman, R.J., Murtha, P., Ingolia, D.E., Yeung, C.Y., Kellems, R.E., 1986. Selection and amplification of heterologous genes encoding adenosine deaminase in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3136-3140.
- Keravala, A., Portlock, J.L., Nash, J.A., Vitrant, D.G., Robbins, P.D., Calos, M.P., 2006. PhiC31 integrase mediates integration in cultured synovial cells and enhances gene expression in rabbit joints. *The journal of gene medicine*. 8, 1008-1017.
- La Bella, F., Ozer, H.L., 1985. Differential replication of SV40 and polyoma DNAs in Chinese hamster ovary cells. *Virus research*. 2, 329-344.
- Murray, E.J., (1991) *Gene Transfer and Expression Protocols*. Springer.

- Raymond, C.S., Soriano, P., 2007. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PloS one*. 2, e162.
- Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., Calos, M.P., 2001a. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Molecular and cellular biology*. 21, 3926-3934.
- Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., Calos, M.P., 2001b. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell Biol*. 21, 3926-3934.
- Zhou, H., Liu, Z.G., Sun, Z.W., Huang, Y., Yu, W.Y., 2010. Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. *J. Biotechnol*. 147, 122-129.
- Ziegler, K., Bui, T., Frisque, R.J., Grandinetti, A., Nerurkar, V.R., 2004. A rapid in vitro polyomavirus DNA replication assay. *Journal of virological methods*. 122, 123-127.

Chapter 6

Conclusions

It is evident that biologics and in particular Mab's have contributed significantly to treating various types of diseases. Our overall goal was to improve the process of isolating high producing Mab cell lines by either decreasing the time needed to amplify the Mab gene to many copies, or integrating many copies of a Mab gene in one step.

We have isolated a CHO clone which is capable of fast, reliable, and reproducible transgene amplification. This was shown to be an improvement over random integration of a transgene followed by gene amplification. This cell line can be utilized to direct any transgene of choice (i.e. for a Mab) for fast, reliable gene amplification. We have also successfully developed a method using fluctuation analysis which can measure gene amplification rates in any host cell being used for recombinant protein production. The method of measuring amplification rates of several clones carrying a ϕ C31 SSR site may prove valuable in the isolation of a potential production host. Despite the initial time investment of this process, it may save a considerable amount of development time once a highly amplifiable clone is isolated.

In an effort towards removing the need to go through lengthy gene amplification we isolated a potential host CHO clone containing over 7000 copies of the ϕ C31 SSR recognition sequence (attP sites) in the genome for potential integration of many copies of a transgene. Despite the isolation of this clone, we were only able to successfully integrate 25 copies of a transgene into these sites, indicative that there is another bottleneck hindering many genomic integration events occurring. This result led us to modifying both the ϕ C31 enzyme and the recognition sites in which it initially binds to

improve genomic transgene integration. We discovered that mutating the flanking DNA sequence around the ϕ C31 recognition sites improved recombination considerably using both an *in vitro* and *in vivo* assay.

Based on our last result we plan to conduct experiments which will select sequences flanking both the attP and attB upstream and downstream that promote the most efficient site-specific recombination. Isolation of mutant flanking sequences and incorporating them in CHO host cell lines may alleviate this recombination bottleneck and allow more transgene copies to be integrated into many genomic sites. We intend to perform this experiment by using an *in vitro* assay which can assess how well ϕ C31 recombination occurs between two molecules (one with an attP site and one with an attB site). Our first step will be to create two libraries of molecules. The first library will be circular DNA molecules containing the standard wild-type ϕ C31 attB site flanked by all single base pair substitutions (SBS) for the 5 base pairs on either side of the this sequence. The second library will be linear DNA molecules containing the wild-type ϕ C31 attP site flanked by SBS for the 5 base pairs on either side of this sequence. These molecules will be mixed and incubated with recombinant ϕ C31 recombinase produced in *E. coli* or by *in vitro* transcription and translation. Successful recombination of one circular attB containing molecule and one linear attP containing molecule will result in a longer linear molecule containing an attL sequence (upstream portion of attP site and downstream portion of attB site), and an attR sequence (upstream portion of attB site and downstream portion of attP site). These recombination products will be isolated by gel electrophoresis on the

basis of their size. We will use high-throughput sequencing to determine which mutant flank combinations gave rise to the most recombinations using specific primers.

We may iterate this process by testing flanking sequence beyond 5 base pairs depending on the obtained results.

Mutant flank combinations which are present in high amounts will then be tested *in vivo* in our CHO host to determine if recombination of a transgene into many genomic locations is improved. The best flank sequences for the attP site can be stably transfected into host CHO cells and amplified to many copies similar to the procedure described in chapter 5. The best flank sequences for the attB site can then be co-transfected with the ϕ C31 vector and recombination into many attP sites can be assessed.

We have addressed the issue of the long development time associated with isolating high producing CHO clones for recombinant protein production. First, we have isolated a clone with the capability of integrating any transgene for a therapeutic protein at a location prone to fast gene amplification. This gives users the ability to isolate a high gene copy number clone in less time than they would if using the standard method of random transfection followed by gene amplification. Users may also utilize our method of measuring gene amplification rates for other mammalian cell hosts by implementing fluctuation analysis. This would allow for the isolation of any host capable of fast gene amplification. Future experiments we propose have the potential to improve our system by removing the need for any gene amplification. Upon isolation of attP and attB flank mutants and implementing them in a CHO host with thousands of sites for recombination, we may even further reduce development time of obtaining high producing clones for end

users. The result of this work, coupled with other developments made in the field of biopharmaceutical development have the ability to continue to decrease the time needed to get biologic drugs to market.