Initial considerations before designing a promoter construct.

David Gould – Biochemical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London EC1M 6BQ, UK. d.j.gould@qmul.ac.uk

Running Title: Initial considerations before designing a promoter construct.

Summary

Before designing a synthetic promoter it can be helpful to think about its final application. Is the study purely an *in vitro* exercise in monitoring short term promoter activity from an episomal vector or does the promoter eventually need to be permanently active and be integrated into the genome or perhaps even to function *in vivo*. The final application will have a bearing on promoter design and vector of choice from the start of the study. In this chapter I highlight some of the vector attributes to consider and features that should be thought about.

Introduction.

Promoters are DNA sequences that determine where transcription of a gene by RNA polymerase is initiated. In general terms the types of promoters that have been utilised are constitutive ubiquitous, tissue specific, regulated promoters (activated by endogenous transcription factors) and inducible promoters (transactivators that are ligand/light responsive). With the increasing demand for production of recombinant proteins especially as pharmaceutical agents, advances in gene therapy and the progress of synthetic biology there

are increasing applications for promoters in biological systems and the demands on their function vary from one application to the next. There are a number of different reasons for assembling synthetic promoters which include the generation of compact promoters, improving promoter function or designing promoters with additional functionality. In each instance it is important to consider the final use of the promoter so that the promoter is constructed with the optimal subunits and is cloned in the correct context with other vector elements.

Vector constraints.

It is important to consider any limitation on promoter size that will be imposed by the vector of choice for gene delivery. Adeno-associated viral vectors have a capacity of approximately 4.8 Kb (1) and the capacity of self-complementary rAAV is 2.4 Kb (2), so compact synthetic promoters will be required, other viral vectors also have size limitations and you should be aware of these before progressing.

Core promoter

Core promoters are a basic feature of all synthetic promoters, they contribute to synthetic promoter strength and can have tissue-specific activity. Often they are a simple TATA box but can contain other elements, the methods to determine their activity are detailed in Chapter 7.

Transcription factor binding sites

In this volume several authors describe the assembly of synthetic promoters through the annealing and ligation of oligonucleotides which result in the alignment of desired transcription factor binding sites (TFBS) upstream of the core promoter. The sequence bound

by transcription factors is conserved but will vary to some extent between binding sites. When different binding sequences are aligned they can be used to generate a sequence logo which is computed based on Shannon entropy {Schneider, 1990 #3659}. In these logos the height of the individual letters represents the frequency of a nucleotide in the particular position of the alignment. The sequence logo will be informative when selecting nucleotides at each position within a TFBS but examination of the literature for functional data is also recommended. In many instances spaces will be introduced between TFBS and it may be important to screen the resulting spacing sequences using a database such as Transfac® to ensure that unwanted TFBS are not inadvertently introduced.

Reporter genes.

There are a variety of reporter genes that are utilised in order to monitor synthetic promoter activity and these are discussed in Chapters 4, 5 and 6. Reporter proteins that are secreted from cells have the advantage that promoter activity can be readily monitored from the same transfected cells over time, whilst intracellular reporters generally require termination of the experiment in order to determine promoter activity. Fluorescent reporters have the added advantage that no substrate is required to measure their activity. Most reporter proteins are foreign when delivered *in vivo* and expressing cells will be targeted by the immune system in immunocompetent animals. The only exception is secreted alkaline phosphatase, an endogenous protein that can be expressed long-term in an immunocompetent host (3).

Cloning approach

Traditionally cloning will utilise restriction enzymes, in which case it is important to ensure that cloning sites are included with the 5' and 3' termini of a synthetic promoter and are not inadvertently introduced internally within the synthetic promoter sequences during their

construction. If the aim is to use the same synthetic promoter in several different vectors it maybe more efficient to use a recombination cloning system. Firstly, synthetic promoters will need to be generated so that they are flanked by the appropriate sequence recognised by the specific recombinase. The synthetic promoter is then incubated with the site-specific recombinase and the supplied holding vector. It is then easy to shuffle the promoter into a variety of 'destination'/'acceptor' vectors by recombination cloning without the need to use restriction enzymes.

The recent use of promoters in complex synthetic transcriptional circuits requires careful planning as it is necessary to incorporate unique annealing sequences into each subunit to ensure correct positioning (Chapter 17). Computer software that can aid in the selection of unique nucleotide sequences is described in Chapter 18.

Maximal gene expression

When high protein expression is required strong promoter activity can be complemented by other vector elements that result in replication and amplification of the expression cassette. Utilising vectors with a viral origin of replication (Ori), a sequence of DNA at which replication is initiated, will result in vector replication in cells expressing appropriate viral proteins. In cells that express the Simian Virus 40 (SV40) large tumor antigen (T-Ag) such as HEK 293T and Cos7, vectors with the SV40 Ori are episomally replicated. Similarly, cells in which Epstein–Barr virus nuclear antigen 1 (EBNA1) is expressed such as HEK 293E will replicate vectors containing an Ori-P.

High levels of gene expression from transfected cells can also be achieved through gene amplification of integrated copies of an expression cassette where dihydrofolate reductase gene (*Dhfr*) is co-expressed. This enzyme breaks down Methotrexate (Mtx), by gradually increasing the concentration of Mtx in the culture media it is possible to amplify expression

of both the *Dhfr* gene and an adjacent expression cassette. This amplification approach is commonly used for recombinant protein expression in Chinese hamster ovary (CHO) cells. An alternative approach that results in spontaneous initiation of gene amplification in transfected cells is achieved by using a vector containing a mammalian replication site and a matrix attachment region (MAR) (4). The advantage of this approach is that no viral proteins need to be expressed in the cells because the necessary components are DNA elements.

Regulated expression systems

When working with regulated promoter systems it is optimal that basal expression is minimal in the absence of promoter activation. Synthetic regulated promoters are often studied in the pGL3 vectors (Promega Corp). In this vector family there is an upstream synthetic poly(A) signal/transcriptional pause site (5) to limit promoter activity from the vector backbone which is ideal for regulated promoters. The pGL3 vectors do not contain a selection gene expressed in eukaryotic cells so if stable clones are required a selection vector should be co-transfected with the promoter vector. A selection vector of choice is one that utilises the herpes simplex virus thymidine kinase promoter as this is devoid of enhancer activity (ie pTK-Hyg from Clontech Laboratories Inc.). The pGL4 family of vectors (Promega Corp) have the advantage that there are a reduced number of consensus transcription factor binding sites within the backbone and there are versions with co-expressed eukaryotic selection gene expressed. If you clone your promoter in a vector that also expresses a selection gene in eukaryotic cells then the promoter that drives its constitutive expression could easily influence the activity of the synthetic promoter in the same vector.

In the case of viral vectors endogenous enhancer and promoter elements could also compromise the activity of a cloned synthetic promoter. In retroviral vectors this can be surmounted by using self-inactivating vectors (6, 7) in which endogenous enhancer and

promoter activity is deleted during reverse transcription so that function of the integrated synthetic promoter is not compromised (8).

When regulated expression cassettes are integrated into the genome it is possible that their function can be affected by the genetic environment into which they are inserted. This may mean that the activity of synthetic promoters could dramatically alter due to interactions with enhancer elements neighbouring the integration site. It is possible to overcome these influences through the use of an insulator element which has positional enhancer blocking activity. The first characterised insulator was found at the 5' end of the chicken β-globin locus and is termed the hypersensitivity site 4 (HS4) insulator or 5'HS4 element (9). Insulation activity is mediated by binding of the protein CCCTC-Binding Factor (CTCF) via a single site within a 250bp core element (10). Through the interaction of other proteins the 5'HS4 also has barrier activity which blocks the propagation of heterochromatic structures and protects against DNA methylation.

Long-term expression

In order for synthetic promoters to display long-term activity it is both necessary for the vector/expression cassette to be maintained in the cells and the promoter needs to remain active. Through the use of integrating viruses it is possible to ensure that expression cassettes are maintained indefinitely, but this can to some extent also been achieved with non-integrating vectors. Scaffold or MARs (S/MARS) are often found in association with chromosomal origins of bidirectional replication and when these DNA elements are linked to expression cassettes they can recruit cellular factors to promote episomal replication and mitotic stability (11). A vector with an S/MAR-element in association with the SV40 origin of replication was shown to replicate at very low copy numbers in CHO cells and was stably maintained without selection for more than 100 generations (12). Similarly, when these

elements were used in a non-integrating lentiviral vector there was no decline in gene expression over a 4 week period in 293T cells (13).

In order to maintain promoter activity one option already mentioned is the 5'HS4 element which can potentially protect a promoter sequence from silencing due to methylation. Another DNA sequence termed Ubiquitous Chromatin Opening Element (UCOE) also has anti-silencing properties by preventing DNA methylation. A recently described minimal 0.7 kb UCOE (CBX3-UCOE) functioned well in combination with viral and tissue-specific promoters without altering promoter specificity (14).

Interaction with the immune system

Promoters are DNA elements and as such are not immunogenic *in vivo* but immunogenicity is a consideration for other vector elements. If selection genes are expressed from eukaryotic promoters then these will be immunogenic *in vivo*. Such vectors can be modified so that foreign genes are not expressed (15) or promoters should be transferred to vectors devoid of foreign transgenes. Expression of reporter genes such as EGFP (16) and luciferase (17) are immunogenic in immunocompetent mice. Similarly, regulated promoter systems that utilise protein components that are foreign will be immunogenic (Chapters 12 and 13), as shown with the tetracycline transactivator rtTA (18).

Most plasmid vectors also contain unmethylated CpG dinucleotides which can activate Toll-like receptor 9 (TLR9) (19) which could be important if your cell of choice express these receptors or if you plan to use the vector *in vivo*. There are plasmid vectors available that are devoid of CpG dinucleotides (InvivoGen) and it is also possible to have genes and vectors synthesised in which alternative codons are used so that all CpG dinucleotides are eliminated. TLR activation also occurs with viral vectors and strategies can be employed for evasion (20).

References

1.   Grieger JC, Samulski RJ (2005) Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. J Virol 79:9933-44

2.   Wu J, Zhao W, Zhong L, Han Z, Li B, Ma W, Weigel-Kelley KA, Warrington KH, Srivastava A (2007) Self-complementary recombinant adeno-associated viral vectors: packaging capacity and the role of rep proteins in vector purity. Hum Gene Ther 18:171-82

3.   Maelandsmo GM, Ross PJ, Pavliv M, Meulenbroek RA, Evelegh C, Muruve DA, Graham FL, Parks RJ (2005) Use of a murine secreted alkaline phosphatase as a non-immunogenic reporter gene in mice. J Gene Med 7:307-15

4.   Shimizu N, Miura Y, Sakamoto Y, Tsutsui K (2001) Plasmids with a mammalian replication origin and a matrix attachment region initiate the event similar to gene amplification. Cancer Res 61:6987-90

5.   Levitt N, Briggs D, Gil A, Proudfoot NJ (1989) Definition of an efficient synthetic poly(A) site. Genes Dev 3:1019-25

6.   Yu SF, von Ruden T, Kantoff PW, Garber C, Seiberg M, Ruther U, Anderson WF, Wagner EF, Gilboa E (1986) Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells. Proc Natl Acad Sci U S A 83:3194-8

7.   Zufferey R, Dull T, Mandel RJ, Bukovsky A, Quiroz D, Naldini L, Trono D (1998) Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J Virol 72:9873-80.

8.   Mohamed H, Chernajovsky Y, Gould D (2016) Assembly PCR synthesis of optimally designed, compact, multi-responsive promoters suited to gene therapy application. Sci Rep 6:29388

9.   Chung JH, Bell AC, Felsenfeld G (1997) Characterization of the chicken beta-globin insulator. Proc Natl Acad Sci U S A 94:575-80

10.   Bell AC, West AG, Felsenfeld G (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98:387-96

11.   Baiker A, Maercker C, Piechaczek C, Schmidt SB, Bode J, Benham C, Lipps HJ (2000) Mitotic stability of an episomal vector containing a human scaffold/matrix-attached region is provided by association with nuclear matrix. Nat Cell Biol 2:182-4

12.   Piechaczek C, Fetzer C, Baiker A, Bode J, Lipps HJ (1999) A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. Nucleic Acids Res 27:426-8

13.   Verghese SC, Goloviznina NA, Skinner AM, Lipps HJ, Kurre P (2014) S/MAR sequence confers long-term mitotic stability on non-integrating lentiviral vector episomes without selection. Nucleic Acids Res 42:e53

14.   Muller-Kuller U, Ackermann M, Kolodziej S, Brendel C, Fritsch J, Lachmann N, Kunkel H, Lausen J, Schambach A, Moritz T et al (2015) A minimal ubiquitous chromatin opening element (UCOE) effectively prevents silencing of juxtaposed heterologous promoters by epigenetic remodeling in multipotent and pluripotent stem cells. Nucleic Acids Res 43:1577-92

15.   Subang MC, Fatah R, Wu Y, Hannaman D, Rice J, Evans CF, Chernajovsky Y, Gould D (2015) Effects of APC De-Targeting and GAr Modification on the Duration of Luciferase Expression from Plasmid DNA Delivered to Skeletal Muscle. Curr Gene Ther 15:3-14

16. Gambotto A, Dworacki G, Cicinnati V, Kenniston T, Steitz J, Tuting T, Robbins PD, DeLeo AB (2000) Immunogenicity of enhanced green fluorescent protein (EGFP) in BALB/c mice: identification of an H2-Kd-restricted CTL epitope. Gene Ther 7:2036-40

17. Limberis MP, Bell CL, Wilson JM (2009) Identification of the murine firefly luciferase-specific CD8 T-cell epitopes. Gene Ther 16:441-7

18. Ginhoux F, Turbant S, Gross DA, Poupiot J, Marais T, Lone Y, Lemonnier FA, Firat H, Perez N, Danos O et al (2004) HLA-A*0201-restricted cytolytic responses to the rtTA transactivator dominant and cryptic epitopes compromise transgene expression induced by the tetracycline on system. Mol Ther 10:279-89

19. Cornelie S, Hoebeke J, Schacht AM, Bertin B, Vicogne J, Capron M, Riveau G (2004) Direct evidence that toll-like receptor 9 (TLR9) functionally binds plasmid DNA by specific cytosine-phosphate-guanine motif recognition. J Biol Chem 279:15124-9

20. Agudo J, Ruzo A, Kitur K, Sachidanandam R, Blander JM, Brown BD (2012) A TLR and non-TLR mediated innate response to lentiviruses restricts hepatocyte entry and can be ameliorated by pharmacological blockade. Mol Ther 20:2257-67