ELSEVIER

Cloning technologies for protein expression and purification

James L Hartley

Detailed knowledge of the biochemistry and structure of individual proteins is fundamental to biomedical research. To further our understanding, however, proteins need to be purified in sufficient quantities, usually from recombinant sources. Although the sequences of genomes are now produced in automated factories purified proteins are not, because their behavior is much more variable. The construction of plasmids and viruses to overexpress proteins for their purification is often tedious. Alternatives to traditional methods that are faster, easier and more flexible are needed and are becoming available.

Addresses
Protein Expression Laboratory, Research Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702, USA

Corresponding author: Hartley, James L (hartley@ncifcrf.gov)

Current Opinion in Biotechnology 2006, **17**:359–366

This review comes from a themed issue on
Protein technologies
Edited by Deb K Chatterjee and Joshua LaBaer

Available online 12th July 2006

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2006.06.011

Introduction

Few proteins are abundant enough to be isolated from their native hosts. Clearly evolution has adjusted the *in vivo* level of each protein to match the needs of the cell. The tasks for the scientist who needs a larger amount of a protein are to contrive ways to increase its production in an appropriate host and to facilitate its purification. Protein overexpression requires three components: a gene, a vector that contains the gene, and an expression host that maximizes the amount and quality of the protein produced by the vector–gene combination.

The choice of cloning method to insert a gene into an expression vector is important, principally because proteins are so varied in their behavior that it is impossible to predict whether a protein of interest will express well, be soluble, easy to purify, possess activity or crystallize. (For example, in our laboratory interferon A is well behaved, interferon B is well behaved, but interferon C, which is comprised of part of A and part of B, is insoluble.) This means that a particular vector, expression host or affinity purification tag might not give the desired results, and it is advisable to have in mind how expression methods could be modified if all does not go as planned. Here, I describe how new cloning technologies expedite the manipulations that are often required to obtain pure, active recombinant proteins.

Cloning for protein expression and purification

The required amount, purity, homogeneity, and activity of the protein of interest usually determine the choice of expression context, vector and host. Usually the first consideration is what host cell to use. Expression in *Escherichia coli* is fast, inexpensive and scaleable, and minimal post-translational modifications make proteins purified from *E. coli* relatively homogeneous and highly desirable for structural studies. However, in our experience most eukaryotic proteins larger than 30 kDa are unlikely to be properly folded when expressed in *E. coli* (JL Hartley *et al.*, unpublished). Expression in mammalian cells is best for activity and native structure (including post-translational modifications), but yields are much lower and costs are high. Eukaryotic cells have mechanisms that limit overexpression [1••] (see also http://videocast.nih.gov/PastEvents.asp, November 16, 2005), so viral systems that subvert these mechanisms are often preferable. The relative safety of insect viruses has made the baculovirus system popular [2•], and this host–vector combination has been used extensively. It is particularly suitable for secretion, as commonly used insect cell lines grow well in serum-free media, which in turn simplifies downstream protein purification. Many signal peptides enable some level of secretion [3•] from insect cells, however, the baculovirus gp64/gp67 signal peptide is usually superior (JL Hartley *et al.*, unpublished).

The first recombinant proteins were overproduced by simply cloning entire DNA segments on multicopy plasmids; that is, the amount of protein was increased by increasing the copy number of the promoter–gene combination. Later on, genes were removed from the influence of their native promoters and inserted downstream of more powerful, inducible promoters. Now that the polymerase chain reaction (PCR) has made gene modification routine, the gene itself and regions on either side can be modified without undue difficulty or expense. Translation starts can be modulated, stop codons can be removed, epitopes, affinity tags, and signal peptides can be added (the latter by overlap PCR [4] if they are long), and entire genes can be synthesized for optimum codon usage. As a result, most protein expression is now done using genes that have been modified through PCR, even if only to add convenient cloning sites.

As most eukaryotic genes contain introns, cloning these genes for protein expression usually begins with cDNAs.

Using a cDNA as a PCR template, the gene of interest is typically amplified (sometimes in multiple versions with various modifications), cloned, sequenced, and then subcloned into one or more expression vectors. This initial cloning into a non-expression vector is often a big time saver for several reasons. Firstly, unintended expression of the gene of interest can cause toxicity and make it difficult to clone some genes directly into an expression vector. Secondly, many expression vectors are large (complicating cloning) or have low copy number (complicating both cloning and sequencing). Thirdly, it is surprising how many genes, even those with good pedigrees, contain sequence errors or single-nucleotide polymorphisms (JL Hartley *et al.*, unpublished). Lastly, errors in PCR primers are common (usually owing to the low fidelity of chemical synthesis). Thus, preliminary cloning into a smaller, higher copy, non-expressing vector facilitates both the cloning itself and also sequence confirmation of the gene and motifs that have been added via PCR primers to improve expression and protein purification.

Cloning sites and technologies

To appreciate the effect of cloning methods on protein expression, it is necessary to catalog briefly the sequences commonly added to genes that affect translation of the gene or which encode domains that make the resulting protein more useful or easier to purify (Table 1). Many of these motifs, along with cloning sites, are added to PCR primers and are thus incorporated into PCR products (typical examples shown in Figure 1). Additional protein expression elements are found in the vectors into which gene cassettes are introduced (Figure 2). The N- and C-terminal cloning sites form the junctions between the genes and the vectors that contain them. Table 2 summarizes the primary cloning methods used in the protein expression literature. For example, in a typical project one might consider which elements in Table 1 would be required for the expression, purification, and utility of the desired protein, and then decide which elements can be incorporated during PCR amplification of the gene (Figure 1) and which would be derived from the cloning vector (Figure 2). These factors would need to be weighed against the availability of the desired vectors for the available cloning technologies (Table 2).

Most protein expression plasmids are constructed with restriction enzymes and ligase (REaL; Figure 3a). After the desired protein configuration (see the elements in Table 1) has been decided, a vector is acquired and the gene is designed to be compatible with the vector (i.e. appropriate restriction sites, tags, reading frame, orientation), usually as part of a PCR amplification step. For

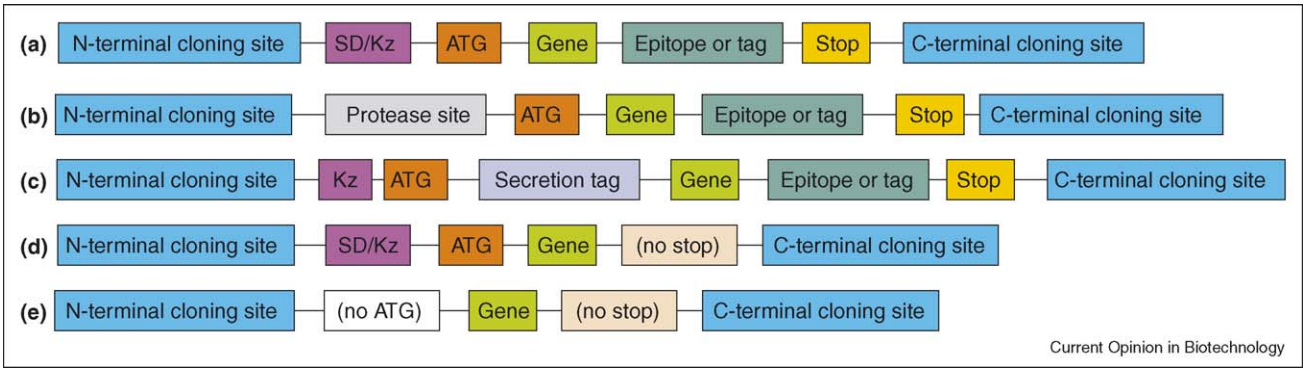
Table 1

Motifs that are commonly added to improve protein expression, purification or usefulness

Element	Location relative to open reading frame	Function	Comment	Refs
Kozak (Kz) (eukaryotic hosts)	From −1 to −6 relative to ATG	Increases recognition of the first ATG by the ribosome	Consensus is GCCRCCATGR, where the purines (R) at −3 and +4 are most important Functions in most or all eukaryotes Can be combined with Shine–Dalgarno	[22•]
Shine–Dalgarno (SD, aka ribosome-binding site, rbs) (*E. coli*)	5 to 13 bases upstream of ATG	Increases recognition of ATG by the ribosome in *E. coli*	Consensus in *E. coli* is GGAGG Less preference for the first ATG than in eukaryotes Can be combined with Kozak	[22•,23•,24]
Epitopes	Usually at the C terminus of genes when expressed in eukaryotic cells	Detection by antibodies (e.g. western blots)	Examples are myc (10 aa), hemagglutinin (9 aa), Flag (9 aa) and His6 (6 aa)	[25]
Affinity tags	Usually at the N terminus in *E. coli* expression, C terminus in eukaryotic expression	Protein purification	His6 is most widely used Others are GST, biotin and *Strep*-Tag	[26]
Protease sites	Usually upstream of the gene	Removal of affinity or solubility tag	Proteases from tobacco etch virus and rhinovirus 3C are the most specific Others are thrombin and factor Xa	[27,28,29•]
Solubility tags	Usually at N terminus	Improves solubility and translation efficiency	Most often used in *E. coli* expression but sometimes helps in eukaryotic hosts	[26]
Secretion tags (signal peptides)	N terminus	Causes secretion to medium or periplasm	Proteins that are cytoplasmic (e.g. GFP) can also be secreted when expressed with a signal peptide	[30–32]
IRES (internal ribosome entry site)	Downstream of first gene	For expression of second gene on same mRNA in eukaryotic cells	Expressing drug resistance (neomycin, puromycin, etc.) from the same message as the gene of interest minimizes silencing	[22•,33•]
Fluorescent protein	Usually at C terminus	Readout for solubility or intracellular localization	Fluorescence increases as protein becomes more soluble	[34]

aa, amino acids; GFP, green fluorescent protein.

Figure 1

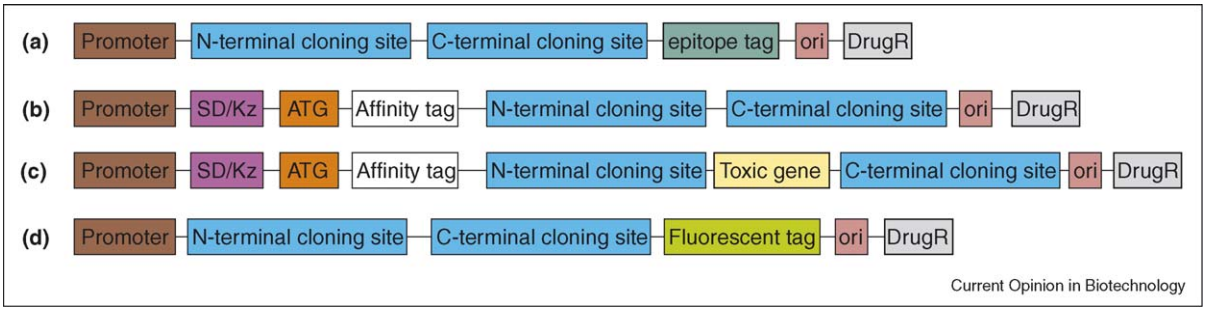


Model PCR products showing combinations of protein expression motifs relative to the cloning sites used to insert the gene into an expression vector. **(a)** PCR product encoding a protein with a C-terminal epitope (for antibody detection or purification) or tag (for purification). **(b)** PCR product encoding a protein with an upstream protease site (for removal of an N-terminal tag derived from the expression vector) and a C-terminal epitope or tag. **(c)** PCR product encoding a protein secretion signal and a C-terminal epitope or tag. **(d)** PCR product encoding a protein without a stop codon, so that a variety of C-terminal tags can be added by the expression vector. **(e)** PCR product encoding a protein that requires the expression vector to supply both the translation start and stop (i.e. the protein will have extra amino acids from the expression vector at both the N and C termini). Kz, Kozak; SD, Shine–Dalgarno sequence.

example, for a PCR product such as that shown in Figure 1b the cloning sites at the N and C termini must be compatible with the chosen cloning vector, absent from the gene itself, and positioned in the correct reading frame and orientation. As the PCR product does not have any translation start signals (Shine-Dalgarno or Kozak) and is designed for subsequent removal of an N-terminal tag (using a protease), a vector similar to those in Figure 2b or c is appropriate, because both supply an N-terminal tag and associated translation starts. Vectors with a multiple cloning site (many restriction sites available for gene insertion) are the most flexible, but any unused sites add extra amino acids. Of course as these amino acids, along with the N-terminal tag, will be removed by the protease step, these extra amino acids are of no consequence unless they contribute to an undesirable trait early in the procedure, such as toxicity or insolubility. All other factors being equal, the vector shown in Figure 2c (where the stuffer between cloning sites contains a toxic gene) is preferable, as any vector that is incompletely cut or that does not otherwise lose its toxic gene will not produce a colony.

REaL is the least expensive, if most labor-intensive, way to clone genes for protein expression, and there are thousands of vectors available. Nevertheless, many of the virtues of the REaL approach are also liabilities. Restriction sites are small and more than 200 are commercially available, which in turn means that a site that would otherwise be used in a cloning scheme might also be found in an inconvenient location within the gene or vector of interest. There are many procedures that can be used to modify DNA ends to make them compatible with a strategy, but these require time and practice and reagents to execute well. A further complication is that DNA ligase

Figure 2



Typical combinations of protein expression elements in expression vectors. All expression vectors supply a promoter upstream of the N-terminal cloning site, which may or may not be controllable. The N- and C-terminal cloning sites can be a single site. **(a)** Provides a C-terminal epitope tag. **(b)** Provides an N-terminal translation start and affinity tag. **(c)** A vector similar to (b) but with the presence of a toxic gene between the cloning sites for selection against empty vector molecules (i.e. vectors that do not contain the gene of interest). **(d)** Vector containing a C-terminal fluorescent protein tag (e.g. green fluorescent protein) for making fluorescent fusion proteins. drugR, drug resistance gene; ori, origin of replication.

Table 2

Cloning methods for protein expression

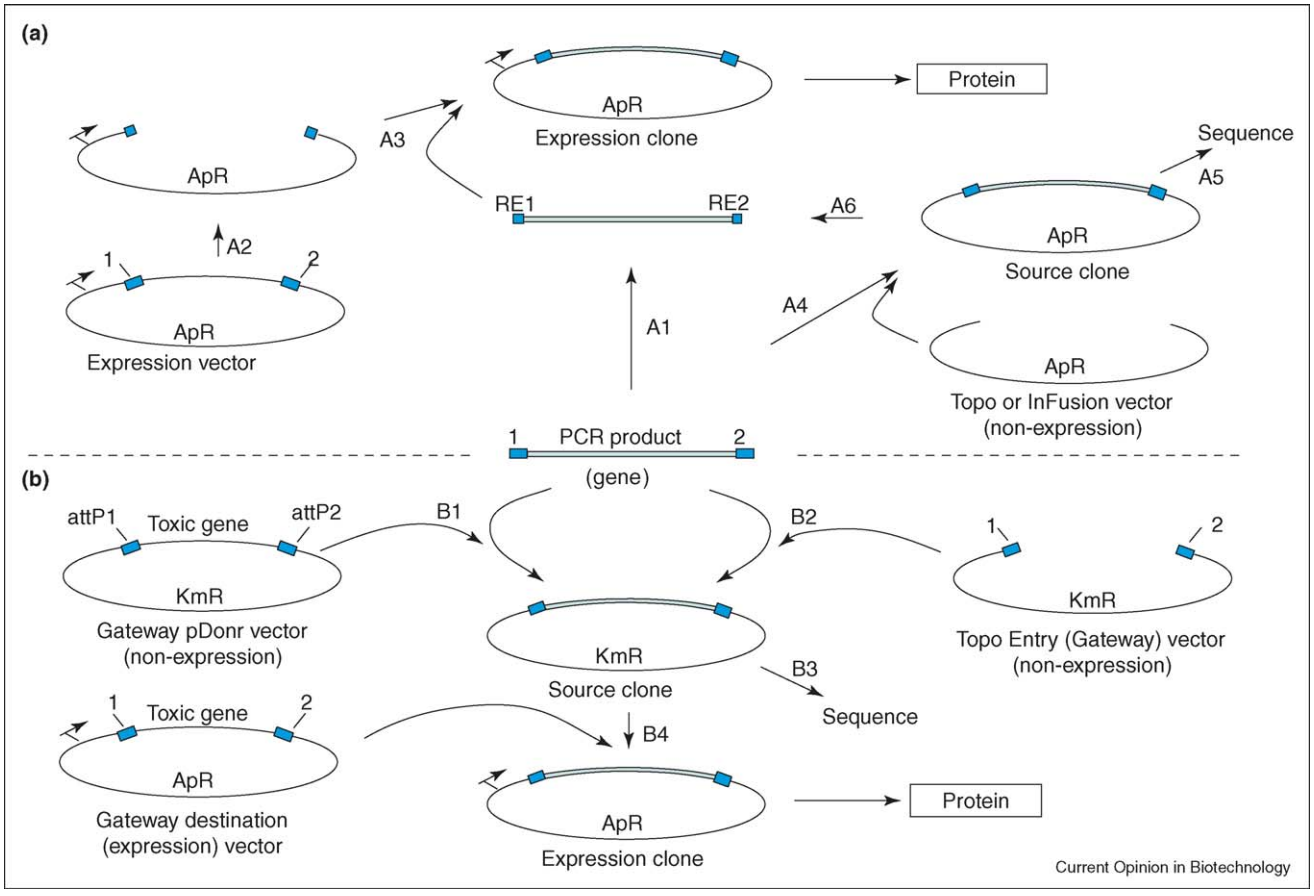
Cloning method	Cloning sites	Size of cloning sites	Comments	Refs	Websites
REaL[a], Flexi Vector	Restriction sites	6–8 bp/ 2–3 amino acids[b]	Vectors with multiple cloning sites can add more base pairs or amino acids Direct cloning into expression vector if desired Largest number of available vectors, but no convention for sites, reading frame or orientation Laborious to subclone to new vectors	[35••]	A
Gateway SSR[c] (Invitrogen)	AttB sites	21–25 bp/ 7–9 amino acids[b]	Large cloning sites Intitial cloning into non-expression vector REaL vectors must be converted Trivial to sublcone into hundreds of available vectors All vectors compatible	[36,37•–39•, 40,41,42••]	B
Topo (Invitrogen)	Trapped topoisomerase	9–16 bp/ 3–6 amino acids[b]	For cloning of PCR products (5′OH required) Vectors must be purchased Some vectors for immediate expression Entry to Gateway technology	[43]	C
Creator SSR (ClonTech)	LoxP sites	34 bp/ 11–12 amino acids[b]	Large sites Initial cloning into non-expression vector REaL vectors must be converted Easy to subclone into available vectors C-terminal fusions must undergo splicing All vectors compatible	[44•,45•,46]	D
InFusion homologous recombination (ClonTech)	Homologies to vector ends	14–15 bp/ 5 amino acids[b]	For cloning PCR products Compatible with any vector No convention for sequences Vectors must be linearized Some vectors available for immediate expression Entry to Creator technology	[47]	E
Ligation-independent cloning (Novagen)	Single-stranded homologies to vector ends	12–15 bp/ 4–5 amino acids[b]	No convention for sequences Compatible with any REaL vector Joining reaction is non-enzymatic (anneal sticky ends) Vectors must be cut Vectors and PCR products must be resected	[6,7]	F
MAGIC	Homologies to vector	~50 bp/ ~17 amino acids[b]	Reactions occur inside *E. coli* cells Vectors must be converted No convention for sequences	[20]	
Overlap PCR	Homologies to other PCR products	12–15 bp	Used to combine PCR products into larger molecules Can be used to make deletions, insertions or add larger domains to genes	[4]	
Gateway MultiSite	AttB sites	21 bp	Sequence variants of attB sites used to clone PCR products Multiple sequence-confirmed clones can be recombined into a vector in a single reaction All reading frames and orientations determined by attB sites in PCR primers	[8•]	G

[a] Restriction enzymes and ligase.
[b] If the cloning sites are translated into amino acids.
[c] Site-specific recombination. Web sites: A, http://www.promega.com/tbs/tm254/tm254.pdf (Flexi Vector at Promega); B, http://www.invitrogen.com/content.cfm?pageid=4072 (Gateway at Invitrogen); C, http://www.invitrogen.com/content.cfm?pageid=4073 (Topo cloning at Invitrogen); D, http://www.clontech.com/clontech/products/families/creator/index.shtml (Creator at ClonTech); E, http://www.clontech.com/clontech/infusion/index.shtml (InFusion at ClonTech); F, http://www.emdbiosciences.com/docs/NDIS/inno05-002.pdf (LIC at Novagen); G, http://www.invitrogen.com/content/sfs/manuals/multisite_gateway_man.pdf.

can join ends that ought to be incompatible, leading to clones that are in the wrong orientation or reading frame. The multitude of protein expression vectors offers a universe of solutions to cloning, expression and purification problems, but their diversity, in both use and structure, makes acquiring and using them a tedious process.

Site-specific recombination (SSR; Table 2 and Figure 3b) is the most widely used alternative to REaL for protein expression. (SSR occurs between specific base sequences, whereas homologous recombination requires runs of identical sequences of no particular sequence.) The enzymes that execute SSR combine cutting and ligation reactions

Figure 3



Comparison of restriction enzymes and ligase (REaL) and site-specific recombination (SSR) by the Gateway method for the construction of clones that express a protein. The starting material for both methods is a PCR product (center) in which the blue squares are either restriction sites or recombination sites. **(a)** The construction of expression clones using REaL. A1, start of the direct REaL path, digestion of the PCR product with restriction enzyme(s). A2, preparation of the expression vector with restriction enzyme(s), phosphatase treatment, etc. A3, ligation of the linear expression vector and the PCR product. A4, start of the indirect REaL path, cloning of the PCR product (including restriction sites) into a non-expression vector. A5, sequence confirmation of the cloned PCR product. A6, preparation of the restriction fragment for ligation into the linear expression vector (step A3). **(b)** The construction of expression clones using SSR. There are two ways to add flanking recombination sites for the Gateway SSR method. If sites 1 and 2 in the initial PCR product are Gateway attB1 and attB2 sites, the PCR product is cloned by SSR with a Gateway pDonr vector (B1). If sites 1 and 2 are plain ends (any sequence) the PCR product is cloned by Topo cloning (B2) into a Gateway entry vector with flanking att sites. (PCR products can also be cloned between recombination sites using REaL; not shown.) The product of either B1 or B2 is a non-expressing source clone. B3, sequence confirmation of the cloned PCR product. B4, subcloning of the gene from the source clone into a Gateway destination (expression) vector by SSR. The arrows on expression vectors represent promoters. ApR, ampicillin resistance. KmR, kanamycin resistance.

in a single, concerted process, which accounts for their exceedingly high accuracy. SSR recognition sites are considerably longer than restriction sites, so that SSR reactions are also extremely specific. The affinities of SSR enzymes for their sites (compared with DNA ligase) make their reactions relatively insensitive to the size of the molecules involved (one SSR baculovirus vector is more than 100 kb in length).

Both commercially available SSR cloning technologies are derived from bacteriophages: Gateway (Invitrogen) from phage lambda and Creator (ClonTech) from phage P1 (Table 2). The use of the Gateway technology is illustrated in Figure 3b; the Creator technology is similar in outline but differs in detail. Although both companies sell a variety of protein expression vectors, they also address the diverse needs of their customers by making it as easy as possible to convert existing vectors to compatibility with their technologies. Particularly for Gateway, hundreds of vectors initially constructed for use with REaL have been made Gateway-compatible, not only for protein expression and purification but also for many *in vivo* purposes. One of the most important advantages of the two SSR cloning technologies is that

both have adopted a convention for reading frame and orientation. In contrast to the multitude of REaL vectors, all Gateway and Creator vectors are compatible (i.e. reading frame and gene orientation) with every gene cloned into their respective systems. And, because of powerful selections and the lack of dependence on free DNA ends, one transfers a gene into a new expression vector by mixing the two uncut plasmids, incubating with SSR enzyme(s), and then transforming *E. coli*. Both companies have very efficient technologies for cloning PCR products into their SSR systems: Topo cloning from Invitrogen and InFusion from ClonTech (Table 2).

Ligation-independent cloning (LIC; marketed by Novagen) [5–7] allows PCR products to be cloned without the need to add restriction sites. LIC involves treating linear vectors and PCR products with an exonuclease/polymerase to make precise single-stranded tails that can then be annealed to each other. The principal disadvantages of LIC are the difficulty in preparing new vectors and the lack of consensus sequences, orientations, and reading frames. Overlap PCR [4] is an easy way to combine two PCR products into a longer molecule so as to alter the structure of a gene (e.g. to add a signal peptide for secretion or to make a deletion). In using Gateway Multi-Site, different recombination sites are added to PCR products which are then cloned and sequenced. SSR enzymes acting at the recombination sites remove multiple (typically three) inserts from their respective non-expressing vectors, link them end to end, and clone them into an expression (destination) vector in a single reaction [8•,9•,10,11•].

Conclusions and recommendations

In light of the above discussion, recommendations for choosing the best cloning and expression strategies are summarised in Box 1. Cloning for protein expression and purification is most conveniently approached through PCR, which facilitates the introduction of necessary modifications to the coding sequence (e.g. the inclusion of expression, purification and utility motifs; Table 1). A cloned and sequenced PCR product can then be moved into expression vectors with restriction enzymes and ligase (REaL), but the greater ease and reliability of cloning methods based on SSR are making them an increasingly popular alternative. As the size of genes and vectors increase, one might consider switching from REaL to SSR, where size matters much less. SSR greatly simplifies the cloning process, dramatically improves reliability, and allows the same insert to be subcloned into multiple expression vectors with very little labor (i.e. supercoiled plasmids are mixed, incubated and transformed). The penalty is larger cloning sites, but in most circumstances this has little relevance. Both the Gateway and Creator systems for SSR offer pros and cons (see Box 1); for a comparison of Gateway and Creator in a high-throughput setting see the article by LaBaer *et al.* [12•]. In a recent advance, Gateway vectors have been modified so that subcloning reactions can be performed inside *E. coli* cells [13•,14•], which could offer a very low cost alternative for SSR cloning.

Box 1 Recommendations for choosing cloning and expression strategies

Cloning recommendations

It is easiest to add cloning sites (if needed) and expression, purification and utility motifs with PCR

Use a PCR template with a verified pedigree

Do not assume that just because a particular plasmid has been used extensively, its sequence is trustworthy

The Mammalian Gene Collection [20] is the most reliable source of cDNAs

Always sequence cloned PCR products

Choice of expression vector

REaL is the least expensive but most labor-intensive way to clone genes for protein expression

Thousands of vectors are available for REaL

Flexi Vector (Promega) and LIC (Novagen) offer alternatives to REaL, but both require digestion of PCR products and vectors

For larger genes and vectors, SSR is recommended over REaL as size is less important

SSR greatly simplifies the cloning process, dramatically improves reliability, and allows the same insert to be subcloned into multiple expression vectors with very little labor

SSR require larger cloning sites, but in most circumstances this has little relevance

Gateway enzymes are expensive: most laboratories scale down the reactions to 5 or 10 μl

The Creator enzyme (Cre recombinase) costs less and you can purify your own [21]

Creator is less flexible than Gateway and fewer vectors are available

Gateway vectors have been modified so that subcloning reactions can be performed inside *E. coli* cells

For large protein expression and purification projects, REaL and LIC can be practical alternatives to SSR because the same vectors are used hundreds of times and a smooth pipeline can be developed

For large protein expression and purification projects, such as those undertaken by the laboratories of the Protein Structure Initiative [15–17], REaL and LIC can be practical alternatives to SSR because the same vectors are used hundreds of times and a smooth pipeline can be developed. MAGIC is an *in vivo* method that employs homologous recombination [18•] and seems well suited to high-throughput production, but thus far it has been applied only to the synthesis of short-hairpin RNA [19].

In summary, recent advances have led to new methods of cloning genes for the overexpression and purification of

proteins. These technologies are faster, easier to use and more flexible. In the future we are likely to witness further improvements, as interest moves from the genome to the proteome and the need to obtain purified proteins intensifies.

Acknowledgements

Work in the author's laboratory is funded with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Isler JA, Skalet AH, Alwine JC: **Human cytomegalovirus infection**
•• **activates and regulates the unfolded protein response**. *J Virol* 2005, **79**:6890-6899.
The data in this paper suggest that, where possible, it is preferable to employ viral expression systems in eukaryotes because viruses have evolved ways of bypassing feedback mechanisms that limit protein overexpression.

2. Kost TA, Condreay JP, Jarvis DL: **Baculovirus as versatile**
• **vectors for protein expression in insect and mammalian cells**. *Nat Biotechnol* 2005, **23**:567-575.
This review not only covers protein overexpression but also considers novel uses of baculovirus particles.

3. Redaelli L, Zolezzi F, Nardese V, Bellanti B, Wanke V, Carettoni D:
• **A platform for high-throughput expression of recombinant human enzymes secreted by insect cells**. *J Biotechnol* 2005, **120**:59-71.
In this study, 85% of human proteins were secreted at some level with their native signal peptides.

4. Horton RM: **Cai 2, Ho SN, Pease LR: Gene splicing by overlap extension: Tailor-made genes using the polymerase chain reaction**. *Biotechniques* 1990, **8**:528-535.

5. Cabrita LD, Dai W, Bottomley SP: **A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production**. *BMC Biotechnol* 2006, **6**:12.

6. Chanda PK, Edris WA, Kennedy JD: **A set of ligation-independent expression vectors for co-expression of proteins in *Escherichia coli***. *Protein Expr Purif* 2005, in press.

7. Haun RS, Moss J: **Ligation-independent cloning of glutathione S-transferase fusion genes for expression in *Escherichia coli***. *Gene* 1992, **112**:37-43.

8. Cheo DL, Titus SA, Byrd DR, Hartley JL, Temple GF, Brasch MA:
• **Concerted assembly and cloning of multiple DNA segments using *in vitro* site-specific recombination: functional analysis of multi-segment expression clones**. *Genome Res* 2004, **14**:2111-2120.
This is a good explanation of how multiple DNA segments can be assembled in a specified order, orientation and reading frame in a single *in vitro* reaction.

9. Sasaki Y, Sone T, Yahata K, Kishine H, Hotta J, Chesnut JD, Honda
• T, Imamoto F: **Multi-gene gateway clone design for expression of multiple heterologous genes in living cells: eukaryotic clones containing two and three ORF multi-gene cassettes expressed from a single promoter**. *J Biotechnol* 2005, [Epub ahead of print].
Gateway MultiSite was used to chain together six DNA segments in the correct order, orientation and reading frame in a vector in a single reaction.

10. Tubb J, Groth AC, Leong L, Emery DW: **Simultaneous sequence transfer into two independent locations of a reporter vector using MultiSite Gateway technology**. *Biotechniques* 2005, **39**:553-557.

11. van Dooren GG, Marti M, Tonkin CJ, Stimmler LM, Cowman AF,
• McFadden GI: **Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum***. *Mol Microbiol* 2005, **57**:405-419.
The authors use Gateway MultiSite to combine a promoter, a gene, a C-terminal fluorescent protein, and a vector in a single reaction.

12. Marsischky G, LaBaer J: **Many paths to many clones: a**
• **comparative look at high-throughput cloning methods**. *Genome Res* 2004, **14**:2020-2028.
This paper compares the Gateway, Creator, and Univector SSR technologies in the context of assembling large libraries of ORFs.

13. House BL, Mortimer MW, Kahn ML: **New recombination**
• **methods for *Sinorhizobium meliloti* genetics**. *Appl Environ Microbiol* 2004, **70**:2806-2815.
The authors describe how to do Gateway subcloning inside *E. coli* cells.

14. Schroeder BK, House BL, Mortimer MW, Yurgel SN, Maloney SC,
• Ward KL, Kahn ML: **Development of a functional genomics platform for *Sinorhizobium meliloti*: construction of an ORFeome**. *Appl Environ Microbiol* 2005, **71**:5858-5864.
A very clear description of how to clone an ORFeome.

15. Lattman E: **The state of the Protein Structure Initiative**. *Proteins* 2004, **54**:611-615.

16. Klock HE, White A, Koesema E, Lesley SA: **Methods and results for semi-automated cloning using integrated robotics**. *J Struct Funct Genomics* 2005, **6**:89-94.

17. Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A: **An expression vector tailored for large-scale, high-throughput purification of recombinant proteins.** *Protein Expr Purif* 2006, in press.

18. Li MZ, Elledge SJ: **MAGIC, an *in vivo* genetic method for**
• **the rapid construction of recombinant DNA molecules**. *Nat Genet* 2005, **37**:311-319.
The authors describe how to subclone a gene into new vectors with *in vivo* homologous recombination. The homologous sequences are completely flexible (in contrast to SSR), but have to be the same in all vectors to be compatible.

19. Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, Siolas D, Hu G, Paddison PJ, Schlabach MR *et al.*: **Second-generation shRNA libraries covering the mouse and human genomes**. *Nat Genet* 2005, **37**:1281-1288.

20. Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF *et al.*: **Mammalian Gene Collection Program Team. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences**. *Proc Natl Acad Sci USA* 2002, **99**:16899-16903.

21. Liu Q, Li MZ, Leibham D, Cortez D, Elledge SJ: **The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes**. *Curr Biol* 1998, **8**:1300-1309.

22. Kozak M: **Regulation of translation via mRNA structure in**
• **prokaryotes and eukaryotes**. *Gene* 2005, **361**:13-37.
The author has very high standards of proof and appears to have read every relevant paper in depth.

23. Betton JM: **High throughput cloning and expression strategies**
• **for protein production**. *Biochimie* 2004, **86**:601-605.
Genes high in GC content might express poorly in *E. coli* owing to mRNA secondary structure. The author describes a strategy to test alternative codons in the amino end of a gene via *in vitro* transcription and translation of PCR products.

24. Olins PO, Rangwala SH: **A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in *Escherichia coli***. *J Biol Chem* 1989, **264**:16973-16976.

25. Jarvik JW, Telmer CA: **Epitope tagging**. *Annu Rev Genet* 1998, **32**:601-618.

26. Waugh DS: **Making the most of affinity tags**. *Trends Biotechnol* 2005, **23**:316-320.

27. Knott JA, Orr DC, Montgomery DS, Sullivan CA, Weston A: **The expression and purification of human rhinovirus protease 3C**. *Eur J Biochem* 1989, **182**:547-555.

28. Kapust RB, Tozser J, Copeland TD, Waugh DS: **The P1′ specificity of tobacco etch virus protease**. *Biochem Biophys Res Commun* 2002, **294**:949-955.

29. • Cheeseman IM, Desai A: **A combined approach for the localization and tandem affinity purification of protein complexes from metazoans**. *Sci STKE* 2005, **2005**:pl1.
Although rhinovirus protease 3C is marketed by Amersham/GE Healthcare (as PreScission protease), there is not a good review of its uses and properties.

30. Kim JY, Fogarty EA, Lu FJ, Zhu H, Wheelock GD, Henderson LA, DeLisa MP: **Twin-arginine translocation of active human tissue plasminogen activator in *Escherichia coli***. *Appl Environ Microbiol* 2005, **71**:8451-8459.

31. Esposito D, Gillette WK, Miller DA, Taylor TE, Frank PH, Hu R, Bekisz J, Hernandez J, Cregg JM, Zoon KC, Hartley JL: **Gateway cloning is compatible with protein secretion from *Pichia pastoris***. *Protein Expr Purif* 2005, **40**:424-428.

32. Georgiou G, Segatori L: **Preparative expression of secreted proteins in bacteria: status report and future prospects**. *Curr Opin Biotechnol* 2005, **16**:538-545.

33. • Jackson RJ: **Alternative mechanisms of initiating translation of mammalian mRNAs**. *Biochem Soc Trans* 2005, **33**:1231-1241.
Jackson has a more functional view of internal ribosome entry site activities than to Kozak.

34. Cabantous S, Pedelacq JD, Mark BL, Naranjo C, Terwilliger TC, Waldo GS: **Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis***. *J Struct Funct Genomics* 2005, **6**:113-119.

35. •• Blommel PG, Martin PA, Wrobel RL, Steffen E, Fox BG: **High efficiency single step production of expression plasmids from cDNA clones using the Flexi Vector cloning system**. *Protein Expr Purif* 2006, **47**:562-570.
The authors show how rare-cutting restriction enzymes can be combined with selections to facilitate the direct cloning of PCR products into expression vectors. They include ways to work around the non-specificity of ligase.

36. Hartley JL, Temple GF, Brasch MA: **DNA cloning using *in vitro* site-specific recombination**. *Genome Res* 2000, **10**:1788-1795.

37. • Kornienko M, Montalvo A, Carpenter BE, Lenard M, Abeywickrema P, Hall DL, Darke PL, Kuo LC: **Protein expression plasmids produced rapidly: streamlining cloning protocols and robotic handling**. *Assay Drug Dev Technol* 2005, **3**:661-674.
Gateway was used to investigate many expression options.

38. • Luan CH, Qiu S, Finley JB, Carson M, Gray RJ, Huang W, Johnson D, Tsao J, Reboul J, Vaglio P *et al.*: **High-throughput expression of *C. elegans* proteins**. *Genome Res* 2004, **14**:2102-2110.
This paper contains a nice description of high-throughput cloning for protein expression using the Gateway technology.

39. • le Maire A, Schiltz M, Braud S, Gondry M, Charbonnier JB, Zinn-Justin S, Stura E: **Crystallization and halide phasing of the C-terminal domain of human KIN17**. *Acta Crystallograph Sect F Struct Biol Cryst Commun* 2006, **62**:245-248.
A nice example of testing several different expression conditions to find a way to produce a crystallizable protein.

40. Niegowski D, Hedrén M, Nordlund P, Eshaghi S: **A simple strategy towards membrane protein purification and crystallization**. *Int J Biol Macromol* 2006, in press.

41. Thao S, Zhao Q, Kimball T, Steffen E, Blommel PG, Riters M, Newman CS, Fox BG, Wrobel RL: **Results from high-throughput DNA cloning of *Arabidopsis thaliana* target genes using site-specific recombination**. *J Struct Funct Genomics* 2004, **5**:267-276.

42. •• Tirat A, Freuler F, Stettler T, Mayr LM, Leder L: **Evaluation of two novel tag-based labelling technologies for site-specific modification of proteins**. *Int J Biol Macromol* 2006, in press.
A custom Gateway entry vector was used to add a Prescission (rhinovirus 3C) protease to the gene of interest, which was then tagged with either biotin or fluorescent dyes. This paper contains very clear descriptions of biotinylation motifs and how to optimize both their performance and the so-called SNAP-tag technology.

43. Shuman S: **Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase**. *J Biol Chem* 1994, **269**:32678-32684.

44. • Park J, Hu Y, Murthy TV, Vannberg F, Shen B, Rolfs A, Hutti JE, Cantley LC, Labaer J, Harlow E, Brizuela L: **Building a human kinase gene repository: bioinformatics, molecular cloning, and functional validation**. *Proc Natl Acad Sci USA* 2005, **102**:8114-8119.
An example of high-throughput cloning of PCR products using InFusion homologous recombination and subcloning into expression vectors with the Creator SSR technology.

45. • Hilson P: **Cloned sequence repertoires for small- and large-scale biology**. *Trends Plant Sci* 2006, **11**:133-141.
A comparison of Gateway, Creator and Univector SSR technologies as applied to ORF libraries. A listing of plant resources that are compatible with SSR cloning is also provided.

46. Colwill K, Wells CD, Elder K, Goudreault M, Hersi K, Kulkarni S, Hardy WR, Pawson T, Morin GB: **Modification of the Creator recombination system for proteomics applications — improved expression by addition of splice sites**. *BMC Biotechnol* 2006, **6**:13.

47. Howorka S, Bayley H: **High-throughput scanning mutagenesis by recombination polymerase chain reaction**. *Methods Mol Biol* 2002, **182**:139-147.