Genetic toolkits to design and build mammalian synthetic systems

Roberto Di Blas^{1,2+}, Annalise Zouein^{1,2,3+}, Tom Ellis^{2,3,4}, Francesca Ceroni^{1,2*}

1. Department of Chemical Engineering, Imperial College London, South Kensington Campus, London (UK)

2. Imperial College Centre for Synthetic Biology, South Kensington Campus, London (UK)

3. Department of Bioengineering, Imperial College London, South Kensington Campus, London (UK)

4. Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge (UK)

+ These authors contributed equally to this work

* Correspondence: f.ceroni@imperial.ac.uk (F. Ceroni)

KEYWORDS

Cloning, Standardisation, DNA construct, Design-Build-Test-Cycle, Mammalian synthetic biology

ABSTRACT

Construction of DNA-encoded programs is central to synthetic biology and the chosen method often determines the time required to design and build constructs for testing. Here we describe and summarise key features of the available toolkits for DNA construction for mammalian cells. We compare the different cloning strategies based on their complexity and the time needed to generate constructs of different sizes, and we reflect on why Golden Gate toolkits now dominate due to their modular design. We look forward to future advances, including accessory packs to cloning toolkits that can facilitate editing, orthogonality, advanced regulation, and integration into synthetic chromosome construction.

MAMMALIAN CLONING TOOLKITS

Synthetic biology is now a mature discipline that is advancing our ability to use biological systems in a wide range of applications, from biosensing[1] to production of biomaterials[2, 3] and therapeutics[4, 5]. This is achieved by applying engineering principles such as the design-build-test cycle, and by ensuring standardisation and thorough characterisation of biological parts used in modular cloning strategies[6, 7]. Fast DNA assembly and rapid design prototyping have become essential tools that enable the production and testing of large libraries of DNA constructs from which the best-performing candidates for further implementation can be identified[8]. Molecular cloning still represents the main means to achieve genetic construct assembly, but DNA synthesis together with novel DNA assembly methods have improved the pace and scale at which genetic constructs are now built and tested[9].

Over the last decade, many synthetic biology cloning ‘toolkits’ have been described and made available to the scientific community in order to ease and accelerate the assembly of complex genetic constructs. Although most of the toolkits currently available are based on bacterial[10-12], yeast[13-16] and plant [17-19] DNA components, toolkits for mammalian synthetic biology have begun to appear. Unsurprisingly, mammalian cloning toolkits have lagged behind microbial equivalents due to the higher complexity of

mammalian biology in which characterisation, standardisation and testing is more complex and time-consuming[20]. The number and quality of available and characterised DNA components for mammalian synthetic biology is simply not yet comparable to that for microbial synthetic biology.

The first mammalian toolkits mainly focused on the development of a modular framework for cloning. However, more recent toolkit publications also include valuable genetic part characterisation results[21, 22]. In some cases, data are provided for genetic part performance in both transiently and stably expressed systems, recognising that integrated genetic constructs behave differently when compared to transiently expressed plasmid DNA[23].

Given the accelerating number of different mammalian modular cloning toolkits, and the crucial role they are playing in enabling and standardising mammalian synthetic biology, we here provide a review of these toolkits and their differences. We classify mammalian toolkits according to the cloning strategy they are based upon and describe the main advantages and disadvantages of each. As well as providing a schematic summary of the overall available toolkit systems, we also examine more in detail the best developed toolkits where a wider number of characterised components are included. This review is designed to provide a useful synopsis for researchers wanting to identify the right cloning strategy for their endeavour.

1. CLONING TOOLKITS

1.1 Golden Gate-based toolkits

Many of the most promising mammalian cloning toolkits are based on the Golden Gate method of DNA assembly [24] (Box 1, Figure 1A). This method has the advantages of being modular, hierarchical, standardised, and relatively simple at the design stage. Automation of Golden Gate DNA assembly is also possible. Golden Gate-based cloning toolkits can be shipped in plate format with standardised DNA parts that include sequence overhangs with required Type IIS restriction sites. This enables immediate use and explains why these toolkits have quickly become standard in many synthetic biology laboratories and have been implemented for a wide variety of host organisms including mammalian cells [18, 19]. There are many widely available mammalian synthetic biology toolkits that utilise Golden Gate-based construction (Table 1). In this section we cover the EMMA, MTK, MOPET, mMoClo, and Multi Kingdom toolkits.

EMMA toolkit. The Extensible Mammalian Modular Assembly (EMMA) toolkit, is the most comprehensive toolkit for mammalian Golden Gate compatible parts currently available[21]. It provides a highly modular method for construction of 3 transcriptional units (TUs) carrying up to 25 parts at-a-time. Each part is specified by a 4-base overhang left after digestion with appropriate enzyme. The use of standardised DNA parts facilitates hierarchical assembly to further increase the number of TUs achieved, although no construct

with more than 3 TUs was described in the original research paper. For the cloning of shorter constructs, connectors spanning across multiple unused positions need to be included. The major disadvantage of the EMMA kit is that in order to build a functional construct, assembly of at least 9 parts occupying different positions is required. This framework complicates the assembly of single TU constructs and lowers DNA assembly reaction efficiency (40% correct hits per a 23-part assembly), in exchange for enabling the cloning of longer and more complex constructs. EMMA comprises a good number of parts for building mammalian synthetic constructs but lacks dedicated parts for CRISPR-based engineering. Many of the DNA parts used in the research paper are not included in the available toolkit plate. This increases the need for users to first clone their own set of genetic parts before implementing them in DNA assemblies, a process often referred to as “domestication”.

MTK toolkit. The Mammalian Toolkit (MTK) is a more recent addition[22]. This comprehensive kit is described with a library of over 300 DNA parts, all of which can be rapidly combined to build complex genetic circuits. The MTK is based on the previously developed Yeast Toolkit (YTK)[13] inspired by MoClo[12], the first Golden Gate-based toolkit. It requires the assembly of only 8 unique parts for a functional transcriptional unit, which in turn can be immediately used in additional sequential DNA assembly rounds to build plasmids carrying up to 9 TUs, thus making the assembly of simple constructs more straightforward when compared to EMMA. Interestingly, parts that have been built for the YTK can be also reused in the MTK, potentially increasing the pool of parts. The kit further expedites the experimental workflow by supporting DNA delivery for both transient and stable expression by the change of only one component in the assembly. It also includes a way to integrate standardised Cas9-sgRNA systems. Depending on how this work is translated into an accessible toolkit plate, it could provide a wealth of genetic parts to many laboratories.

MoPET toolkit. The Modular Protein Expression Vector (MOPET) toolkit is a MoClo-based toolkit designed for expression of recombinant proteins in eukaryotic systems[25]. Although in the reference paper a comprehensive collection of peptide tags, linkers and signal peptides is presented, the toolkit is not readily available as a distributable kit, and this makes its adoption difficult. Should this kit be made available in the future, it would represent a valuable asset for building combinatorial libraries of hard-to-express proteins carrying different tags and linkers, which can be then screened for optimised expression.

mMoClo toolkit. mMoClo is a toolkit based on MoClo but adapted for use in mammalian cells. With respect to MoClo, it introduces an additional ‘part positioning’ vector to accommodate insulating elements, which are often important for designing mammalian synthetic constructs. As with the MoPET toolkit, this system, while promising, is not advertised for distribution, limiting its use[26].

Multi-Kingdom toolkit. The Multi Kingdom (MK) toolkit uses a multi-kingdom platform[27]. Six Level I plasmids are cloned into Level II plasmids, and then hierarchically assembled in Level III vectors. Level II and Level III plasmids are of six different series differing for plasmid origins of replication that allow cloning into different organisms. Moreover, this toolbox benefits from containing additional plasmids that make it compatible with Gateway cloning, which is described below.

1.2 Gibson and Gateway Assembly based toolkits

Cloning toolkits based on Gibson Assembly benefit from it being a scar-less cloning method capable of combining multiple DNA fragments in one quick isothermal reaction (Box 1, Figure IB).

GMAP toolkit. The Gibson-based Modular Assembly Platform (GMAP) is a toolkit that allows stable integration of constructs by lentiviral vector as well as by targeted integration into the Rosa26 locus (Table 2)[28]. The toolbox facilitates the construction of relatively simple genetic constructs (e.g., 2 TUs) with the possibility to increase the complexity of the constructs by modifying the receiver plasmid with additional parts using traditional cloning methods. The requirement to modify plasmids for more complex constructs makes this toolkit inconvenient. Modularity is achieved here by use of five common overlapping sequences to specify part positions. However, these overlapping sequences remain between DNA parts in the final construct effectively leaving large ‘scars’ between modules. Despite its efficiency in the assembly of large DNA fragments, very few Gibson-based toolkits have been developed so far, probably due to the lack of modularity inherent to Gibson cloning. On the contrary, Gateway Assembly can be more easily implemented in toolkits, the most interesting being the pMVP and pMAGIC toolkits (Box 1, Figure IC).

pMVP is a multisite Gateway-based platform useful to build multi-component vectors[29]. Although pMVP enables the construction of small constructs of maximum 4 parts, the kit comprises a collection of parts useful for more advanced applications, such as protein tagging, as well as the cloning of fusion proteins. Furthermore, it includes many destination vectors that can be used to enable both transient and stable expression. This gives the user the freedom to choose between many delivery strategies according to their needs. The **pMAGIC toolkit** can be used with pMVP destination vectors to build useful constructs for epigenetic control of gene expression and CRISPR engineering. Other Gateway Assembly-based toolkits have been developed for more specific applications and these are summarised in Table 2.

1.4 Chaining Cloning-based toolkits

The first chaining-based cloning method developed was BioBrick Assembly[30] (Box 1, Figure ID). BioBricks were however not suitable for cloning of fusion proteins as the scars would cause frame-shifts in the protein assembly and insertion of a stop codon[31]. Bgl-Bricks[31] were then developed to address this problem by

use of a different combination of restriction enzymes (BglII and BamHI), resulting in a scar sequence coding for a glycine-serine dipeptide, which is commonly used for fusion linkers. Collections of BioBrick plasmids for specific uses have been developed; for example, Constante and colleagues[32] developed a BioBrick library for mammalian cell engineering. Such curated collections are very useful to synthetic biologists, as the large and diverse BioBricks Registry of Standard Biological Parts is hard to navigate for high quality DNA parts[33].

MXS-chaining toolkit. The MluI, XhoI, SalI (MXS) toolkit is based on the three named restriction sites that allow iterative chaining of individual parts via chaining cloning[34]. Since MXS is based on chaining cloning, building long arrays of a single module in a single reaction is very straightforward, with no limit to the number of allowed repetitions. This is particularly useful when trying to embed multiple transcription factor binding sites in a synthetic promoter, for example. Since it relies on chaining-based cloning, the assembly of long genetic circuits is extremely time consuming. The low number of parts originally included relegates MXS to simple applications, making it unsuitable for the construction of multigene constructs for mammalian cells. A summary of available chaining cloning-based toolkits is provided in Table 2.

2 Cloning Toolkit Implementation into Workflows

Construction of plasmids is an essential part of the Design-Build-Test cycle adopted widely in engineering and design. When required for the testing of genetic constructs, the chosen method of construction can determine how rate-limiting the design and build stages are. Here we compare the number of available parts for the design phase and time investment required to implement representative kits from the various cloning strategies described above (Figure 1).

One of the biggest time investments for all kits is the necessary domestication of new parts, which takes at least 4 days for each new part (1 day for part PCR amplification/gene synthesis, domestication and cloning into a plasmid in *E. coli*, 1 day for single colony inoculation and growth, and 2 days for plasmid purification and sequence validation). Domestication therefore increases the generation time of a desired construct at the build stage.

While some kits do not have limits on the length of construct, some kits such as pMVP and GMAP can only accommodate two transcriptional units. Generally, the bigger the construct, the more complex the design stage appears. Despite greater design considerations with increasing construct size, Golden Gate toolkits, namely EMMA and MTK, appear to be less impacted by construct size at the build and validation stages. MTK is slightly more efficient compared to EMMA, especially when increasing the number of TUs, due to a simpler design strategy and increased number of available parts. It also provides a higher initial number of parts which make the design stage more efficient.

3. TOOLKITS FOR SPECIFIC APPLICATIONS

While the toolkits described so far were designed to facilitate modern cloning methods, *ad hoc* designs may be needed for a given applications and in this section we provide an overview of available tailored toolkits.

3.1 CRISPR-based toolkits

The CRISPR-Cas9 technology has drastically improved our ability to genetically engineer living cells[35, 36]. The simplicity and versatility of engineering this system has accelerated its use in a wide variety of functions from creation of gene knockouts and knock-ins to tuneable levels of expression. Therefore, it is not surprising that CRISPR toolkits have been designed to facilitate its use. Most toolkits have been developed to allow insertion of DNA fragments and editing at specific genomic locations [43, 44] but other systems have been described for alternative applications, such as the imaging of DNA 3D conformational changes [42]. Information about available CRISPR toolkits is given in Table 3. Although CRISPR-based methods have proven to be a useful tool for manipulating DNA by only requiring the design of guide RNAs, the use of Cas9 has been shown to have off-target effects. While more complex at the design stage, Transcription activator-like effector (TALE) DNA binding proteins may have more desirable specificity, with less prominent off-target effects[37, 38].

3.2 TALE toolkits

Many toolkits for the assembly of tailor-made transcription activator-like (TAL) effectors with custom DNA specificity have been made available to the scientific community[39]. We provide a summary of the available TALE-based toolkits in Table 3. Detailing these is out of the scope of this review, as of today many of these systems have been replaced by more flexible and tuneable alternatives, including the CRISPR-based toolkits described above.

3.3 Other toolkits

Further cloning toolkits have been developed to provide frameworks for additional functionalities useful for mammalian synthetic biology, such as The Composable Mammalian Elements of Transcription (**COMET**). The COMET toolkit has been designed to build orthogonal Zinc Finger (ZF)-based transcription factors for synthetic induction or repression of gene expression[40]. The ZF-based TFs are easily engineerable to obtain chemically inducible TFs. This standalone toolkit will likely be of great use for mammalian synthetic biology as it will enable user to engineer tunability of gene expression into their cells.

Cloning of RNA-binding proteins [41] and engineering of protein-protein interaction [42] have also been recently demonstrated and can provide control of protein intracellular localisation, tuning of transcriptional

rate, and the engineering of inducible control of gene expression. In Table 4 we discern the main features of these and other toolkits developed for specific applications, such as for generating barcoded cells to aid *in vivo* high throughput screening (HTS) of small molecules.

CONCLUDING REMARKS

What does a perfect toolkit look like?

The attractiveness of any cloning strategy depends on several key features, and the implementation of these in a toolkit is what accelerates its adoption into labs, making it easier and faster to share DNA parts and assemble them into desired constructs for testing.

The first and foremost feature is availability. Some of the toolkits presented in this review are either not readily available via sharing sites like Addgene or are only available as separate plasmids rather than as a distributable plate-format kit, making it difficult to get all the parts. A toolkit that is not readily available greatly limits its use in the community and this is reflected in the current difference in the adoption of the accessible MoClo toolkits for bacteria, yeasts, and plants, compared to their mammalian counterpart, mMoClo, which is not distributed by Addgene. Most MoClo kits have “extension packs” (see below), each comprising several ready-to-use parts for the cloning of complex constructs[10, 11, 13, 15, 18, 43], whereas none of these exist for mMoClo.

The second feature is the number of DNA parts present in the kit. While there must be a limit on the number and type of parts that a toolkit can carry, it is also true that a careful design of the cloning workflow can allow the initial toolkit to then be later enriched with many parts by extension packs. These extensions can add on comprehensive libraries useful for diverse purposes (i.e., CRISPR-based gene integration, protein tagging, etc.), so long as they are compatible with the original cloning toolkit. This is exemplified by the most recent published toolkits, such as EMMA and MTK, whose cloning workflow can be adapted to match a user’s needs, and for which the delivery of extension packs can be envisioned. Extension packs are particularly useful when new parts become available and the widespread use of the original toolkit is still desirable. For example, some toolkits include parts for tuneable expression using the traditional TetR system[30, 34]. With the development of more novel tuneable gene regulation systems such as the aforementioned COMET system, their implementation will likely be accelerated by being fully compatible with a widely used toolkit.

An important feature critical for uptake of a toolkit is the level of expertise required to use it and the speed at which it can make desired constructs. While today’s cloning toolkits are generally very straightforward to use even by non-expert users, their speed is a crucial factor, especially when the cloning of large and complex constructs is sought. For example, the MXS toolkit – as well as other chaining cloning-based toolkits - are

greatly limited by their serial nature, as the time required to build synthetic constructs inevitably increases with their complexity.

An interesting further feature would be broad compatibility, so that DNA parts coming from different organisms can be easily re-used or domesticated to join a given toolkit. The direct compatibility of MTK with its yeast counterpart YTK is of great value for groups that work on both yeast and mammalian organisms. Similarly, the multi-kingdom toolkit, although being Golden Gate-based, can also support easy domestication of Gateway cloning parts.

Hybrid cloning frameworks have also been developed where two different cloning methods are combined in one unique workflow. Although this may seem unnecessarily complicated, such frameworks have the advantage of benefitting from the strengths of both cloning methods and partially overcome their weaknesses too. For example, Torella and colleagues[44] reported on the construction of insulated circuits by combining BioBrick assembly and Gibson assembly methods while Guye and colleagues[45] combined Gateway and Gibson Assembly to obtain a framework ideal for the cloning of very large constructs. Here the reliance on Gateway assembly enables increased the modularity in the first cloning steps, and the high efficiency of Gibson cloning for long DNA fragments aids for the final assembly round. By using this cloning workflow, the authors succeeded in cloning a 64 Kbp synthetic construct.

What is missing in cloning toolkits?

One of the biggest problems of today's mammalian cloning toolkits is the scarcity of parts readily available to the user. Even for more comprehensive toolkits like EMMA, the DNA parts provided within the distributed kit are much less in number than the ones presented in the paper. This greatly limits the applications of the toolkits and is probably the biggest barrier to their adoption amongst the mammalian synthetic biology community.

Further work is also required to increase the realm of what is possible in this field (see Outstanding Questions). Important additions to toolkits would include a comprehensive library of DNA parts for inducible gene expression with low background levels, standardised tools for epigenetic control, and a variety of mammalian recombinases that are well-characterised. Moreover, in certain experiments, it is favourable to have parts that work independently. Work in this direction has already created sets of orthogonal regulated promoters by transplanting DNA parts of prokaryotic origin into mammalian cells[46-48]. Such additions to established toolkits would allow the construction of complex genetic circuits with reduced effort needed at the design and build stages.

A longer-term goal of synthetic biology is to build much larger synthetic constructs, chromosomes[49] and genomes, and to do so systematically[50-52]. As our ability to build larger and more complex synthetic DNA

constructs improves-along with novel genome editing tools[53] - the realm of what is possible will be expanded and the pace of discovery accelerated.

Towards this, the use of PCR and enzymes such as exonucleases becomes less desirable in assembly workflows as they are prone to introducing mutation into DNA sequence[54]. Construction methods avoiding mutation-prone steps will likely predominate at the larger scales. In this direction, it is likely that the development and improvement of mammalian artificial chromosomes (MACs) will be a significant technology to link with in the future[55]. Understanding the design rules for construction of MACs could help decipher fundamental attributes of mammalian chromosomes, while also enabling building of complex constructs not yet possible using any of the cloning strategies discussed here.

MACs have been designed by both bottom-up and top-down approaches, yet in both cases the current methods require specialist workflows that are time-consuming and have low yields[56-59]. Once synthetic biology principles such as modularity, standardisation and characterisation are applied to MAC generation, the energy barrier for the user will be reduced and high-throughput and complex DNA construction will become easier and faster. This will greatly increase our engineering capabilities in mammalian systems and open up new lines of research and technology development. Importantly, such revolutionary advances may greatly benefit from newer and more efficient DNA synthesis and assembly methods, allowing researchers to design, build and test large DNA constructs at low costs[54].

Author Contribution

RDB, AZ and FC conceived the paper. RDB, AZ, TE and FC wrote and edited the manuscript.

Conflict of interests

Authors declare no conflict of interest.

FIGURE AND CAPTIONS

Figure 1: Benchmarking of representative cloning toolkits. On the left, the part availability from each representative toolkit is given. For this, the score was assigned where a value of one, two or three was given to toolkits having a low, medium or high repertoire of compatible parts, respectively. On the right, the time to build and test constructs of increasing complexity (1, 3 and 6 transcriptional units) is shown. Time score was calculated by converting the time required (in days) into a 10-point score system with 5 points assigned to each phase. Higher score means more time is required for the completion of the specific phase. In particular, the build phase comprises part domestication, assembly reaction, and bacterial transformation, while the validate phase includes single colony inoculation and either digestion analysis or sequence verification. Assumptions have been made considering i) UK standard delivery time (1 day for sequencing after sample shipping, 1 day for primer receipt after order placement) and ii) use of normal growth rate *E. coli* strains for sub-cloning.

Figure I. Schematic of representative cloning strategies. Golden Gate (A), Gibson (B), Gateway (C) and Chaining (D) assembly methods are shown.

Text box 1. Golden Gate, Gibson, Gateway and Chaining cloning overview

Golden Gate Cloning was first described as a “one pot” reaction and relies on the use of Type IIS restriction enzymes (e.g., BsaI and BsmBI). Type IIS are restriction enzymes which cut outside their asymmetric recognition sequence, thus allowing scarless assembly of DNA parts. Two sequential parts are joined in predefined order when each part in its origin vector is digested with the same Type IIS enzyme exposing specific complementary sequences. Standardised implementation of Golden Gate is based on the MoClo kit (Figure 1A). Advantages include parts can then be easily ligated together in an overhang-dependent manner, with the overhang sequence being independent of the enzyme recognition sequence and the *one-pot* assembly of many parts. However, type IIS restriction sites are frequently present in mammalian promoters and protein-coding regions, making their use time consuming [45]. Also, promiscuous annealing can occur[60].

Gibson Assembly relies on overlapping primers with homologous sequences[61]. After part amplification, primers are chewed back by the T5 exonuclease, leaving single-stranded complementary sequences. Gaps are then repaired *in vitro* by DNA polymerisation and ligation (Figure 1B). The whole reaction occurs isothermally. While there is no need for part domestication, using Gibson requires avoiding sequence similarity within overhangs and sequences that are difficult to amplify by PCR. It is also not suitable for small DNA part (<200 bp) assembly.

Gateway Cloning utilises 6-8 bp proprietary recognition sequences (*att* sequences) and recombination enzymes (BP clonase and LR clonase) for site-specific recombination. Part entry vectors are generated by a BP recombination reaction, which then can be cloned inside destination vectors by means of a LP

recombination reaction[62] (Figure IC). With Creator and Univector[30], Gateway was developed to advance design and building of mammalian constructs and it is now used in many organisms[63, 64]. Gateway is highly efficient and broadly compatible but cost of the reagents and presence of long scar sequences between functional parts are included in the drawbacks.

Chaining Cloning methods uses Type II restriction enzymes to combine DNA fragments flanked by the corresponding restriction sites. Since isocaudomers are adopted for digestion of intra-part regions, after ligation, scars are regenerated that cannot be recognised by any of the restriction enzymes (Figure ID). The newly assembled DNA fragment can then participate in another digestion-ligation reaction, increasing the complexity of the final construct with sequential cloning steps. While being highly efficient and useful in cloning arrays of repetitive modules, chaining methods require a long time for the serial cloning of complex constructs.

Glossary

- **Addgene:** Non-profit plasmid repository that distributes plasmids constructs.
- **Assembly:** Reaction in which many DNA parts are added together; restriction enzymes facilitate the fitting together of parts into a final construct in a (usually) pre-set order.
- **BioBrick:** Standardised DNA parts flanked by specific restriction enzymes (Prefix: EcoRI, XbaI; Suffix: SpeI, PstI) allowing iterative assembly of larger DNA constructs.
- **Design-Build-Test-Learn Cycle:** iterative cycle widely implemented in metabolic engineering. Starting from an initial pool of designs (Design), these are built and implemented in the system of interest (Build), and their functionality tested (Test). The Test phase returns valuable information to further better the quality of the designs which can then be imputed again into the cycle.
- **Domestication:** The process in which a sequence of interest is introduced into a toolkit usually by the addition of prefix and suffix sequences.
- **Gene Editing:** Procedure aimed at modifying the genome with mutations, insertions, or deletions at specific locations.
- **Library:** Collection of DNA parts (e.g., promoters, genes, terminators) that can be used as starting material for the cloning of larger constructs.
- **MACs:** Mammalian Artificial Chromosomes. Long DNA sequences (linear or circular) for assembly of large constructs or the study of chromosomes.
- **Modular:** A method with a simplified design stage, as constructs are created from standardised units. This simplifies processes and reduces chances of errors.
- **MoClo:** Modular Cloning. A method of making golden gate technology modular.
- **Transcriptional unit:** Sequence of DNA which is transcribed into RNA molecule. It includes promoters and termination sequences.
- **Off-target effects:** Unwanted and non-specific changes in the genome sequence that may arise when using gene editing technologies, such as CRISPR-Cas9.
- **Orthogonality:** Having little or no cross reactivity with other components in a system, referring to sequences in this context.
- **TFs:** Transcription factors are proteins that bind promoters and alter gene expression output from a transcriptional unit.
- **YTK:** Yeast Toolkit, a MoClo-based toolkit for the high-throughput cloning of constructs for yeast engineering.

Table 1 – Golden Gate toolkits

Name	Purpose	Main parts available	Cloning Efficiency	Parts for inducible gene expression	Parts for CRISPR engineering	Parts for site-specific recombination	Delivery method			Expression	Availability *	References
							Expression plasmid	Retroviral	Lentiviral			
EMMA toolkit	Cloning of constructs for mammalian cell engineering applications	4 promoters, 3 polyA, 2 reporter genes, 1 selection marker, miscellaneous (peptide linkers, peptide tags, p2A, IRES) * *	40% - 100% (Depends on # of parts)	NO*	NO	NO*	X	–	X	Transient and stable	95 plasmids Addgene #1000000119	[21]
MTK toolkit	Cloning of constructs for mammalian cell engineering applications	> 10 promoters, 3 polyA, > 9 reporter genes, 3 resistance genes, miscellaneous (peptide linkers, peptide tags, p2A, IRES, transgenes, HR for knock-in)	75% - 100%	YES	YES	YES	X	X	X	Transient and stable	341 plasmids Addgene https://www.addgene.org/browse/article/28197510	[22]
MoPET	Cloning of C- and N-terminal tagged proteins for eukaryote cell engineering		>95%	NO	NO	NO	X	–	–	Transient	No	[25]
Multi-kingdom toolkit	Cloning of constructs for multi-kingdom cell engineering		Not published	NO	NO	NO	X	–	–	Transient	No	[27]

* Up to date pricing information can be found on the Addgene website by following the published kit numbers

** Other parts may be available upon request from corresponding authors.

Table 2 – Chaining, Gibson and Gateway toolkits

Name	Purpose	Main parts available	Cloning Efficiency	Parts for inducible gene expression	Parts for CRISPR engineering	Parts for site-specific recombination	Delivery method			Expression	Availability *	References
							Expression plasmid	Retroviral	Lentiviral			
Chaining Toolkits												
MXS-chaining toolkit	Cloning of constructs for mammalian cell engineering applications	10 promoters, 3 polyA, 14 reporter genes, 4 resistance genes, miscellaneous (degradation tags, p2A, linkers, transgenes)	>90%	YES	NO	YES	X	–	–	Transient and stable	55 plasmids Addgene #1000000064	[34]
Gibson Toolkits												
GMAP toolkit	Cloning of integrative constructs for mammalian cell engineering applications	8 promoters, 10 reporter genes, 3 resistance genes (transgenes)	40% to 99% dep on n° of parts (see Suppl. fig 1)	YES	NO	YES	–	X	X	Stable	47 plasmids Addgene (https://www.addgene.org/browse/article/17904/)	[28]
Gateway Toolkits												
White et al.	Cloning of retroviral vectors for in vivo expression in neuronal cells	4 promoters, 5 reporter genes, miscellaneous (transgenes)	Not published	NO	NO	NO	–	X	X	Stable	25 plasmids Addgene (https://www.addgene.org/browse/article/4511/)	[65]
pGCS Kit	Cloning of C- and N-terminal tagged proteins for cell engineering in vertebrates	pCS2+ backbones carrying various protein tags (HA, MYC, FLAG, GST, His, eGFP epitope)	Not reported; selection strategies adopted	NO	NO	NO	X	–	–	Transient	12 plasmids Addgene #1000000107	[66]
MuLE toolkit	Cloning of lentiviral constructs for mammalian cell engineering applications	7 promoters, 6 reporter genes, 2 resistance genes	25-90%	YES	YES	YES	–	–	X	Stable	96 plasmids Addgene #1000000060	[67]
Yoshimatsu et al.	Cloning of knock-in vectors for mammalian cell engineering applications	7 reporter genes, 4 resistance genes, 2 negative selection cassettes	Not reported; selection strategies adopted	NO	YES	YES	X	–	X	Stable	Request from the corresponding authors	[68]
pMVP and pMAGIC toolkits	Cloning of constructs for mammalian cell engineering applications, including epigenome editing	6 promoters, 5 reporter genes, 5 resistance genes, miscellaneous (protein tags, p2A, IRES, transgenes, dCas9 fusion proteins)	Not reported; selection strategies adopted	YES	YES	YES	X	X	X	Transient and stable	191 plasmids Addgene #1000000155	[29]

* Up to date pricing information can be found on the Addgene website following the published kit numbers

Table 3 – Gene Editing toolkits

CRISPR Based Toolkits							
Name	Availability *	General Purpose	Key Features			References	
CRISPRainbow	Addgene #1000000079	Labelling of genomic loci	-Simultaneous tagging of 6 different loci - Combines 3 pairs of RNA hairpins/binding domains -Optimised for DNA labelling			[69]	
CRISPaint	Addgene # 1000000086	Modular integration into genomic sites	-Modular integration of DNA to user defined locus by use of cellular NHEJ, with or without antibiotic selection -Complete system consists of 3 plasmids: Target selector, frame selector and universal donor -Only Target selector needs to be made per experiment -Contains, luciferase (x2), fluorescent proteins (x4), epitope tags (x5), destabilization tag and protein-protein interaction tag.			[70]	
Multiplex CRISPR/Cas9 Assembly Systems	Addgene #1000000055 + Multiplex CRISPR dCas9/FokI-dCas9 Accessory Pack Addgene #1000000062	Cloning of plasmids for multiplexed genome editing	-Generation of cassettes with up to 7 gRNAs and Cas9 nickase or nuclease - FokI-dCas9 and guide plasmids for increased specificity and editing up to 2 genomic loci. -Golden Gate based method (2 steps) -gRNA sequences are generated by annealing oligos -gRNA expressed from U6 promoter -Can be used in cell lines and mice embryos			[71]	
GeneWeld	Addgene #1000000154	Integration into genomic sites	- Vector series for microhomology-mediated end joining knock-ins in zebrafish and mammalian cells			[72]	
TAL Based Toolkits							
Name	Availability *	General Purpose	Key Features				References
			Cloning Method	RVD Types	Number of RVDs	Other	
STAR	Addgene # 1000000117	Assembly of TALEs	Gibson (Hierarchical)	Classical and non-RVD variations	16.5	Destination vector toolbox contains tools to activate, repress or control epigenetics of target	[73]
FusX	Addgene #1000000063	Assembly of TALEs	Golden Gate (2 stages)	NI, NN, HD, NK, NN, NH, NG	14.5-16.5	Backward compatible with Golden Gate TALEN and TAL effector kit 2.0	[74]
Ding et al.	Addgene # 1000000034	Generation of TALENs for creation of disease models	Golden Gate	HD, NG, NI, NN and NK	15	Optimised for mammalian delivery Library of 832 tetramer or trimer TAL repeats	[75]
Sakuma et al.	Accessory kit Addgene #1000000030	Assembly of TALENs	Golden Gate based	HD, NI, NG, NN NH and NK	31	Mammalian cell evaluation	[76]
LIC TAL effector assembly kit	Addgene # 1000000023	Assembly of TALEs	Hierarchical two-step process.	NG, NI, HD, and NK	18.5	High throughput Library contains 2mer and 5mer combinations	[77]
Golden Gate TALEN and TAL Effector kit 2.0	Addgene # 1000000024	Assembly of TALEs	Golden Gate (2 stages)	HD, NI, NG, NK and NN	12-31	Yeast based evaluation	[78]
Zhang et al.	Addgene #1000000019	Assembly of TALENs	Golden Gate like	HD, NI, NG, NK and NN	13	Steps include PCR of modules	[37]

* Up to date pricing information can be found on the Addgene website following the published kit numbers

Table 4 – Other toolkits

Name	Availability *	General Purpose	Key Features	References
Donahue et al. (COMET Toolkit)	Addgene # 138717- 138747 138749-138940 139212-139278	Toolkit of orthogonal TFs and Promoters	<ul style="list-style-type: none"> - 44 activating and 12 inhibitory zinc-finger transcription factors (TFs) - Inducible TFs - 83 cognate promoters with YB_TATA minimal promoter - Framework readily accommodates new parts. - Suitable for use in transient or stable expression systems 	[40]
Lebar et al. (NICP set)	Addgene # 135983- 135989 (7 parts)	Engineering of protein interactions	<ul style="list-style-type: none"> - CC pairs can be used for multiplexable localization of response - Gene transcription can be controlled using interaction strength and variable stoichiometry - Inducible by small molecules - Orthogonal - Each CC is 33 amino acids long - Minimal off-target binding - 6 pairs of CCs described in paper. One pair P3-P4 available in toolkit. 	[42]
Grüner et al.	Addgene # 83678 - 83773	Platform for multiplexed screening of molecules	<ul style="list-style-type: none"> - In-vivo screening of 100-1000 of compounds - Uses molecular barcoding and high-throughput sequencing - Kit includes 96 distinct barcodes 	[79]
Dodev et al.	Addgene # 50365 –61889	Expression of recombinant antibodies	<ul style="list-style-type: none"> - One-step assembly of antibody DNAs into one mammalian expression vector - Starting from DNAs encoding the desired variable and constant regions - Assembly of antibodies of different isotypes and specificity. Kit contains IgG, IgA, IgE and IgM expression vectors 	[80]
Abil et al. (PUF Assembly Kit)	Addgene #1000000051	Cloning of custom RNA binding proteins	<ul style="list-style-type: none"> - Transient delivery and expression - Plasmids comprising 4 variations of each of the 8 PUF repeat modules - Cloning of PUF proteins with custom RNA binding specificity 	[41]

* Up to date pricing information can be found on the Addgene website following the published kit numbers

Table 5 – Comparison of cloning methods

<i>Cloning Method as implemented in toolkits</i>	<i>Pros</i>	<i>Cons</i>
Golden Gate	-Simple design stage -Reduced verification when all parts domesticated	-Frequency of Type IIs restriction sites inside parts -4bp scars between parts*
Gibson	-Highly efficient in cloning large constructs	-Large scars between parts*
Chaining	-Easy assembly of repetitive sequences	-Long time required for the cloning of larger constructs
Gateway	-High efficiency -Broad compatibility	-High cost -Scars between parts

*Although Golden Gate and Gibson can enable scarless cloning, their implementation in cloning toolkits can only happen at the expense of introducing scar sequences between parts.

References

1. Slomovic, S. et al. (2015) Synthetic biology devices for in vitro and in vivo diagnostics. *Proc Natl Acad Sci U S A* 112 (47), 14429-35.
2. Austin, M.J. and Rosales, A.M. (2019) Tunable biomaterials from synthetic, sequence-controlled polymers. *Biomater Sci* 7 (2), 490-505.
3. Qian, Z.G. et al. (2020) Synthetic biology for protein-based materials. *Curr Opin Biotechnol* 65, 197-204.
4. Caliendo, F. et al. (2019) Engineered Cell-Based Therapeutics: Synthetic Biology Meets Immunology. *Front Bioeng Biotechnol* 7, 43.
5. Wu, M.R. et al. (2019) Engineering advanced cancer therapies with synthetic biology. *Nat Rev Cancer* 19 (4), 187-195.
6. Carbonell, P. et al. (2018) An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Commun Biol* 1, 66.
7. Nielsen, J. and Keasling, J.D. (2016) Engineering Cellular Metabolism. *Cell* 164 (6), 1185-1197.
8. Chao, R. et al. (2017) Engineering biological systems using automated biofoundries. *Metab Eng* 42, 98-108.
9. Casini, A. et al. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nat Rev Mol Cell Biol* 16 (9), 568-76.
10. Moore, S.J. et al. (2016) EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. *ACS Synth Biol* 5 (10), 1059-1069.
11. Iverson, S.V. et al. (2016) CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth Biol* 5 (1), 99-103.
12. Weber, E. et al. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS One* 6 (2), e16765.
13. Lee, M.E. et al. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth Biol* 4 (9), 975-86.
14. Shaw, W.M. et al. (2019) Engineering a Model Cell for Rational Tuning of GPCR Signaling. *Cell* 177 (3), 782-796 e27.
15. Obst, U. et al. (2017) A Modular Toolkit for Generating Pichia pastoris Secretion Libraries. *ACS Synth Biol* 6 (6), 1016-1025.
16. Perez-Gonzalez, A. et al. (2017) Adaptation of the GoldenBraid modular cloning system and creation of a toolkit for the expression of heterologous proteins in yeast mitochondria. *BMC Biotechnol* 17 (1), 80.
17. Hahn, F. et al. (2020) A modular cloning toolkit for genome editing in plants. *BMC Plant Biol* 20 (1), 179.
18. Engler, C. et al. (2014) A golden gate modular cloning toolbox for plants. *ACS Synth Biol* 3 (11), 839-43.
19. Sarrion-Perdigones, A. et al. (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One* 6 (7), e21622.
20. Mathur, M. et al. (2017) Mammalian synthetic biology for studying the cell. *J Cell Biol* 216 (1), 73-82.
21. Martella, A. et al. (2017) EMMA: An Extensible Mammalian Modular Assembly Toolkit for the Rapid Design and Production of Diverse Expression Vectors. *Acs Synthetic Biology* 6 (7), 1380-1392.

22. Fonseca, J.P. et al. (2019) A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells. *ACS Synth Biol* 8 (11), 2593-2606.
23. Ede, C. et al. (2016) Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. *ACS Synth Biol* 5 (5), 395-404.
24. Engler, C. et al. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3 (11), e3647.
25. Weber, E. et al. (2017) Modular Protein Expression Toolbox (MoPET), a standardized assembly system for defined expression constructs and expression optimization libraries. *PLoS One* 12 (5), e0176314.
26. Duportet, X. et al. (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res* 42 (21), 13440-51.
27. Chiasson, D. et al. (2019) A unified multi-kingdom Golden Gate cloning platform. *Sci Rep* 9 (1), 10131.
28. Akama-Garren, E.H. et al. (2016) A Modular Assembly Platform for Rapid Generation of DNA Constructs. *Sci Rep* 6, 16836.
29. Haldeman, J.M. et al. (2019) Creation of versatile cloning platforms for transgene expression and dCas9-based epigenome editing. *Nucleic Acids Res* 47 (4), e23.
30. Shetty, R.P. et al. (2008) Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* 2, 5.
31. Anderson, J.C. et al. (2010) BglBricks: A flexible standard for biological part assembly. *J Biol Eng* 4 (1), 1.
32. Constante, M. et al. (2011) A biobrick library for cloning custom eukaryotic plasmids. *PLoS One* 6 (8), e23685.
33. Peccoud, J. et al. (2008) Targeted development of registries of biological parts. *PLoS One* 3 (7), e2671.
34. Sladitschek, H.L. and Neveu, P.A. (2015) MXS-Chaining: A Highly Efficient Cloning Platform for Imaging and Flow Cytometry Approaches in Mammalian Systems. *PLoS One* 10 (4), e0124958.
35. Moon, S.B. et al. (2019) Recent advances in the CRISPR genome editing tool set. *Exp Mol Med* 51 (11), 1-11.
36. Adli, M. (2018) The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9 (1), 1911.
37. Zhang, F. et al. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* 29 (2), 149-53.
38. Lee, H.B. et al. (2016) Genome Engineering with TALE and CRISPR Systems in Neuroscience. *Front Genet* 7, 47.
39. Moore, R. et al. (2014) Transcription activator-like effectors: a toolkit for synthetic biology. *ACS Synth Biol* 3 (10), 708-16.
40. Donahue, P.S. et al. (2020) The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nat Commun* 11 (1), 779.
41. Abil, Z. et al. (2014) Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *J Biol Eng* 8 (1), 7.
42. Lebar, T. et al. (2020) A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nat Chem Biol* 16 (5), 513-519.
43. Gantner, J. et al. (2018) Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system. *PLoS One* 13 (5), e0197185.
44. Torella, J.P. et al. (2014) Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acids Res* 42 (1), 681-9.
45. Guye, P. et al. (2013) Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res* 41 (16), e156.
46. Gossen, M. and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* 89 (12), 5547-51.
47. Kallunki, T. et al. (2019) How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells* 8 (8).
48. Stanton, B.C. et al. (2014) Systematic transfer of prokaryotic sensors and circuits to mammalian cells. *ACS Synth Biol* 3 (12), 880-91.
49. Annaluru, N. et al. (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science* 344 (6179), 55-8.
50. Boeke, J.D. et al. (2016) GENOME ENGINEERING. The Genome Project-Write. *Science* 353 (6295), 126-7.
51. Gibson, D.G. et al. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319 (5867), 1215-20.

52. Hutchison, C.A., 3rd et al. (2016) Design and synthesis of a minimal bacterial genome. *Science* 351 (6280), aad6253.
53. Liu, R.M. et al. (2019) Synthetic chimeric nucleases function for efficient genome editing. *Nat Commun* 10 (1), 5524.
54. Hughes, R.A. and Ellington, A.D. (2017) Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harb Perspect Biol* 9 (1).
55. Martella, A. et al. (2016) Mammalian Synthetic Biology: Time for Big MACs. *ACS Synth Biol* 5 (10), 1040-1049.
56. Kazuki, Y. et al. (2011) Refined human artificial chromosome vectors for gene therapy and animal transgenesis. *Gene Ther* 18 (4), 384-93.
57. Logsdon, G.A. et al. (2019) Human Artificial Chromosomes that Bypass Centromeric DNA. *Cell* 178 (3), 624-639 e19.
58. Kotzamanis, G. et al. (2005) Construction of human artificial chromosome vectors by recombineering. *Gene* 351, 29-38.
59. Kouprina, N. et al. (2018) Human Artificial Chromosome with Regulated Centromere: A Tool for Genome and Cancer Studies. *ACS Synth Biol* 7 (9), 1974-1989.
60. Potapov, V. et al. (2018) Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth Biol* 7 (11), 2665-2674.
61. Gibson, D.G. et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6 (5), 343-5.
62. Hartley, J.L. et al. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res* 10 (11), 1788-95.
63. Kirchmaier, S. et al. (2013) Golden GATEway cloning--a combinatorial approach to generate fusion and recombination constructs. *PLoS One* 8 (10), e76117.
64. Atanassov, I. et al. (2009) A simple, flexible and efficient PCR-fusion/Gateway cloning procedure for gene fusion, site-directed mutagenesis, short sequence insertion and domain deletions and swaps. *Plant Methods* 5, 14.
65. White, M.D. et al. (2011) A Molecular Toolbox for Rapid Generation of Viral Vectors to Up- or Down-Regulate Neuronal Gene Expression in vivo. *Front Mol Neurosci* 4, 8.
66. Wang, H.Y. et al. (2016) Construction of a series of pCS2+ backbone-based Gateway vectors for overexpressing various tagged proteins in vertebrates. *Acta Biochim Biophys Sin (Shanghai)* 48 (12), 1128-1134.
67. Albers, J. et al. (2015) A versatile modular vector system for rapid combinatorial mammalian genetics. *J Clin Invest* 125 (4), 1603-19.
68. Yoshimatsu, S. et al. (2019) A versatile toolbox for knock-in gene targeting based on the Multisite Gateway technology. *PLoS One* 14 (8), e0221164.
69. Ma, H. et al. (2016) Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat Biotechnol* 34 (5), 528-30.
70. Schmid-Burgk, J.L. et al. (2016) CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat Commun* 7, 12338.
71. Sakuma, T. et al. (2014) Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep* 4, 5400.
72. Wiersen, W.A. et al. (2020) Efficient targeted integration directed by short homology in zebrafish and mammalian cells. *Elife* 9.
73. Gogolok, S. et al. (2016) STAR: a simple TAL effector assembly reaction using isothermal assembly. *Sci Rep* 6, 33209.
74. Ma, A.C. et al. (2016) FusX: A Rapid One-Step Transcription Activator-Like Effector Assembly System for Genome Science. *Hum Gene Ther* 27 (6), 451-63.
75. Ding, Q. et al. (2013) A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* 12 (2), 238-51.
76. Sakuma, T. et al. (2013) Efficient TALEN construction and evaluation methods for human cell and animal applications. *Genes Cells* 18 (4), 315-26.
77. Schmid-Burgk, J.L. et al. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol* 31 (1), 76-81.
78. Cermak, T. et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39 (12), e82.

79. Gruner, B.M. et al. (2016) An in vivo multiplexed small-molecule screening platform. *Nat Methods* 13 (10), 883-889.
80. Dodev, T.S. et al. (2014) A tool kit for rapid cloning and expression of recombinant antibodies. *Sci Rep* 4, 5885.