BASIC: a new Biopart Assembly Standard for Idempotent Cloning provides accurate, single-tier DNA assembly for synthetic biology

Marko Storch[1†], Arturo Casini[1†], Ben Mackrow[1], Toni Fleming[2], Harry Trewhitt[1], Tom Ellis[3], Geoff S. Baldwin*[1]

1. Department of Life Sciences & Centre for Synthetic Biology & Innovation, Imperial College, South Kensington Campus, London, SW7 2AZ, UK
2. Dr Reddy's Chirotech Centre, Milton Road, Cambridge, CB4 0PE, UK
3. Department of Bioengineering & Centre for Synthetic Biology & Innovation, Imperial College, South Kensington Campus, London, SW7 2AZ, UK

ABSTRACT

The ability to quickly and reliably assemble DNA constructs is one of the key enabling technologies for synthetic biology. Here we define a new Biopart Assembly Standard for Idempotent Cloning (BASIC), which exploits the principle of orthogonal linker based DNA assembly to define a new physical standard for DNA parts. Further, we demonstrate a new robust method for assembly, based on type IIs restriction enzyme cleavage and ligation of oligonucleotides with single stranded overhangs that determine the assembly order. It allows for efficient, parallel assembly with great accuracy: 4 part assemblies achieved 93% accuracy with single antibiotic selection and 99.7% accuracy with double antibiotic selection, while 7 part assemblies achieved 90% accuracy with double antibiotic selection. The linkers themselves may also be used as composable parts for RBS tuning or the creation of fusion proteins. The standard has one forbidden restriction site and provides for an idempotent, single tier organisation, allowing all parts and composite constructs to be maintained in the same format. This makes the BASIC standard conceptually simple at both the design and experimental levels.

INTRODUCTION

The ability to build newly-designed DNA constructs easily, quickly and with high accuracy is one of the key enabling technologies of Synthetic Biology[1, 2] and the adoption of a standard format for the assembly of genetic components is part of this vision[1, 3]. The BioBrick standard[4] is a restriction-ligation-based format and its usefulness stems from the principle of idempotency, where assembled parts retain the prefix and suffix of the original, enabling successive rounds of hierarchical cloning. However, a recent survey[5] has highlighted that most synthetic biology researchers now use Gibson's isothermal method for their DNA assembly[6]. This suggests that the advantages of being able to assemble five or more fragments of DNA in parallel and having no forbidden sequences or scars outweighs the usefulness of a widely-adopted standard in the eyes of many researchers.

Although the Gibson method can be adapted to a physical standard framework using synthetic sequences to guide assembly[3, 7, 8], it is mostly used 'ad hoc', with customised parts that are generally prepared via PCR amplifications. This has led to a return to bespoke assembly, where each reaction requires design, optimisation and verification. Furthermore, reliance on PCR can compromise fidelity through errors in amplification and is inefficient for very long sequences or those containing high GC content and repeat sequences. PCR is also difficult to implement in an automated workflow because reactions for individual parts have to be optimised and verified. A recent approach excludes PCR[7], but requires upstream cloning to define downstream assembly order, thus extending the workflow.

Aside from BioBricks, alternative restriction-based standards have been developed, including GoldenBraid[9, 10] and MoClo[11], which are based on the Golden-Gate [12] protocol that employs type IIs restriction enzymes. A common feature of these approaches is that the entry vector of a part defines its position in the final destination vector, so that changing the order of the parts requires an additional round of cloning. Both MoClo and GoldenBraid adopt a tiered approach, which takes advantage of the consistent layout of transcription units: in the first tier of assembly, where elementary parts such as promoters, ORFs and terminators are assembled into transcription units, a fixed and predefined part order is adopted. In this way, the first round of assembly never requires

changing entry vectors. The same strategy cannot be used in the second tier of assembly, where transcription units are assembled into multigene constructs, since it is usually necessary to retain complete freedom of design. Here MoClo adopts a parallel approach, which requires cloning in a different vector for each possible position, while GoldenBraid adopts a sequential approach that minimises the number of vectors necessary but only allows pairwise assembly. It was previously suggested that the relative advantages of MoClo and GoldenBraid were mutually exclusive[9].

To address the limitations of current assembly technologies, we have developed BASIC (Biopart Assembly Standard for Idempotent Cloning), to bring together six key concepts: standard reusable parts; single-tier format (all parts are in the same format and are assembled using the same process); idempotent cloning; parallel (multipart) DNA assembly; size independence; automatability. Our previous assembly strategy was based on Modular Overlap Directed Assembly with Linkers (MODAL)[3], which introduced the concept of computationally derived orthogonal linkers[13]. To address these key concepts we have developed a new method based on robust restriction/ligation reactions to ligate orthogonal oligonucleotide linkers with single stranded overhangs that define the assembly order. To further address many of the requirements of assembling DNA parts and biological pathways[1] we have enabled hierarchical cloning within a single-tier format and demonstrated that the linkers themselves can be used as composable parts encoding RBS sequences or peptide linkers for fusion proteins. This has been achieved within a standard format that facilitates re-use of both linkers and parts.

RESULTS & DISCUSSION

Specification and Design. The core of the BASIC physical DNA standard is constituted by the integrated prefix and suffix sequences (*i*P/*i*S), which were designed to be back compatible with our previous MODAL strategy[3] where they can act as PCR priming sites; alternatively we here define the BASIC assembly method, based on simple robust reactions. The *i*P and *i*S sequences were also designed to ensure compatibility with the creation of fusion proteins, either by BASIC or MODAL, by optimising the amino acid coding of both the short BASIC scars and the full *i*P/*i*S sequences (Figure 1a).

To avoid PCR, yet retain the advantages of overlap directed DNA assembly, we have revisited pre-PCR methods, where oligonucleotides were routinely ligated onto DNA ends to either provide restriction enzyme sites, or compatible sticky ends to direct molecular cloning[14]. The BASIC standard defines two inward-facing BsaI recognition sites to release the parts from a storage vector, leaving a 4 bp scar on the prefix end and a 6 bp scar on the suffix (Figure 1b). Digestion yields different 4 bp overhangs at the prefix and suffix, enabling end-specific ligation. Ligation of partially double-stranded oligonucleotide DNA linkers is performed simultaneously with BsaI digestion (Figure 1c). Non-ligated oligonucleotide linkers are then removed by a purification step to yield linker-adapted parts. Final assembly is achieved by annealing the linker-adapted parts in an ionic buffer at elevated temperature. No ligase is required in the final step and the nicked plasmid generated is readily repaired *in vivo* following transformation. Full details of the protocol and optimisation of the method are provided in Online Supplementary Information.

Figure 1

Linker sequences to guide assembly were an expanded set of 7 linkers based on our previously-used 40% GC content linkers designed by R2oDNA Designer[3, 13] (Supplementary Tables 3&4). We split each 45 bp linker sequence across 2 parts, with each containing a 12 bp double stranded region on

the outer side and sharing the central 21 base single stranded overlap region as top and bottom strands (Figure 1b). Double-stranded regions at the sites of linker ligation are necessary for efficient activity of T4 DNA ligase[15]. The use of 21 bp overlaps enables elevated temperatures during final assembly by complementary annealing, facilitating the kinetics, thermodynamics and specificity of the homology-search process.

Evaluation of efficiency. To evaluate the efficiency and accuracy of our assembly method, and benchmark against our previous Gibson-based work, we generated a number of parts in BASIC format with *i*P and *i*S sequences flanking the part of interest (Supplementary Table 2). This formatting step only ever has to be performed once for any part since the storage plasmid carries no positional information for the DNA assembly process, which is directed through the subsequent choice of linkers. Parts prepared include those that are essential for cell survival (origin of replication (MB1), kanamycin resistance (Kan) chloramphenicol resistance (Cm), combined origin and kanamycin (Kan-MB1), and others that produce fluorescent proteins (GFP and RFP; both as expression cassettes and as separate open reading frame parts).

Benchmarking DNA assembly reactions were performed by creating plasmids in a modular format from this parts library. Constructs comprising 2 to 6 component parts with a single antibiotic marker, and 2 to 7 parts with double antibiotic selection were chosen for evaluation (Figure 2a). All final constructs (apart from D2) contain a fluorescent reporter, and accuracy of assembly was thus evaluated by observing the correct expression of reporters whilst assessing each construct's ability to replicate and confer the appropriate antibiotic resistance.

Figure 2

BASIC assembly reactions were performed four times for each of the 11 designated test constructs following an optimised protocol (Online Supplementary Information). Assembly efficiency was determined from the number of colonies and accuracy as the percentage of colonies with the correct antibiotic resistance expressing the correct fluorescent reporters (Figure 2).

The results of the DNA assembly benchmarking reveal that the efficiency of assembly decreases exponentially with the number of parts involved (Figure 2b). However, even with 6 or 7 parts, reactions routinely returned between 40 and 150 colonies, while 3-4 part assembly routinely returned more than 1000 colonies, demonstrating the overall efficiency of the process.

The more critical measure of DNA assembly is accuracy. With single antibiotic selection there is the possibility that the storage plasmid that carries either the Kan-MB1 composite part or the Kan cassette can return a viable non-fluorescent colony if it is not completely digested in the first step of the protocol. The assembly efficiency decreased exponentially with increasing number of parts, but the number of incorrect assemblies (which includes both white background colonies and colonies with the wrong fluorescent reporters) remained relatively constant. The incorrect assemblies thus became a larger proportion of the colony count, decreasing accuracy (Figure 2c). To address this we included a second antibiotic resistance cassette, chosen so that the final construct could be selected using double antibiotic selection without any of the starting constructs conferring resistance. This significantly reduces the proportion of incorrect assemblies, indicating that these arise largely through carryover of storage plasmids when only a single antibiotic marker is used (Figure 2c). The double antibiotic selection strategy thus provides a significant improvement in the accuracy of BASIC assembly and was therefore adopted as the standard method in subsequent assemblies.

Since the orthogonal linker sequences provide positional watermarks in the final assembly, they may be used to validate assemblies since they provide ideal PCR primer sites. This strategy was used to evaluate the 5-part assembly, demonstrating the flexibility in re-ordering parts simply by changing the linker combinations ligated to each part. We assessed the assembly order of these reactions as well as the seven part construct by performing PCR reactions with a forward primer for the first linker and reverse primers for each of the other 4 or 6 linkers. The PCR products exhibit the anticipated ladder of increasing size demonstrating the correct order and presence of each part in the assembly (Supplementary Figure 1). This provides a useful screening method for DNA assembly verification and because the linkers are standardised, the PCR verification primers are also standardised (Supplementary Table 13). Because the DNA assembly workflow starts with plasmid DNA and does not involve PCR amplification, there is less of an imperative to sequence the final construct following positional verification of the parts, which is especially useful when constructing pathways and libraries.

Hierarchical assembly. In many cases it is advantageous to assemble a limited number of parts together in a module and then combine different modules to create more complex systems or to re-use modules in different assemblies. The single-tier approach of BASIC therefore requires an idempotent method by which the *i*P and *i*S sequences can be recapitulated during DNA assembly. The objective therefore was to encode *i*P and *i*S on linkers attached during DNA assembly, whilst avoiding any modification to the protocol. To achieve this we investigated DNA methylation as a strategy to protect the BsaI site from digestion during the assembly process.

The cognate DNA methyltransferase of the BsaI restriction modification system is a C-5 methyltransferase, but its target within the BsaI recognition sequence is not known[(16)]. We have therefore determined the pattern of methylation protection through *in vitro* digestion of fluorescently labelled oligonucleotides, with each of the 4 cytosine residues within the recognition site methylated in turn. The restriction digests clearly reveal that methylation of the bottom strand only partially protects the DNA from digestion, while methylation of either cytosine in the top strand effectively protects the DNA from digestion by BsaI (Supplementary Figure 2). We therefore propose a general single-tier workflow for BASIC, where *i*P and *i*S are recapitulated around the constructed cargo during assembly by methylation of specified linker oligonucleotides to avoid cleavage during the combined digestion/ligation step (Figure 3a).

Figure 3

To demonstrate this approach we separately constructed GFP and RFP expression cassettes from individual parts encoding a constitutive promoter (J23102) and RBS-ORFs for GFP and RFP: these cassettes were then used in a second round of assembly to construct a dual fluorescence plasmid (Figure 3b). Parallel reactions were also performed with non-methylated linkers to benchmark the efficiency of the idempotent assembly compared to standard linkers [for a detailed list of assembly order see Supplementary Table 7]. The 4-part first round of assembly proceeded with 99% accuracy and an efficiency that was only 10% lower than that with standard linkers (Figure 3c). The expression cassette constructs were then successfully used for construction of the dual reporter plasmid. This demonstrates that methylation of a single cytosine in the BsaI recognition sequence provides sufficient protection against BsaI digestion to enable an idempotent strategy without modification of the protocol. Maintaining the same protocol for all stages of assembly and for all parts ensures an easy workflow for both bench-scale work and automation.

Linkers as composable parts encoding RBS sequences. One feature of synthetic biology is the ability to rationally compose parts to provide either tuneable or predictable behaviour. Using custom RBS sequences to regulate protein translation has become increasingly common[17-20]. The use of synthetic linker sequences provides the opportunity to encode small parts within the linker, such as RBS sequences. In line with our modular standardised approach to DNA assembly, we chose to tune the output of fluorescent reporters by encoding known RBS sequences of different strengths with the expectation that local sequence context would provide additional variability[19]. Four RBS sequences were selected from the iGEM Parts Registry, and encoded onto the double stranded portion of the prefix linker (Figure 4a; Supplementary Table 8). Two linker overhang sequences were designed using R2oDNA Designer software[13] that are suitable for assembly with the 4 RBS sequences used. These two linker overhangs are orthogonal to the other linkers used in this paper and thus it was possible to generate a library of RBS sequences that can be incorporated in two different locations within a single assembly.

Figure 4

To evaluate the tunability of protein expression using RBS-linkers, 4-part assemblies were performed with the four different strength RBS linkers to join a constitutive promoter to a GFP ORF part without an RBS, but with a start codon adjacent to its *i*P (Figure 4b). Two sets of assemblies were performed to evaluate the degree of variation caused by the minor context change produced by changing the overhang sequence in the two sets of RBS linkers (Supplementary Table 8). The four RBS sequences clearly give distinct levels of GFP expression, while there is no significant difference due to the overhang sequence context of linker 1 vs. linker 2 (Figure 4c).

Additionally, we evaluated the potential to perform combinatorial library assembly by including multiple RBS linkers for a single part using a combination of RBS1 and RBS3 in one instance and of all four RBS linkers in the other. To evaluate the combinatorial RBS assemblies, a number of individual colonies were randomly selected from a quadrant of the plate and grown out in culture. Comparison of expression levels for assemblies with a single RBS linker demonstrate that each colony tested exhibited a fluorescence expression within the expected range for the RBS sequences used. An even distribution of all RBS sequences included was also observed, demonstrating that there was no obvious bias between the RBS sequences chosen. All possibilities of RBS variants in the library construction were found within a relatively small number of colonies analysed. Furthermore none of the randomly selected colonies for either the specific or library constructions were incorrect, again demonstrating the overall accuracy of the assembly process.

The constructed sequences were computationally evaluated for predicted expression strength using the reverse mode of the RBS Calculator[18] (Figure 4c; Supplementary Table 15). It is interesting to note that the expected levels of expression follow the anticipated order for RBS 1-3, while RBS 4 gives significantly lower than expected output. For the different linker contexts, the computational analysis predicted a significant difference in protein output when the RBS sequences were combined with the different linker sequences. However, experimental results demonstrate that there was minimal variation in the protein output when the RBS sequences were placed in different linker contexts (Figure 4; Supplementary Table 15). While RBS calculator tools provide a reasonable correlation between prediction and output on a larger sample size[21], our results demonstrate that with a small population accurate prediction remains difficult.

Creation of fusion proteins. To further develop the BASIC approach we designed linkers that enable the fusion of protein parts during the assembly process. The *i*P and *i*S sequences were already

optimised to be compatible with fusion proteins. To realise this we created linkers to provide complete read through of coding sequence to generate peptide sequences that can join two in frame protein ORFs (Fig. 5a). A GFP ORF part was generated omitting the stop codon and with the final codon in frame to *i*S, while an RFP ORF was generated without an RBS and with the Met start codon in frame to *i*P (Supplementary Table 2). Three fusion linkers have been designed to encode peptide fusions with different properties including both flexible and alpha-helical sequences (Supplementary Table 11). Their codon usage was balanced to avoid nucleotide repeats and the sequences were validated with R2oDNA Designer software to ensure compatibility with BASIC.

Figure 5

Constructs expressing GFP and RFP cassettes singly and on the same plasmid were then constructed in addition to test constructs with GFP fused in frame to RFP using the peptide fusion linkers. To demonstrate the functionality of the linkers, cells expressing the protein fusions were grown to mid-log phase and their protein expression analysed by SDS-PAGE, which revealed that all three constructs containing fusion linkers expressed stable GFP-RFP fusion proteins (Figure 5b).

Conclusion. BASIC comprises both a standard format for DNA parts and a new method for efficient parallel assembly. Our standardised assembly reactions can be benchmarked against our previous 4 part assemblies performed using Gibson reactions with the same orthogonal linkers defining the junctions[(3)]. Our previously published Gibson 4-part assembly gave 75% accuracy[(3)], while the similar 4-part BASIC assembly reported here gave 93% accuracy with single antibiotic selection and 99.4% accuracy with double antibiotic selection, and 7-part assembly gave 90% accuracy with double antibiotic selection.

The single-tier format retains the greatest degree of flexibility and simplicity, while the presence of only one forbidden restriction sequence minimises adoption requirements. Operations such as changing the position of a part, or even reversing the direction of a promoter or ORF can easily be accomplished by simply changing the linkers. Additionally the assembly workflow is completely PCR-free, which greatly enhances its reliability, reduces the chances of introducing sequence errors and avoids the limitations of PCR such as repeat sequences or difficult to amplify sequences.

While double antibiotic selection provides a significant improvement in accuracy for larger assemblies, high accuracy and efficiency can be maintained for smaller assemblies of up to four parts with only single antibiotic selection. The mode of implementation can therefore be chosen by the user based on their specific requirements. Alternative strategies to reduce background may also be employed, such as PCR amplification of the part containing the selectable marker, followed by DpnI digestion. In our view these minor improvements on an already very high accuracy did not outweigh the benefits of a uniform workflow for all parts.

The use of orthogonal sequences to direct assembly, together with the BASIC protocol offers significant advantages over existing DNA assembly technologies. We have demonstrated that it is possible to position the same promoter part in different locations with great accuracy and no loss of efficiency. This would not be possible with a scarless method, such as the original Gibson protocol[(6)], the recently reported ligase cycling reaction method or paperclip [(22, 23)], because the repeated DNA sequence homology would misdirect parts in the final assembly.

The ability to assemble parts as small as 153 bp is also of significance and utility: small parts are known to be problematic with assembly methods that rely on exonuclease digestion as they can readily be digested. However, small parts are frequently required for essential functions such as

promoters and BASIC other restriction-ligation based methods[1, 24] can assemble these without problems. BASIC linkers also provide a means to encode biological functions for even smaller parts: we have demonstrated here that RBS parts can be composed on the adapter regions of the oligonucleotide linkers, or the whole linker can be used to code for peptide sequences that generate fusion proteins.

Verification of the final construct in DNA assembly is a critical component of the workflow. The orthogonal linkers employed in BASIC provide effective watermark sequences for this purpose with ideal PCR primer properties. The standardisation of these components enables assembly verification using a limited set of standardised primers with a standard protocol. Despite reductions in sequencing costs at the genome scale, sequence verification of whole plasmid constructs remains costly in terms of both time and money. But since PCR is not used in BASIC, there is less of an imperative to sequence the final construct following verification of presence and position of the DNA parts.

In common with other standards, adoption of BASIC can facilitate the sharing and re-use of parts and this is enhanced here by the single-tier format of our approach. It would also be possible to reuse parts designed for other type IIs methods like Golden Gate simply by changing the linker ligation overhang sequence. While a significant number of oligonucleotides are required for this method, their standardisation and long-term viability means that economies of scale rapidly accumulate as more people within a single laboratory or group of laboratories adopt the methodology (an evaluation of cost is provided in Supplementary Table 16). Furthermore, the robustness and reliability of all the steps in the BASIC protocol will facilitate translation of the workflow to an automated liquid handling platform.

METHODS

BASIC assembly protocol. A full protocol for laboratory use is provided as online supplementary material.

Methods. Full details of all other materials and methods are provided in online supplementary information.

REFERENCES

1. **Ellis, T., Adie, T., and Baldwin, G. S. (2011) DNA assembly for synthetic biology: from parts to pathways and beyond, *Integrative Biology 3*, 109-118.**
2. **Arkin, A. (2008) Setting the standard in synthetic biology, *Nat Biotechnol 26*, 771-774.**
3. **Casini, A., MacDonald, J. T., De Jonghe, J., Christodoulou, G., Freemont, P. S., Baldwin, G. S., and Ellis, T. (2014) One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy, *Nucleic Acids Res 42*, e7.**
4. **Knight, T. Draft Standard for Biobrick Biological Parts.**
5. **Kahl, L. J., and Endy, D. (2013) A survey of enabling technologies in synthetic biology, *J Biol Eng 7*, 13.**
6. **Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nature methods 6*, 343-345.**
7. **Torella, J. P., Boehm, C. R., Lienert, F., Chen, J. H., Way, J. C., and Silver, P. A. (2013) Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly, *Nucleic Acids Res*.**
8. **Guye, P., Li, Y., Wroblewska, L., Duportet, X., and Weiss, R. (2013) Rapid, modular and reliable construction of complex mammalian gene circuits, *Nucleic Acids Res 41*, e156.**

9. Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juarez, P., Fernandez-del-Carmen, A., Granell, A., and Orzaez, D. (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules, *PLoS One 6*, e21622.
10. Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., and Orzaez, D. (2013) GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology, *Plant physiology 162*, 1618-1631.
11. Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs, *PLoS One 6*, e16765.
12. Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes, *PLoS One 4*, e5553.
13. Casini, A., Christodoulou, G., Freemont, P. S., Baldwin, G. S., Ellis, T., and Macdonald, J. T. (2014) R2oDNA Designer: Computational Design of Biologically Neutral Synthetic DNA Sequences, *ACS synthetic biology 10.1021/sb4001323*.
14. Sambrook, J., E.F.Fritsch, and Maniatis, T. (1989) *Molecular Cloning A Laboratory Manual*, 2nd Edition ed., Cold Spring Harbor Laboratory Press, New York.
15. Horspool, D. R., Coope, R. J., and Holt, R. A. (2010) Efficient assembly of very short oligonucleotides using T4 DNA Ligase, *BMC research notes 3*, 291.
16. Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2010) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes, *Nucleic Acids Res 38*, D234-236.
17. Barrick, D., Villanueba, K., Childs, J., Kalil, R., Schneider, T. D., Lawrence, C. E., Gold, L., and Stormo, G. D. (1994) Quantitative analysis of ribosome binding sites in E.coli, *Nucleic Acids Res 22*, 1287-1295.
18. Salis, H. M. (2011) The ribosome binding site calculator, *Methods Enzymol 498*, 19-42.
19. Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression, *Nat Biotechnol 27*, 946-950.
20. Na, D., and Lee, D. (2010) RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression, *Bioinformatics 26*, 2633-2634.
21. Reeve, B., Hargest, T., Gilbert, C., and Ellis, T. (2014) Predicting translation initiation rates for designing synthetic biology, *Front Bioeng Biotechnol 2*, 1.
22. deKok, S., Stanton, L. H., Slaby, T., Durot, M., Holmes, V. F., Patel, K. G., Platt, D., Shapland, E. B., Serber, Z., Dean, J., Newman, J. D., and Chandran, S. S. (2014) Rapid and reliable DNA assembly via ligase cycling reaction, *ACS Synth. Biol. 3*, 97-106.
23. Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A., and French, C. E. (2014) PaperClip: rapid multi-part DNA assembly from existing libraries, *Nucleic Acids Res 42*, e154.
24. Chao, R., Yuan, Y., and Zhao, H. (2014) Recent advances in DNA assembly technologies, *FEMS yeast research*.

ASSOCIATED CONTENT

Supporting Information

Supplementary information, figures and tables are provided to give further details on protocol development and methods. In addition a stand-alone lab protocol is also provided. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

*To whom correspondence should be addressed:
g.baldwin@imperial.ac.uk
†These authors contributed equally to this work

ACKNOWLEDGEMENT

The authors would like to thank the following organisations for funding: EU (KBBE.2011.5-289326; IEF Fellowship 628019); BBSRC (BB/J020036/1); EPSRC (EP/J02175X/1). The authors do not declare any conflict of interest.

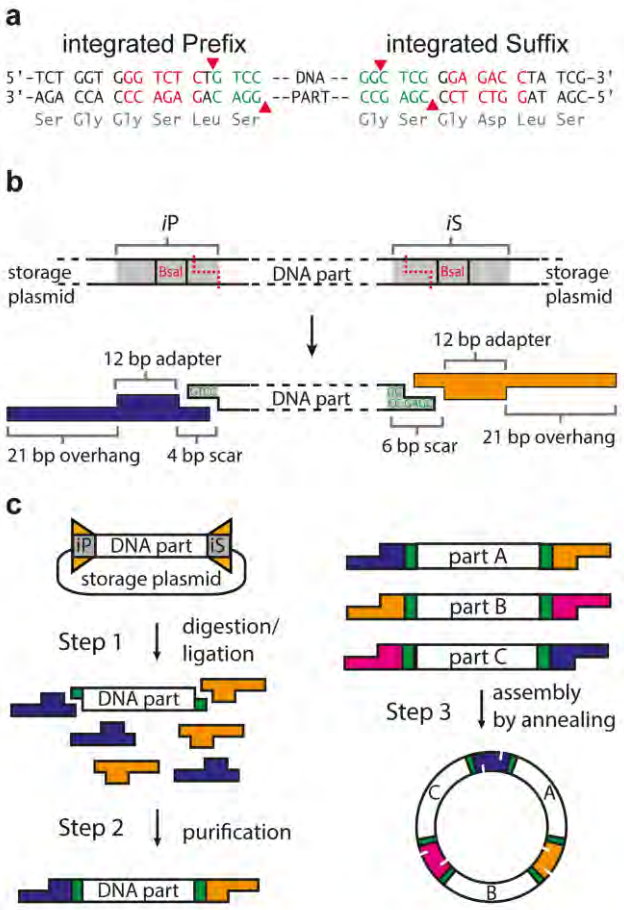

Figure 1. BASIC standard and method**. a**) Sequence of the integrated prefix (*i*P) and suffix (*i*S); the BsaI recognition sequence is shown in red with the cut sites marked (red triangles); the amino acid translation for each codon in the iP/iS are shown. **b**) During the assembly process *i*P and *i*S are cut to produce different sticky ends that enable differential ligation of linkers onto each end. **c**) The BASIC assembly workflow: Step 1: linkers are attached by simultaneous digestion and ligation. Step 2: unligated excess linkers are removed via magnetic bead purification. Step 3: purified linker-adapted parts are mixed and annealed in an ionic buffer to generate the final construct. [Protocol provided in Online Supplementary Information.]

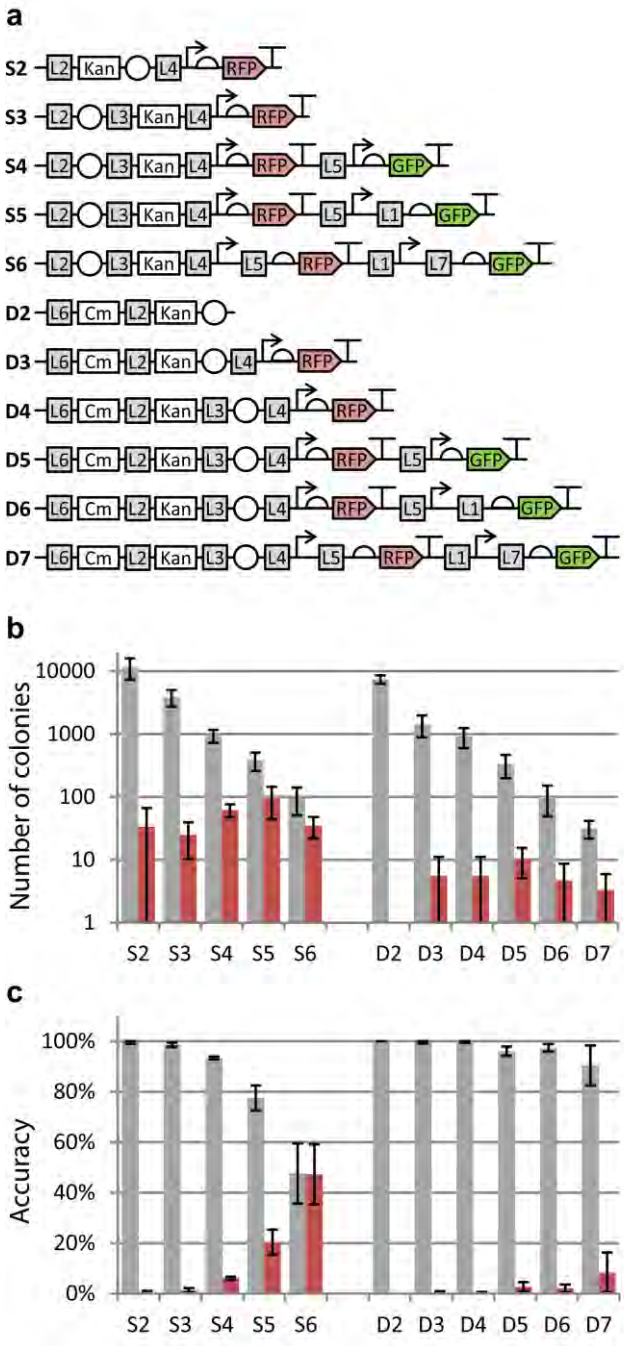

Figure 2. BASIC allows for highly efficient multi part assembly. **a**) Benchmarking DNA assembly reactions were performed creating constructs with 2 to 6 parts using single antibiotic selection (S2-S6) and 2 to 7 parts with double antibiotic selection (D2-D7) [Supplementary Table 1]. **b**) The number of colonies returned from each assembly is shown as the average of 4 repeat reactions with standard error of the mean (SEM; grey bars); the total number of incorrect assemblies that either had no fluorescence or incorrect fluorescence profiles are also shown (red bars). **c**) The accuracy of each assembly reaction was assessed as the percentage of colonies with the correct fluorescence profile for the designed assembly (grey bars); percentage of incorrect assemblies are also shown (red bars). All data is shown as the average of 4 repeat reactions with SEM.

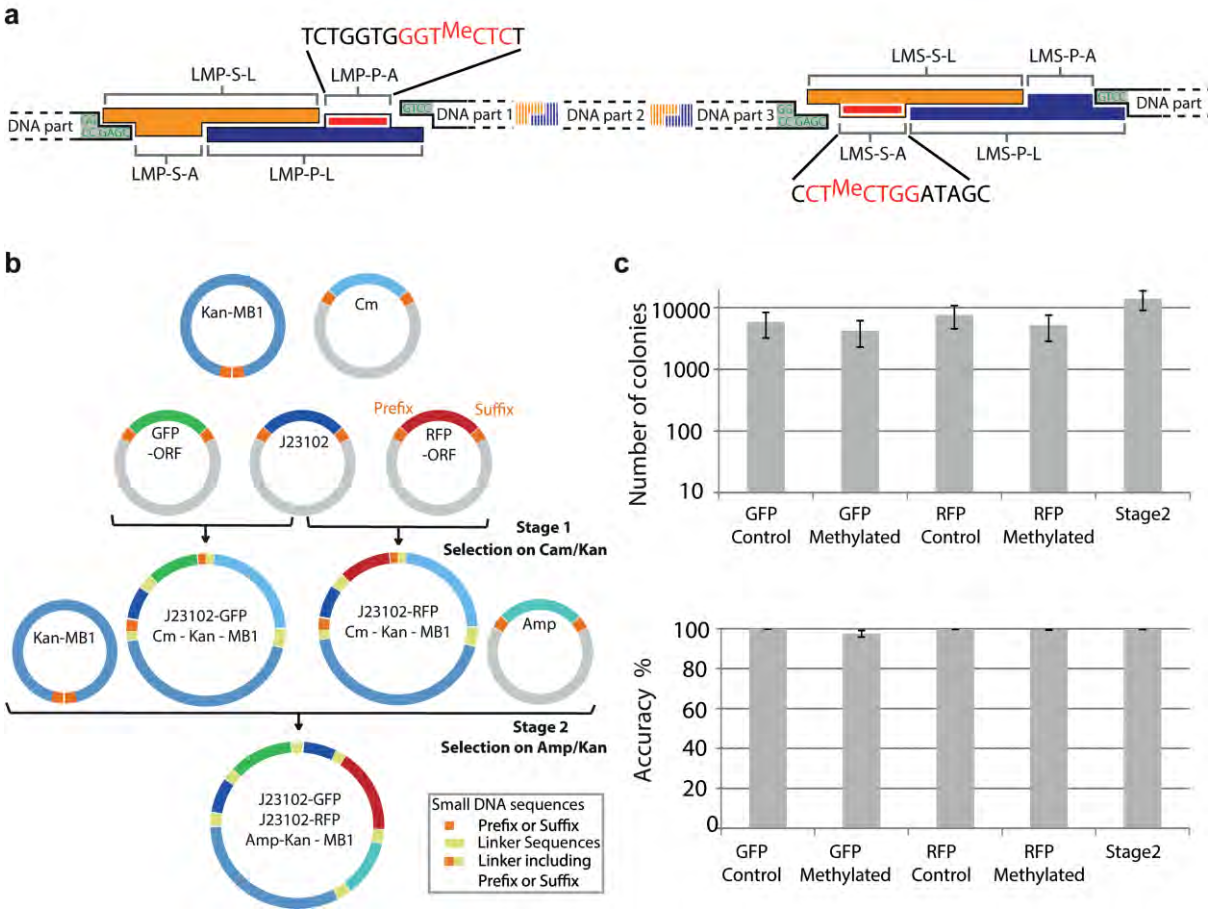

Figure 3. Hierarchical assembly using methylated linkers. **a**) Linker design to recapitulate *i*P and *i*S adjacent to the parts being assembled. The methylated cytosine is located on the adapter oligonucleotide, which prevents digestion of the linker during the assembly process. **b**) Workflow to test idempotent DNA assembly using methylated linkers: in stage 1 GFP and RFP expressing cassettes are assembled flanked by *i*P and *i*S, backbone Kan-MB1 and Cm parts are located outside of *i*P and *i*S and so are not carried through in subsequent assembly rounds. In stage 2 the previously assembled expression cassettes are used to assemble a double fluorescence reporter. **c**) Data from assembly reactions is shown for reactions with methylated linkers, control reactions with non-methylated linkers and stage 2 reactions. Data shown is the average of 4 repeat reactions with SEM for the number of colonies returned and the accuracy, determined as the percentage of colonies with the correct fluorescence profile.

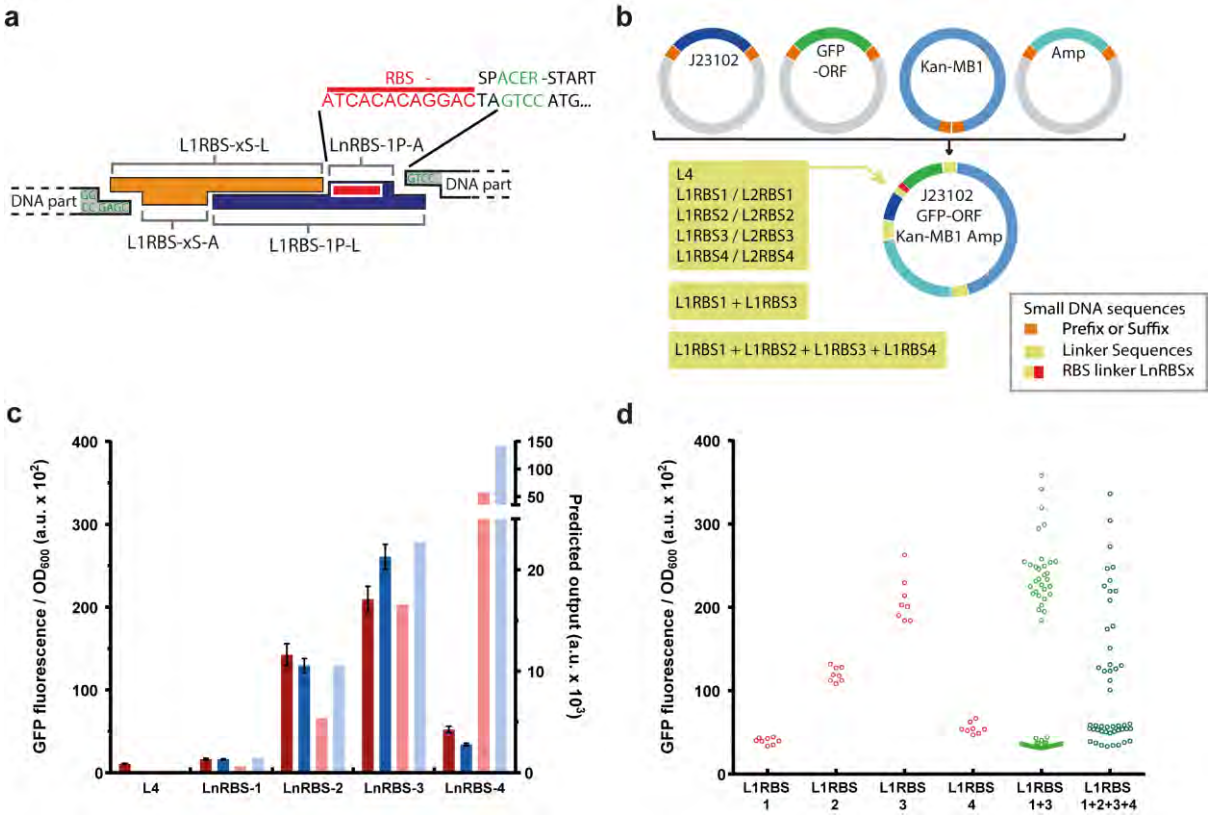

Figure 4. Tuning translation with RBS linkers. **a**) RBS sequences were encoded on the double stranded portion of the ligated linker oligonucleotide with a spacing region to ensure efficient translation [Supplementary Table 8]; the single stranded overlap of the linker does not encode the RBS and multiple RBS sequences were encoded with the same linker homology (LnRBSx, where n denotes the homology type within a series of x different RBS sequences). **b**) Assembly strategy for constructs to test 4 RBS sequences within two different linker contexts and a control linker that does not encode an RBS (L4). **c**) GFP expression was evaluated after 6h growth and is shown normalised to OD_{600} for no RBS control (L4) and RBS1 to RBS4 with linker 1 (dark red bars) and linker 2 (dark blue bars). Predicted expression levels were calculated for all 4 RBS sequences in both linker contexts using the RBS calculator[18] and these are plotted for linker 1 (light red bars) and linker 2 (light blue bars). **d**) Assembly reactions were performed with single RBS linkers and also combinations of both two (RBS1&3) and four (RBS1-4) linkers to create a library of expression variants. Expression levels for randomly selected colonies of these assemblies are shown as a dot plot.

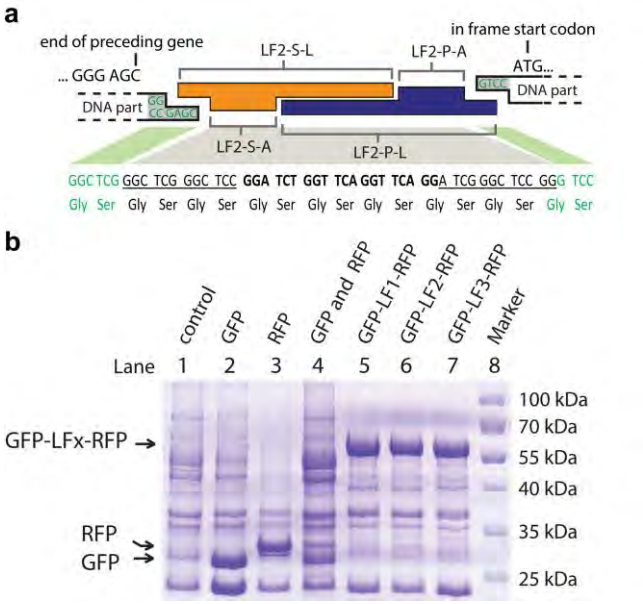

Figure 5. Creating fusion proteins with fusion linkers. **a**) Linkers were designed to provide an in frame polypeptide sequence to fuse two protein sequences, where the upstream gene had no stop codon and the downstream gene was in frame with *i*S. **b**) Constructs were created using 3 different fusion linkers [Supplementary Table 10] between GFP- and RFP-ORFs. SDS-PAGE shows the expression of GFP and RFP in separate cells (lanes 2 and 3) and separately in the same cells (lane 4); the 3 fusion constructs of GFP and RFP are shown in lanes 5-7.