Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination

Lin Zhang1,2, Guoping Zhao1,2,3 & Xiaoming Ding1

1Department of Microbiology and Microbial Engineering, School of Life Sciences, Fudan University, Shanghai 200433, China, 2Key Laboratory of Synthetic biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China, 3Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China.

We describe a site-specific recombination-based tandem assembly (SSRTA) method for reconstruction of biological parts in synthetic biology. The system was catalyzed by Streptomyces phage \( \phi \) BT1 integrase, which belongs to the large serine recombinase subfamily. This one-step approach was efficient and accurate, and able to join multiple DNA molecules in vitro in a defined order. Thus, it could have applications in constructing metabolic pathways and genetic networks.

Synthetic biology is more ambitious than conventional genetic engineering, and aims to design and reconstruct biological systems or even entire bacterial genomes. An essential technique in synthetic biology is the physical assemble of multiple small DNA fragments into large constructs with a defined order and orientation1,2. Traditional methods rely on restriction enzyme digestion followed by ligation, and while this approach works well for the insertion of a single DNA sequence into a vector, as well as being time-consuming, it is often hard to find enough distinct restriction sites for the cloning of multiple DNA fragments. Two techniques, BioBrick\textsuperscript{TM} developed at MIT3 and BglBrick developed at UC Berkeley4, standardized the process of DNA assembly by using standard restriction sites. Biological parts were flanked by four restriction sites (two on each side) and the assembly products of the first round can subsequently be used in the next round, thus permitting the assembly of multiple parts.

The BioBrick\textsuperscript{TM} method is limited by a number of ‘forbidden sites’. This has resulted in the development of a series of ligation-independent cloning (LIC) strategies that offered sequence-independent assembly. These include SLIC (sequence and ligase independent cloning)\textsuperscript{5}, In-Fusion\textsuperscript{TM} (Clontech) technologies\textsuperscript{6}, the ‘Gibson’ isothermal assembly method and CPEC (circular polymerase extension cloning)\textsuperscript{8,9}. These methods are dependent on overlapping homologous sequences at the ends of the DNA fragments1,2. SLIC, In-Fusion\textsuperscript{TM} and the Gibson isothermal assembly method use a mechanism called ‘chew back and anneal’\textsuperscript{2,7,10,11}, while CPEC relies on annealing of complimentary DNA regions and a single high-fidelity PCR cycle\textsuperscript{9}. Using the isothermal assembly method, Gibson and his colleagues successfully assembled three 5-kb DNA molecules using fragments of only 40 bp overlaps. This method was also used to assemble a 900-kb molecule in vitro\textsuperscript{7}. The Gibson method was also effective in gene synthesis, and an entire 16.3-kilobase mouse mitochondrial genome was synthesized through three rounds of assembly from 600 overlapping 60-mers12.

Several competing approaches using type I\( II \)s restriction enzymes have also been developed, such as Golden Gate cloning\textsuperscript{11,14} and Pairwise Selection Assembly (PSA)\textsuperscript{15}. Since type I\( II \)s restriction enzymes cut outside of their recognition sequence, through proper design of the cleavage sites this approach is scarless\textsuperscript{14}. Using Golden Gate cloning, Weber et al\textsuperscript{13} assembled a 33-kb DNA molecule from 44 individual parts in three steps. Blake et al\textsuperscript{15} used PSA to reconstruct a completely synthetic 91-kb molecule in six rounds of assembly. In general, assembly of DNA fragments occurs on three scales – ‘parts to genes’, ‘genes to pathway’ and ‘pathways to genomes’ – and each assembly strategy described above has its own advantages and disadvantages at different scales1,2.

We have developed an alternative strategy for the assembly of multiple DNA segments in one reaction based on \( \phi \) BT1 integrase-mediated recombination in vitro, termed site-specific recombination-based tandem assembly (SSRTA). We previously established a site-specific recombination system based on Streptomyces phage \( \phi \) BT1 integrase-mediated integration in vitro\textsuperscript{16}, and identified 16 pairs of non-compatible \( attB \) and \( attP \) recombination sites\textsuperscript{17}. The \( \phi \) BT1 integration system is highly efficient and accurate because no recombination can occur between
att sites with different central di-nucleotides. The SSRTA method is sequence-independent, and excellent for the combinatorial assembly of multiple biological parts in a defined order, even with those with high GC contents (70%). The utility of SSRTA was demonstrated in the assembly of an entire epothilone biosynthetic gene cluster (62.4 kb) from ten individual parts.

Results

The recombination efficiencies of φBT1 integrase-catalyzed reactions between attB and attP sites in vitro were over 90% in our previous studies and could therefore be used for establishing a robust DNA assembly approach. This system contains 16 pairs of mutated recombination sites, the purified φBT1 integrase and a simple buffer (see Materials and Methods). Once the target DNA modules have been flanked by a pair of non-compatible att sites, multiple fragments can be assembled in one reaction simultaneously at 30°C in vitro (Fig. 1). Because site-specific recombination occurs by precise breakage-joining events and do not involve any DNA synthesis or loss, this method is highly accurate, and no errors are introduced after assembly.

We selected the epothilone biosynthetic gene cluster to test this SSRTA method. Epothilone polyketides are promising anti-cancer drugs with remarkable microtubule-stabilizing activity. The total length of the gene cluster is nearly 56 kb, and consists of six open reading frames (epoA to epoF) and ten modules, including one loading module (LM), one non-ribosomal peptide synthetase (NRPS) module and eight polyketide synthase (PKS) modules. Modules 3 to 6 form a whole ORF defined as epoD, while modules 7 and 8 form epoE (see Supplemental Fig. S1). EpoK, a cytochrome P450 involved in epoxidation of epothilones C and D to A and B, was not included in this work. The purified genomic DNA of the epothilone-producing mycobacterial strain Sorangium cellulosum So0157-2 was used as the template for PCR amplification of the modules. Due to the high GC content (69.5%) of the epothilone biosynthetic gene cluster, we adopted an ‘entry clone’ strategy to place individual modules into a series of vectors containing the appropriate pairs of recombination sites beforehand, instead of engineering the recombination sites into the 5’-terminus of PCR primers.

We constructed a series of ‘entry vectors’ with the apramycin resistance gene (aac(3)IV) flanked by a pair of recombination sites. We also put two XcmI recognition sites close to the recombination sites. After the vectors were digested with XcmI, the target DNA modules were inserted into the ‘entry vectors’ through TA cloning (Fig. 2a). The whole epothilone biosynthetic gene cluster was divided into small parts to facilitate amplification by high-fidelity PCR (PrimerSTAR™ HS DNA polymerase, Takara) (Supplemental Fig. S1). Through PCR amplification and sub-cloning, we constructed nine independent ‘entry clones’ which contained all the essential genes for epothilone production. The ribosomal binding site sequences were fused in the 5’-termus of each open reading frame (ORF) by PCR amplification (Supplemental Fig. S1). After linearization by restriction enzymes, these ‘entry clones’ were ready for tandem assembly in vitro. It should be noted that the genes in each entry clone can be replaced by any other desired module.

The construction process was divided into two steps: first we combined apoD DNA sequences from five individual clones and removed the scar sequences inside the ORF. Then we assembled all the epothilone biosynthetic genes from seven individual clones in one reaction. First we tested the ability to combine apoD from four ‘entry clones’ (Fig. 2c). The plasmids pTA0006-M3-apoH, pTA0613-M4, pTA1303-M5, pTA0315-M6 and pZLE10 were linearized (Supplemental Fig. S2) and incubated with φBT1 integrase overnight (or over 8 hrs) at 30°C. As shown in Fig. 3a, multiple assembly products were obtained (the correct, complete assembly product was a 29.8-kb plasmid). The in vitro reaction products were transformed into E. coli strain DH10B by electroporation. After selection with apramycin and neomycin, resistant clones were isolated and the plasmid DNA was extracted. PCR analysis of the linker regions between modules and restriction enzyme digestion indicated that the five segments had assembled correctly (Fig. 3b, c, Supplemental Fig. S3 and S4). We next explored the possibility of combining seven segments in one-step reaction to generate pZLE10-epo (Fig. 2d, Supplemental Fig. S2 and S4). As shown in Fig. 3d-f, we obtained the assembled plasmid (62.4 kb) successfully. Thus, this tandem assembly system can be used to assemble at least seven DNA segments into a circular molecule in vitro.

Discussion

DNA assembly provides a means for extensively changing the genetics of entire biological pathways and cells. This may be useful for creating organisms capable of producing more than overproduce individual proteins – a common goal of cut-and-paste cloning techniques. Several successful approaches have been described that achieve this aim, such as BioBrick standard assembly, the ‘Gibson’ isothermal assembly method and the Golden Gate assembly method. In this manuscript we have demonstrated the utility of the SSRTA method by reconstructing the epothilone biosynthetic gene cluster. Seven independent DNA fragments were assembled into a circular plasmid (62.4 kb) through a one-step in vitro incubation. An ideal assembly method should have no forbidden sites, would be suitable for combinatorial assembly of different parts from standard modules, and importantly, would allow assembly of the parts in a defined order.

The SSRTA method described here is based on site-specific recombination between a series of non-compatible attB and attP sites and so concerns about forbidden sites are not relevant. It is notable that the epothilone biosynthetic gene cluster contains ten direct repeat sequences larger than 100 bp, including a 554-bp direct repeat. Many
antibiotic biosynthetic gene clusters contains such direct repeats. Homology-based assembly methods may not be appropriate in these cases, but SSRTAT has proved successful. Furthermore, each ‘entry vector’ could be used to build libraries of biological parts. These parts could be reconstructed in a defined order through combinatorial assembly (Supplemental Fig. S5). PCR products could also be directly used in this SSRTA method. Details of the pairs of non-compatible recombination sites and the sequences of PCR primers are described in Supplemental Fig. S6. The shortcoming of the SSRTA method is that it introduces scar sequences (attR, 42 bp) between modules after assembly, although scar sequences between genes may not be problematic.

Synthetic biology holds promise in the development of cheaper drugs and ‘green’ biofuels, efficient environmental remediation and targeted therapies for diseases. This SSRTA method provides an alternative way to rebuild genetic pathways and networks based on a phage-encoded large serine recombinase. Since nearly ten of these recombinases have been biochemically characterized, this method could be widely adopted in other systems. This would also allow applications that combine multiple SSRTAs.

Methods

System design and construction of the ‘entry vectors’. A major feature of this SSRTAT system is the use of pairs of non-compatible recombination sites. Non-compatible sites cannot recombine with each other, and the stringent control that this offers is the main strength of the system (Fig. 1). To assemble the full-length epothilone biosynthetic gene cluster (see supplemental Fig. S1) from six independent clones, seven pairs of attB/attP sites are needed. We chose pairs 0, 3, 6, 7, 12, 13 and 15, the sequences of which are given in our previous report. These recombination sites were used to construct seven ‘entry vectors’, named pTA0006, pTA0613, pTA1307, pTA0712, pTA1203, pTA0315 and pTA1303 (Fig. 2a). pTA0006 was generated using primers PT1 (containing attB0 and PTF containing attP0) to amplify the apramycin-resistance gene (aac(3)IV), and the PCR product was inserted into pMD19-T by TA cloning. Similarly, primers PTB6 (containing attB6) and PTP13 (containing attP13) were used to construct pTA0613; primers PTB13 (containing attB13) and PTP7 (containing attP7) were used to construct pTA1307; primers PTB7 (containing attB7) and PTP12 (containing attP12) were used to construct pTA0712; primers PTB12 (containing attB12) and PTP3 (containing attP3) were used to construct pTA1203; primers PTB3 (containing attB3) and PTP15 (containing attP15) were used to construct pTA0315; and primers PTB13 (containing attB13) and PTP3 (containing attP3) were used to construct pTA1303 (Supplemental Table S1 and S2). The apramycin-resistance gene was cloned into the ‘entry vectors’ in the same orientation in each case, as shown in Fig. 2a.

Construction of the ‘entry clones’ and destination vector. The complete epothilone biosynthetic gene cluster was divided into 17 parts for PCR amplification (Supplemental Fig. S1). To clone epoA, a 4289-bp DNA fragment was amplified using primers 0a0 and 0b0, and then inserted into pTA0006 by TA cloning to generate pTA0006-M0; for epoB, a 4256-bp DNA fragment was amplified using primers 0a1 and 0b1, and then inserted into pTA0613 to generate pTA0613-M1; for epoC, a 5525-bp DNA fragment was amplified using primers 0a2 and 0b2, and then inserted into pTA1307 to generate pTA1307-M2; for module 3, a 4766-bp DNA fragment was amplified using primers 0a3 and 0b3, and then inserted into pTA0006 to generate pTA0006-M3. Module 4 was divided into two parts, M4a (primers 0a0/ZE05, 3525 bp) and M4b (primers ZE06/04b, 3220 bp); these were amplified and cloned into pTA0613 separately, and then combined into one plasmid to generate pTA0613-M4 by enzyme digestion and ligation. Modules 5, 6, 7 and 8 were each divided into two parts. M5a (primers 0a0/ZE07, 3607 bp) and M5b (primers ZE08/05b, 2767 bp) were PCR amplified and cloned into pTA1303 separately, and then combined into one plasmid to generate pTA1303-M5. M6a (primers 0a6/ZE09, 2490 bp) and M6b (primers ZE10/06b, 2611 bp) were PCR amplified and cloned into pTA1303 separately, and then combined into one plasmid to generate pTA1303-M6. M7a (primers 0a7/ZE11, 3506 bp) and M7b (primers ZE12/07b, 2683 bp) were PCR amplified and cloned into pTA1203 separately. M8a (primers 0a8/ZE13, 2649 bp) and M8b (primers ZE14/08b, 3188 bp) were PCR amplified and cloned into pTA1203 separately. Plasmids containing genes from modules 7 and 8 were combined together by ligation after enzyme digestion to generate pTA1203-epoD. Module 9 was divided
Described in Supplemental Fig. S2. DNA segments under 10 kb were isolated by 0.8% Tris-HCl, 100 mM KCl, 50 mM NaCl, 2 mM EDTA and 1 mM DTT. Heterologous assembly, the five attB sites were digested with XbaI, and NewB to release a 3402-bp fragment; these two fragments were ligated to generate pZLE12. pZLE12 was assembled gene cluster in Streptomyces was constructed as follows. A PCR fragment (primers ZE03/ZE04, 1956 bp) containing the epoA promoter, the terminator from NdeII to release a 4441-bp fragment containing the aphII gene (conferring resistance to kanamycin and neomycin) was inserted into pTA0006-M3 through Spel digestion and ligation (see Supplementary Table S1 and S2).

The destination vector pZLE10 used for integration and expression of the assembled gene cluster in Streptomyces was digested with XbaI and BglII to release a 1962-bp fragment. Plasmid T-Bxbatt1-XbaI to release a 3713-bp fragment and plasmid pSET152 was digested with NewB and XbaI to release a 3402-bp fragment; these two fragments were ligated to generate pZLE10. pZLE10 was digested with XbaI and NewB to release a 3713-bp fragment and plasmid pSET152 was digested with NewB and XbaI to release a 3402-bp fragment; these two fragments were ligated to generate pZLE10 (see Supplemental Table S2).

**Tandem assembly of epoD and the complete epothilone biosynthetic gene cluster.**

The enzymes used to linearize the ‘entry clones’ for in vitro assembly procedures are described in Supplemental Fig. S2. DNA segments under 10 kb were isolated by 0.8% agarose gel electrophoresis in 1× TAE buffer, and purified using an agarose gel DNA extraction kit (Generay); segments larger than 10 kb were purified by phenol-chloroform extraction and ethanol precipitation. For pZLE10-epoD assembly, the five DNA segments containing modules 3, 4, 5 and 6 and pZLE10 were incubated with φBTI integrase overnight (or over 8 hrs) at 30°C. The buffer contained 10 mM Tris-HCl, 100 mM KCl, 50 mM NaCl, 2 mM EDTA and 1 mM DTT. Heterologous expression, purification and preservation of φBTI integrase were described in our previous work28. The tandemly assembled products were purified by phenol-chloroform extraction and ethanol precipitation, and analysis by pulsed-field gel electrophoresis (PFGE) performed on 1% agarose gels (Biolab) in 0.5×TBE buffer, with the following parameters: switch time, 1–6 seconds; run time, 11 hours; angle, 120°; and voltage gradient, 6 V/cm. Recombination events were detected by PCR analysis. The assembled products were transformed into E. coli strain DH10B by electroporation (25 μF capacitance, 1.8 kV/mm, and 200 Ohm resistance). Transformants were selected with apramycin and neomycin. The plasmid from surviving clones were identified by PCR using primers ES53/54, then isolated and purified. The plasmid was identified by further PCR reactions of the linker regions between modules and by restriction digestion (Fig. 3b, c, and Supplemental Fig. S3). Correct clones were propagated for the next round of assembly. For assembly of the total epothilone biosynthetic gene cluster, the seven DNA segments containing epoA, epoB, epoC, epoD, epoE, epoF and pZLE10 were incubated with φBTI integrase, and the subsequent steps were as described above.

1. Hillson, N. J. in *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology* 295–314 (Springer-Verlag, Berlin, 2011).
2. Ellis, T., Adie, T. & Baldwin, G. S. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol.* 3, 109–118 (2011).
3. Smolke, C. D. Building outside of the box: iGEM and the BioBricks Foundation. *Nat. Biotechnol.* 27, 1099–1102 (2009).
4. Anderson J. C., D. J., Leguia M., Wu G. C., Goler J. A., Arkin A. P., Keasling J. D. Bglbricks: a flexible standard for biological part assembly. *J. Biol. Eng.* 4(1):1 (2010).
5. Elledge, S. J. & Li, M. Z. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* 4, 251–256 (2007).
6. Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* 35, e45 (2007).
7. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345 (2009).
8. Tian, J. D. & Quan, J. Y. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* 6, 242–251 (2011).
9. Quan, J. Y. & Tian, J. D. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* 4 (2009).
10. Freeman, G. J., Zhu, B. G., Cai, G. F. & Hall, E. O. In-Fusion™ assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. BioTechniques 43, 356–359 (2007).
11. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. In-Fusion BioBrick assembly and re-engineering. Nucleic Acids Res. 38, 2624–2636 (2010).
12. Gibson, D. G., Smith, H. O., Hutchinson, C. A., Venter, J. C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. Nat. Methods 7, 901–905 (2010).
13. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. PLoS ONE 6 (2011).
14. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. PLoS ONE 3 (2008).
15. Blake, W. J. et al. Pairwise selection assembly for sequence-independent construction of long-length DNA. Nucleic Acids Res 38, 2594–2602 (2010).
16. Zhang, L., Ou, X. J., Zhao, G. P. & Ding, X. M. Highly efficient in vitro site-specific recombination system based on Streptomyces phage φBT1 integrase. J. Bacteriol. 190, 6392–6397 (2008).
17. Zhang, L. et al. DNA cleavage is independent of synopsis during Streptomyces phage φBT1 integrase-mediated site-specific recombination. J. Mol. Cell Biol. 2, 264–275 (2010).
18. Tang, L. et al. Cloning and heterologous expression of the epothilone gene cluster. Science 287. 640–642 (2000).
19. Ogura, H. et al. Epok, a cytochrome P450 involved in biosynthesis of the anticancer agents epothilones A and B: substrate-mediated rescue of a P450 enzyme. Biochemistry 43, 14712–14721 (2004).
20. Gong, G. L. et al. Mutation and a high-throughput screening method for improving the production of Epothilones of Sorangium. J. Ind. Microbiol. Biotechnol. 34, 615–623 (2007).
21. Shen, B., Du, L. C., Sanchez, C., Chen, M. & Edwards, D. J. The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. Chem. Biol. 7, 623–642 (2000).
22. Leadlay, P. F. et al. Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. Nat. Biotechnol. 25, 447–453 (2007).

Acknowledgments
We thank Prof. Yuezhong Li (Shandong University, China) for providing the Sorangium cellulosum strain So0157-2.
This work was supported by grants from the National High Technology Research and Development Program of China (863 Program) (grant number 2007AA021206), National Basic Research Program of China (973 Program) (grant number 2012CB711102) and the National Natural Science Foundation of China (grant numbers 30600009, 30830002).

Author contributions
L.Z., G.-P.Z. and X.-M.D. designed research; L.Z. performed research; L.Z. and X.-M.D. analyzed data; and L.Z., G.-P. Z. and X.-M.D. wrote the paper.

Additional information
Supplementary information accompanies this paper at http://www.nature.com/scientificreports

License: This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/

How to cite this article: Zhang, L., Zhao, G. & Ding, X. Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination. Sci. Rep. 1, 141; DOI:10.1038/srep00141 (2011).