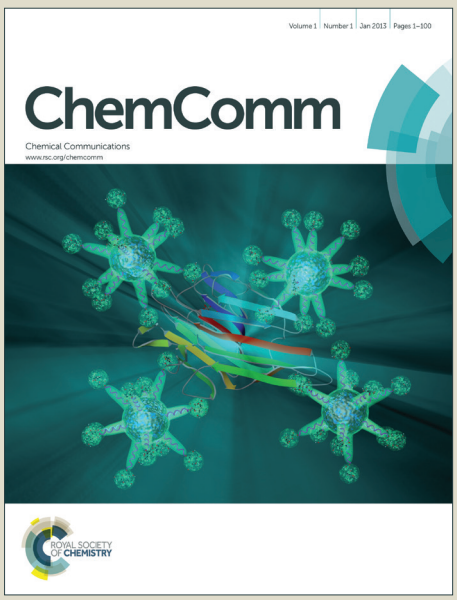
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

DNA-Templated Synthesis of Encoded Small Molecules by DNA Self-Assembly

Cite this: DOI: 10.1039/x0xx00000x

Cheng Cao,[a] Peng Zhao,[a] Ze Li,[a] Zitian Chen,[c] Yanyi Huang,[c] Yu Bai,[a,*] and Xiaoyu Li[a,b,*]

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

We report a novel method for the synthesis of DNA-encoded libraries without the need for discrete DNA template. Reactant DNAs self-assemble to enable chemical reactions and photo-cleavage transfers the product to the DNA terminus, making it suitable for the subsequent affinity-based selection and hit deconvolution.

Originally in 1992, Brenner and Lerner proposed a visionary concept of using DNA to record chemical reactions in combinatorial libraries.[1] During the past two decades, many strategies on DNA-encoded library (DEL) synthesis and selection have been developed.[2-13] Today DELs can be prepared with thousands of millions of compounds,[3, 14] and library selection and hit decoding can be feasibly accomplished[14-19] to discover novel binders against biological targets.[8, 9] Pharmaceutical companies have also employed DELs in drug discovery programs.[14, 20-23]

DNA-templated synthesis (DTS), developed by Liu and co-workers,[8, 24-26] is an important DEL synthesis strategy. Based on sequence complementarity, DTS uses many DNA templates of different sequences to direct chemical reactions with multiple sets of "reagent DNAs" carrying library building blocks. A DEL can be prepared with a large pool of templates through multi-step DTS. However, this "one-template, one-compound" method requires a unique DNA template for each library compound, resulting in challenges and difficulties in sequence design and template preparation, especially for large libraries. Previously, two DEL strategies that do not need a template pool were reported: ESAC by Neri and co-workers, which is specific for fragment-based libraries,[13] and the YoctoReactor by Vipergen, which confines DTS within the tight spaces of multi-way DNA junctions.[20] Recently we reported a "universal template" strategy capable of synthesizing DELs with just a single template regardless the size of the library.[27] Here we report a new library synthesis method that does not require any discrete DNA template.

Our approach is shown in Figure 1b. We use a simple "T" architecture to assemble reactant DNAs to proximity but also with flexibility to support DTS with various building blocks.[20, 28, 29] After chemical reactions, ligase connects DNA strands so that the DTS product is encoded by the original reagent DNAs. Typically, DTS

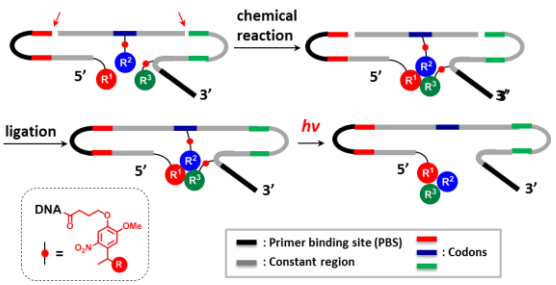


Fig. 1 Scheme for DEL synthesis without DNA template. A typical 3-step synthesis with 3 sets of reagent DNAs is shown. Red dots indicate photo-cleavable linkers (structure is shown on the lower left). Red arrows indicate enzyme ligation sites.

synthesizes compounds at the DNA terminus, presumably more suitable for selections against protein targets (i.e. less interference from the DNA tags). In order to achieve this, we incorporated photo-cleavable *o*-nitrobenzyl linkers in R^2 and R^3 (Figure 1 and S1),[30] while R^1 connects with DNA via a non-cleavable amide linkage; after chemical reactions and enzymatic ligation, without the need for additional cleavage reagents, irradiation readily cleaves both *o*-nitrobenzyl linkers and the final DTS product is obtained at the 5'-terminus of the encoding DNA.

We initiated our study with an *N*-acylthiazolidine synthesis, which can be assembled with an aldehyde, an aminothiol, and a carboxylic acid.[29, 31] We first prepared reagent DNAs (**RD1**, 41-nt; **RD2**, 26-nt; **RD3**, 47-nt; Figure 2a and S3); **RD1**/**RD2** contains an *Afl*II digestion site so that the majority of DNA can be removed afterwards to facilitate MS characterization. These reagent DNAs were mixed and subjected to a typical DMT-MM-mediated acylation condition, along with a series of control experiments to validate the reaction specificity. As shown in Figure 2b, the incubation of **RD2** and **RD3** gave a lower motility band, which presumably is the unacylated thiazolidine (lane 2).[29] With the addition of **RD1** and DMT-MM, a band matching the expected length of the *N*-acylthiazolidine product (114-nt) was observed (lane 3), which can be cleaved after irradiation (lane 4). Importantly, negative control reactions did not generate the same band (lane 5-9), proving the 114-nt product formation requires all three reagent DNAs with

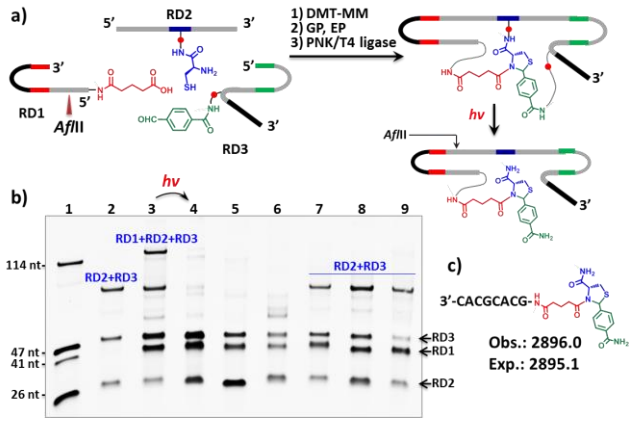


Fig. 2 a) Synthesis scheme of an *N*-acylthiazolidine. GP: gel purification; EP: ethanol precipitation. PNK: T4 polynucleotide kinase. b) Reactions analyzed by denaturing PAGE. Lane 1: DNA standards; lane 2: incubation of **RD2** and **RD3**; lane 3: reaction of **RD1**, **RD2**, **RD3** and DMT-MM; lane 4: lane 3 after irradiation; lane 5-9: same as lane 3 but with a control RD2 (no aminothiol), with mismatched **RD3**, with mismatched **RD1**, without DMT-MM, or with Ac_2O capping before amidation. **RD1-3**: 1 μM each; amidation: 50 mM DMT-MM, 25 °C, 12 hours; irradiation: 365 nm, 0 °C, 15 min. c) MALDI-MS characterization of the final product after *Afl*II digestion. See ESI† for details.

complementary sequences. The 114-nt band in lane 3 was gel-purified and then subjected to PNK/T4 DNA ligase-mediated ligation to afford the *N*-acylthiazolidine product (Figure S4), which was characterized by MALDI-MS after *Afl*II digestion (Figure 2c).

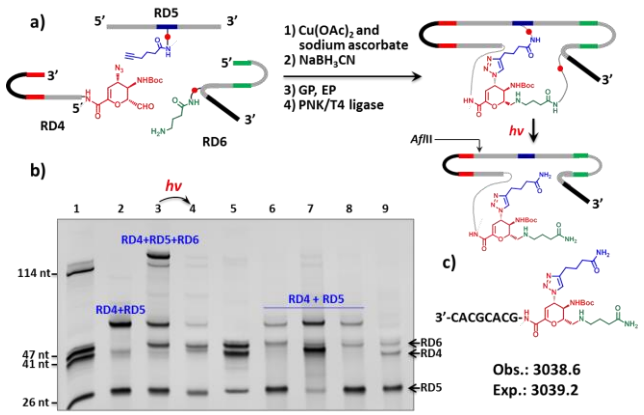


Fig. 3 a) Scheme for the dihydropyran decoration. GP: gel purification; EP: ethanol precipitation. b) Denaturing PAGE analysis. Lane 1: DNA standards; lane 2: reaction of **RD4** and **RD5** with $Cu(OAc)_2$/sodium ascorbate; lane 3: reaction of **RD4**, **RD5**, and **RD6** with $Cu(OAc)_2$/sodium ascorbate and $NaBH_3CN$; lane 4: lane 3 after irradiation; lane 5-9: same as lane 3 but with mismatched **RD4**, with mismatched **RD6**, with a **RD4** without the aldehyde appendage (Figure S6), no $NaBH_3CN$, with a **RD4** without the scaffold (Figure S6). Reagent DNAs: 1 μM each; Click reaction: 500 μM $Cu(OAc)_2$, 500 μM sodium ascorbate, 25 ºC, 3 hours; reductive amination: 50 mM $NaBH_3CN$, 25 ºC, 4 hours; irradiation: 365 nm, 0 °C, 15 min. c) MALDI-MS characterization of the final product after *Afl*II digestion.

Next, we performed a DNA-templated decoration of a dihydropyran scaffold with this method (Figure 3). This scaffold is based on the core structure of a class of approved ant-influenza drugs such as oseltamivir, zanamivir, and laninamivir.[32] We attached four chemically orthogonal appendages on the scaffold reasoning that either of these appendages can be conjugated to DNA while the other three could be diversified with library building blocks. We prepared the scaffold following a report by Smith, Itzsteinz, and co-workers (Figure S2)[33] and then conjugated it to DNA via the carboxylic acid handle to afford **RD4**. **RD5** and **RD6** were prepared

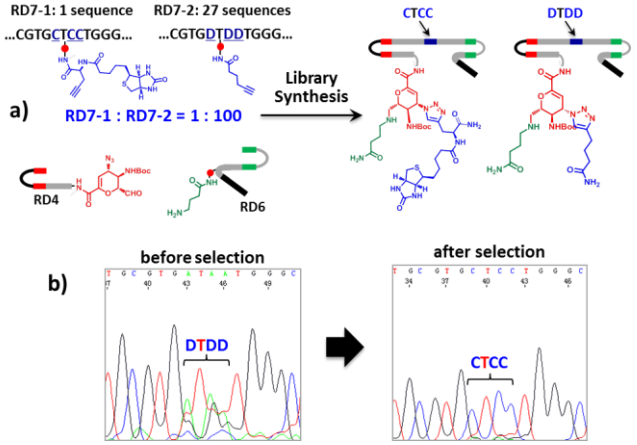


Fig. 4 a) Model library synthesis scheme. **RD7** contains mixed sequences. Reaction conditions are the same as in Figure 3. b) Sequencing results of the model library before and after the selection against immobilized streptavidin. Encoding sites before and after the selection were compared as marked in the figure. See ESI† for details.

bearing a photo-cleavable alkyne and amine, respectively. These reagent DNAs were assembled and subjected to the typical Click reaction and reductive amination conditions. As shown in Figure 3b, **RD4**'s azide reacted with **RD5**'s alkyne, catalyzed by copper (lane 2), and then further generated the product in the presence of **RD6** and $NaBH_3CN$ (114-nt band; lane 3). As expected, this band can be cleaved under irradiation (lane 4). Negative control reactions again indicate that the product formation requires all three reagent DNAs (lane 5-9). After gel purification, enzymatic ligation and *Afl*II digestion (Figure S5), we characterized the product by MALDI (Figure 3c). In addition, since the final product is encoded by the ligated reagent DNAs, which retains the duplex structure and has the "scar" from photo-cleavage, we also validated that the encoding DNA is still compatible with PCR amplification (Figure S4 and S5).

Next, we prepared a model library (Figure 4a). In this library, **RD7-1** has a 3-base codon ("**CTCC**"; T is the small molecule conjugation site) encoding a biotinylated *L*-propargylglycine. **RD7-2** contains mixed sequences ("**DTDD**", D = A, G, or T); for simplicity, they only encode one amino acid (5-hexynoic acid). We mixed **RD7-1** with excess of **RD7-2** (1:100) as the "**RD7**". **RD7** was combined with **RD4**/**RD5** and then subjected to the same reaction conditions as in Figure 3. After synthesis, the library was selected against immobilized streptavidin,[27] which should capture library members containing the biotinylated *L*-propargylglycine. Selected compounds were eluted, PCR-amplified and sequenced (Figure 4b). Results show that the biotin-encoding "**CTCC**" was distinctly enriched after selection. This result is corroborated by the sequencing result of the opposite strand (Figure S8).

Furthermore, we prepared another library containing 18 *N*-acylthiazolidines (Figure 5), in which a biotin-containing building block is encoded by "TAG" at the **RD1** position (**RD1-3**). The ratio of **RD1-3**/(**RD1-1** + **RD1-2**) is 1:400 in the library synthesis. This library was prepared with the same procedure as the Figure 2 and then was also selected against streptavidin. Selection results were decoded by high throughput sequencing (Figure 5b) and again sequences containing the biotin-encoding "TAG" have been significantly enriched (336-fold in average, see ESI† for details). In addition, we have also characterized a portion of the library compounds by mass spectrometry (ESI-MS; Figure S9). Collectively, these results have demonstrated the applicability of our method in the synthesis and selection DNA-encoded libraries with structural and sequence diversities.

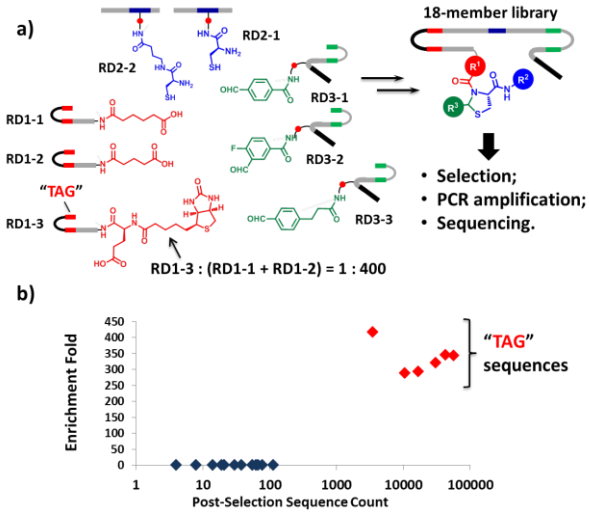


Fig. 5 a) Preparation and selection of an 18-member *N*-acylthiazolidine library, which was prepared with the same procedure as in Figure 2. The library was selected against immobilized streptavidin. Selected compounds were eluted, PCR-amplified and decoded by high throughput sequencing and the result is shown in b). Enrichment fold = (post-selection fraction)/(pre-selection fraction) based on sequence counts before and after the selection. See the ESI† for details.

In conclusion, we have developed a novel DNA-templated synthesis method capable of synthesizing DNA-encoded small molecules without the need for discrete DNA template. This method may further streamline the development of DNA-encoded libraries and accelerate the interrogation of biological targets in drug discovery with DELs.

This work was supported by Ministry of Science and Technology Basic Research Program (2011CB809100), NSFC (21272016, 91013003, 21175008, and J1030413), the Beijing Nova Program (2010B002), and the Doctoral Fund of Ministry of Education of China (20120001110083).

Notes and references

[a] Key Laboratory of Bioorganic Chemistry and Molecular Engineering of the Ministry of Education, Beijing National Laboratory of Molecular Sciences (BNLMS), College of Chemistry and Molecular Engineering, Peking University, Beijing, China 100871. Email: xiaoyuli@pku.edu.cn; yu.bai@pku.edu.cn
[b] Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, China 518055.
[c] Biodynamic Optical Imaging Center (BIOPIC) and College of Engineering, Peking University, Beijing, China 100871
† Electronic Supplementary Information (ESI) available: materials and general methods, experimental details, library selection and sequencing methods and enrichment fold calculations. See DOI: 10.1039/c000000x/

1 S. Brenner and R. A. Lerner, *Proc. Nat. Acad. Sci.*, 1992, **89**, 5381.
2. J. Nielsen, S. Brenner, K.D. Janda, *J. Am. Chem. Soc.*, 1993, **115**, 9812.
3. M. C. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett and M. A. Gallop, *Proc. Nat. Acad. Sci.*, 1993, **90**, 10700.
4. D. R. Halpin and P. B. Harbury, *Plos Biol.*, 2004, **2**, 1015-1021.
5. J. Scheuermann, C. E. Dumelin, S. Melkko and D. Neri, *J. Biotechnol.*, 2006, **126**, 568.
6. S. K. Silverman, *Angew. Chem. Int. Ed.*, 2010, **49**, 7180.
7. M. A. Clark, *Curr. Opin. Chem. Biol.*, 2010, **14**, 396.
8. R. E. Kleiner, C. E. Dumelin and D. R. Liu, *Chem. Soc. Rev.*, 2011, **40**, 5707.
9. L. Mannocci, M. Leimbacher, M. Wichert, J. Scheuermann and D. Neri, *Chem. Commun.*, 2011, **47**, 12747.
10. J. P. Daguer, C. Ciobanu, S. Alvarez, S. Barluenga and N. Winssinger, *Chem. Sci.*, **2011**, 2, 625.
11. S. J. Wrenn, R. M. Weisinger, D. R. Halpin and P. B. Harbury, *J. Am. Chem. Soc.* 2007, **129**, 13137.
12. N. Winssinger, K. Gorska, M. Ciobanu, J. P. Daguer and S. Barluenga, *Methods Mol. Biol.*, 2014, **1050**, 95.
13. S. Melkko, J. Scheuermann, C. E. Dumelin and D. Neri, *Nat. Biotechnol.*, 2004, **22**, 568.
14. M. A. Clark, *et al.*, *Nat. Chem. Biol.*, 2009, **5**, 647.
15. L. Mannocci, Y. X. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko and D. Neri, *Proc. Nat. Acad. Sci.*, 2008, **105**, 17670.
16. F. Buller, L. Mannocci, Y. Zhang, C. E. Dumelin, J. Scheuermann and D. Neri, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 5926.
17. F. Buller, Y. Zhang, J. Scheuermann, J. Schafer, P. Buhlmann and D. Neri, *Chem. Biol.*, 2009, **16**, 1075.
18. R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai and D. R. Liu, *J. Am. Chem. Soc.*, 2010, **132**, 11779-11791.
19. F. Buller, M. Steiner, K. Frey, D. Mircsof, J. Scheuermann, M. Kalisch, P. Buhlmann, C. T. Supuran and D. Neri, *ACS Chem. Biol.*, 2011, **6**, 336.
20. M. H. Hansen, P. Blakskjaer, L. K. Petersen, T. H. Hansen, J. W. Hojfeldt, K. V. Gothelf and N. J. V. Hansen, *J. Am. Chem. Soc.*, 2009, **131**, 1322.
21. J. S. Disch, G. Evindar, C. H. Chiu, C. A. Blum, H. Dai, L. Jin, E. Schuman, K. E. Lind, S. L. Belyanskaya, J. Deng, F. Coppo, L. Aquilani, T. L. Graybill, J. W. Cuozzo, S. Lavu, C. Mao, G. P. Vlasuk and R. B. Perni, *J. Med. Chem*,. 2013, **56**, 3666.
22. R. K. Thalji, *et al.*, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 3584.
23. H. Encinas, *et al.*, *J. Med. Chem.* **2014**, *57*, 1276.
24. Z. J. Gartner and D. R. Liu, *J. Am. Chem. Soc.*, 2001, **123**, 6961.
25. Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder and D. R. Liu, *Science*, 2004, **305**, 1601.
26. B. N. Tse, T. M. Snyder, Y. H. Shen and D. R. Liu, *J. Am. Chem. Soc.*, 2008, **130**, 15611.
27. Y. Li, P. Zhao, M. Zhang, X. Zhao and X. Li, *J. Am. Chem. Soc.*, 2013, **135**, 17727.
28. Z. J. Gartner, R. Grubina, C. T. Calderone and D. R. Liu, *Angew. Chem. Int. Ed.*, 2003, **42**, 1370.
29. X. Y. Li, Z. J. Gartner, B. N. Tse and D. R. Liu, *J. Am. Chem. Soc.*, 2004, **126**, 5090.
30. C. P. Holmes, *J. Org. Chem.*, 1997, **62**, 2370.
31. C. T. Calderone and D. R. Liu, *Angew. Chem. Int. Ed.*, 2005, **44**, 7383.
32. M. von Itzstein, *Nat. Rev. Drug Discov.*, 2007, **6**, 967.
33. P. W. Smith, I. D. Starkey, P. D. Howes, S. L. Sollis, S. P. Keeling, P. C. Cheery, M. von Itzstein, W. Y. Wu and B. Jin, *Eur. J. Med. Chem.*, 1996, **31**, 143.