

Solid-phase DNA-encoded library synthesis: a master builder's instructions

In the format provided by the
authors and unedited

Supplementary Information:

Solid-Phase DNA-Encoded Library

Synthesis: A Master Builder's Instructions

Anjali Dixit¹ and Brian M. Paegel^{1,2*}

¹Department of Pharmaceutical Sciences and

²Departments of Chemistry and Biomedical Engineering, University of California, Irvine, Irvine, 92697, CA, United States.

*Corresponding author(s). E-mail(s): bpaegel@uci.edu;
Contributing authors: dixitaa@uci.edu

Supplementary Information Table 1

xx	5' to 3'			
	x1xx, x3xx, x5xx		x2xx, x4xx, x6xx	
	1xxx[+]	1xxx[-]	2xxx[+]	2xxx[-]
1	TGGAAAGT	AGTTTCCA	CGTGGTAA	TTAGGAGG
2	ACGGAGGA	TGCTCGT	AACGTCAA	TTGAGGTT
3	TTGGAGTT	AAGTCCAA	AATCCGAT	ATGGGATT
4	AAGGAGGT	ACCTCCTT	AACCTAC	GTAGGGTT
5	AGAAAGCA	TGCTTTCT	ATCCTCTC	GAGAGGAT
6	ACAGAACT	AGTTCTGT	CATTCAA	TTGAAATG
7	TAAGGAGT	ACTGCTTA	CGCCTTCA	TGAAGGCG
8	ATGGGAGT	ACTCCCAT	CGTTCCTG	CAGGAACG
9	TGAAGGAA	TTGCTTCA	TTCTTCAT	ATGAAGAA
10	TTGAGGAT	ATGCTCAA	TCCTCTTA	TAAGAGGA
11	AAAGAAAC	GTTCTTTT	AAGCTTGG	CGAAGGTT
12	AAAGGACA	TGCTCTTT	AAGTCCGG	CGGGAGTT
13	AAAGGGAT	ATGCTCTT	AAGTCTTT	AAAGAGTT
14	AAGAAACC	GTTTTCTT	AATCTCA	TGAGGATT
15	AAGAGGCA	TGGCTCTT	AATCTCCC	GGGAGATT
16	ACAAAGAC	GTCTTTGT	AATCTTGT	ACAAAGTT
17	ACAAGAAA	TTTCTTGT	AATTCGA	TGGAAATT
18	ACAAGGCT	AGCCTTGT	ACCTCCT	AGGAGGTT
19	ACAAGGTA	TAGCCTGT	ACCTTGA	TGAAGGTT
20	ACGAAGA	TCTTTGTT	ACCTCCAA	TTGAGAGT
21	ACGAGATT	AATCTGTT	AGCTGTCC	GGAGAGGT
22	ACGAGGGC	GGCCTCGT	ACCTCGCG	CGAAGGTT
23	ACGGAATC	GATTCGGT	ACTCCGCG	GGGGAGTT
24	ACGGGAAG	CTTCCCGT	ACTCCTTT	AAAGGAGT
25	AGAAAGCC	GGTCTTCT	ACTTCCTG	AGGGAGTT
26	AGGAAGGG	CCCTTCCT	AGCCCTCT	AGAGGGCT
27	AGGGAAT	ATTTCCTG	AGCTCCTG	GAGGAGCT
28	ATAAGGGA	TCCTCTAT	AGTCTCTA	TAGAGACT
29	ATAGAGGC	GGGCTCTAT	AGTTCTGT	AGGAAAGT
30	CAAAAGCT	AGTCTTTG	AGTTCTAT	ATGAAGCT
31	CAAAAGGAC	GTGCTTTG	ATGCTCTA	TAAGGGAT
32	CAAGAAGA	TCTTCTTG	ATCTCGGT	ACGGAGAT
33	CAGAGTCT	GAGTCTTG	ATCTCCA	TGGAGAT
34	CAGAAGGA	TCCTCTCG	ATTCTGGG	CCGAGAT
35	CAGAGAAA	TTTCTCTG	ATTTCCAC	GTGGAAT
36	CAGGGACG	CGTCCCTG	CACCTGA	TCAGGGTG
37	CGAAAGCT	AGTTTGGG	CACCTGGG	GGGAGGTT
38	CGAGGAGC	CTCCTGGG	CACCTTAT	ATAAGTTG
39	CGGAGGCG	GGCTGGGG	GAGTGGAT	ATGAGGTT
40	CGAGAACG	GGTTCTCG	CATCCCTA	TAGGGATG
41	CGAGGAGG	CGTCTCTG	CCCTCCGG	CGGAGGGG
42	CGAGGGCA	TGCCCTCG	CCCTCTTA	TAGAAGGG
43	CGGGAATA	TATTCGGG	CCCTTTGG	CGAAAGGG
44	CGGAGGCT	AGTCCCGG	CCCTCAT	ATGAGAGG
45	CTGAAGCC	GGGTTGAG	CGTTCCCG	GGAAAGGG
46	CTGGAAGC	GTTTCCAG	CGCTCCCA	TGGGAGCG
47	GAAGAGGT	AGGCTGTC	CGTCCGAG	GTGGAGGG
48	GAGGAACA	TGTTCTCT	CGTCTCGG	CCGAGAGG
49	GAGGGAAT	ATTCCCTC	CGTCTCGG	CGGAGAGG
50	GCAAGGGG	CCCTTTGG	CTCCCTCG	CGAGGGAG
51	GCAAGAAA	TTCTCTGG	CTTCCGGT	ACGGGAAG
52	GCAAGACC	GGTCTCTG	GACTCCGG	CGGAGGTC
53	GGGGAAGT	ACTTCCGG	GGCCTCGG	CGGAGGGC
54	GGGGGATA	TATCCCGG	GGCCTTCC	GGAAAGGG
55	GGGAAGAA	TCTGTTCC	GGCTCCTT	AAGGAGGG
56	GGGAGAGT	ACCTCTCC	GCTCCCTG	CAGGAGGC
57	GGAGGATT	AATGCTGG	GGGCTTAA	TTAGGGGG
58	GGAGGGAC	GTCCCTCC	GGCTCTCG	CGAGAGCC
59	GGGAAAGC	GGTTTCCC	GGCTTCCC	GGGAAGCC
60	GGGAAGAA	TTCTTCCC	GGTCCCGA	TGGGAGCC
61	GGGAGGTC	GAGCTCCC	GGTTTCGG	CGAAAGCC
62	GTAAGTCT	GACTTTAC	GTTTCCCG	CGGGAAGC
63	GTAGAAGC	GCTTCTAC	TACCTCTT	AAGAGGTA
64	GTAGGAAA	TTTCTCTAC	TACTCTGA	TGAAGTA
65	GTAGGGCT	AGCCCTAC	TCCTCTAC	GTGAGGGA
66	GTGGAGGA	TCTCCAC	TCCTTTGT	ACAAAGGA
67	TAAGAGCA	TGCTCTTA	TCTCCTCC	GGAGGAGA
68	TCAAAGTT	AACTTTGA	TCTCCTC	GAGGAAGA
69	TGGGAGAC	GTCTCCGA	TGTCCCTT	AAGGAGCA
70	TGGGAGCT	AGTCCCGA	TGTCTTCT	AGAAAGCA
71	TTAGGGTC	GAGCCTAA	TGTTCTAA	TTGAAGCA
72	TTGAGAGA	TCTCTCAA	TTCCCTAT	ATAGGGAA

Supplementary Information Table 1. Sequences of Coding Oligonucleotides

Supplementary Information Table 2

	11xx	22xx	13xx	24xx	15xx	26xx
[+]	/5Phos/ATGG	/5Phos/TCA	/5Phos/GTT	/5Phos/CTA	/5Phos/TTC	/5Phos/CGC
[-]	/5Phos/TGA	/5Phos/AAC	/5Phos/TAG	/5Phos/GAA	/5Phos/GCG	/5Phos/AGGC

Supplementary Information Table 2. Sequences of overhangs in oligonucleotide sequences by position

Supplementary Information Table 3

OS[+]	/5Phos/ATACCATGGCATATGCCGCCGCCGCTTCGTCCTTCTCAGCGAC
OS[-]	/5Phos/CCATGTCGCTGAGAAGGACGAAGGCGGCGGCGGCATATGCCATGGTATGG
CS[+]	/5Phos/GCCTCCCAAACNNNNNNNGTTTGCCCGCCAGTTGTTGTGCCAC
CS[-]	GTTTGGG
Sanger FOX primer	CGCCAGGGTTTTCCAGTCACGACCAACCACCCAAAACCACAAACCCAAACCCCAA ACCCAACACACAACAACAGCCGACGCCTTCGTCCTTCTCAGCGA
Entry primer	CGCCAGGGTTTTCCAGTCACGAC

Supplementary Information Table 3. Sequences of non-coding oligonucleotides

Supplementary Information Table 4

Post-Cycle Ligation	200 μ m Tags (MPB)	% RSD	10 μ m Tags (MPB)	% RSD
CuAAC	6E7	71	6E5	28
OS	5E7	26	1E5	8
11xx	2E7	11.7	1E5	32
22xx-13xx	> 7E8	N/A	1E6	9
24xx-15xx	5E8	23	1E6	21

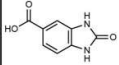
Supplementary Information Table 4. Example DNA tag quantitation during DEL synthesis for a 3-cycle library.

Supplementary Information Table 5

Post-Cycle Ligation	200 μ m Tags (MPB)	% RSD	10 μ m Tags (MPB)	% RSD
CuAAC	6E7	64	9E4	11
OS	3E7	2	1E5	96
11xx	1E8	49	5E5	27
22xx	3E8	13	7E4	128
13xx	6E8	15	2E6	79
24xx-15xx	> 7E8	N/A	5E6	43
26xx	6E8	27	2E5	43

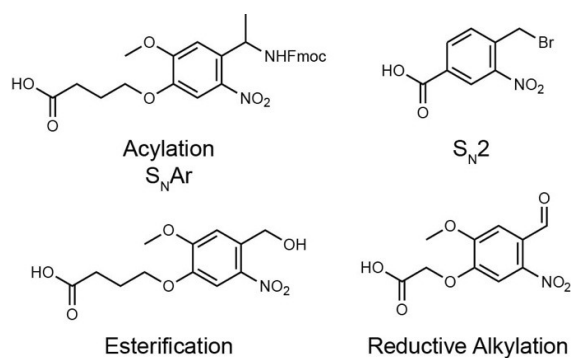
Supplementary Information Table 5. Example DNA tag quantitation during DEL synthesis for a 2-cycle library.

Supplementary Information Table 6

Structure	Solubility	Mass Observed LC-MS	Mass Observed Chemistry	Pass?	Structure	Solubility	Mass Observed LC-MS	Mass Observed Chemistry	Pass?
		No		No		Yes		Yes	
		Yes		Yes		Partial	Yes		Yes
		No	No	No		Yes		Yes	
		No	No	No		Yes		Yes	
		Yes		Yes		No	No	No	No
		Yes		Yes		Yes		Yes	
		Yes		Yes		Yes		Yes	
	No	No	No	No		Yes		Yes	
		Yes		Yes		Yes		Yes	
		Yes		Yes		Yes		Yes	

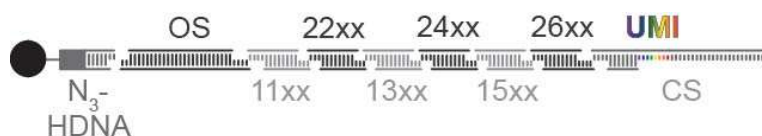
Supplementary Information Table 6. Examples of BB QC analysis.

Supplementary Information Figure 1



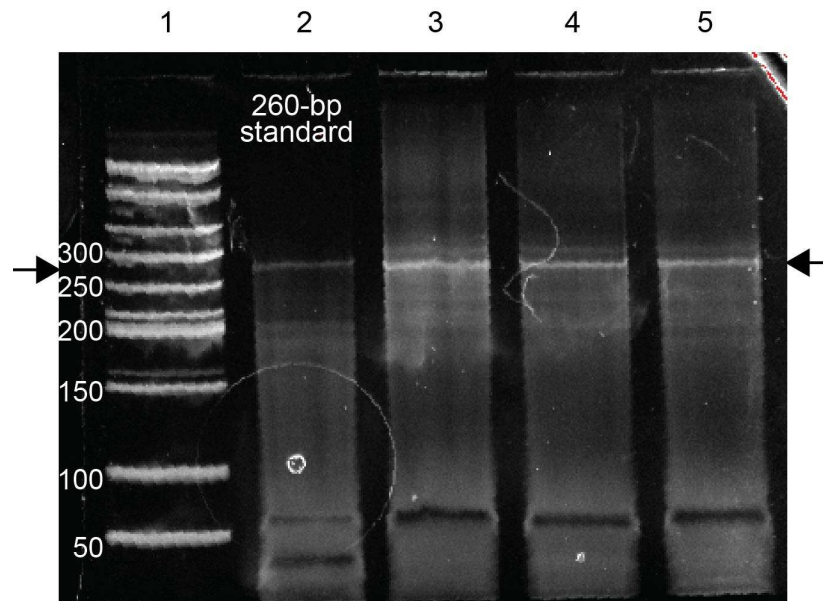
Supplementary Information Figure 1. Examples of photocleavable acids for use in DEL synthesis. The Fmoc-protected amine functional group supports reactions such as acylations and reductive animations (left, top). Bromo-photolinkers are reactive in nucleophilic substitutions (right, top). Hydroxymethyl-linkers react in Steglich esterifications⁴³ (left, bottom). Photolinkers containing an aldehyde are reactive in reductive alkylations (right, bottom).

Supplementary Information Figure 2



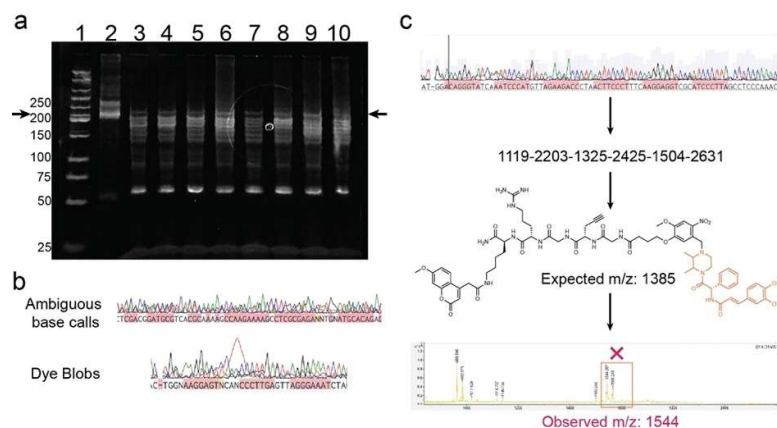
Supplementary Information Figure 2. Schematic of a full solid-phase encoding tag. N₃-HDNA is installed via CuAAC. The **OS[±]** OP, shared between all library members, contains the forward primer binding site for PCR amplification, and a 4-bp overhang for ligation to the first encoding position sequence 11xx. 11-bp coding positions 11xx, 22xx, 13xx, 24xx, 15xx, and 26xx follow; the sequences of these regions will vary depending on the encoding scheme of the library. Each position has a unique 3-bp overhang for specificity in ligation (11xx contains a 4-bp overhang for driving selective ligation to the **OS[±]**), followed by the 8-bp coding region. The final OP **CS[±]** contains a 4-bp overhang for driving selective ligation to 26xx, a brief dsDNA region for enhancing T4 DNA ligase activity, an 8-nt degenerate sequence unique molecular identifier (UMI), and a constant region (ssDNA) containing the reverse primer binding site for PCR amplification.

Supplementary Information Figure 3



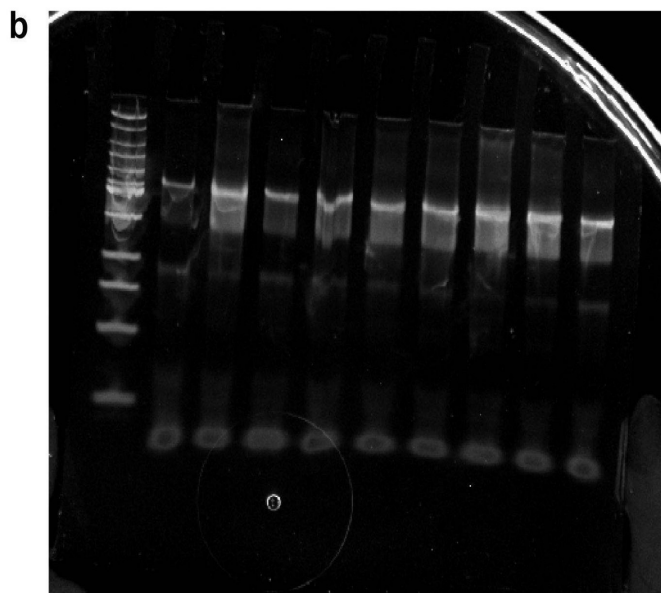
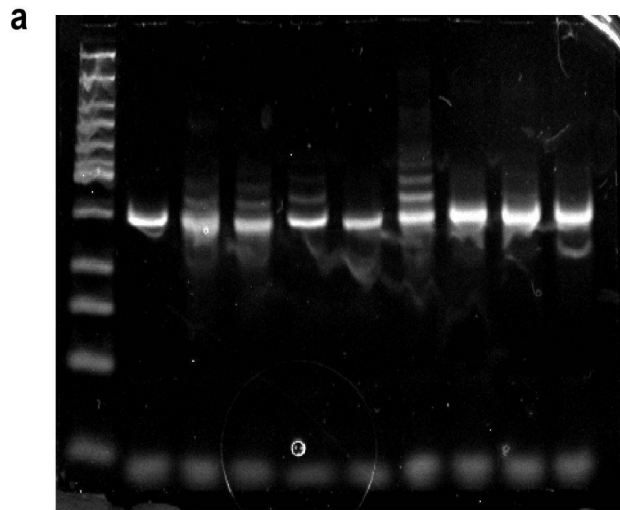
Supplementary Information Figure 3. Example PAGE analysis of 1500-bead 10- μ m screening bead amplicons prepared for next- generation DNA sequencing. PAGE analysis of qPCR amplicons from screening beads indicates the presence of the expected 260-bp product (lanes: 1 = Low MW Ladder, 2 = 260-bp standard amplicon, 3-5 = 1500-bead amplicons).

Supplementary Information Figure 4



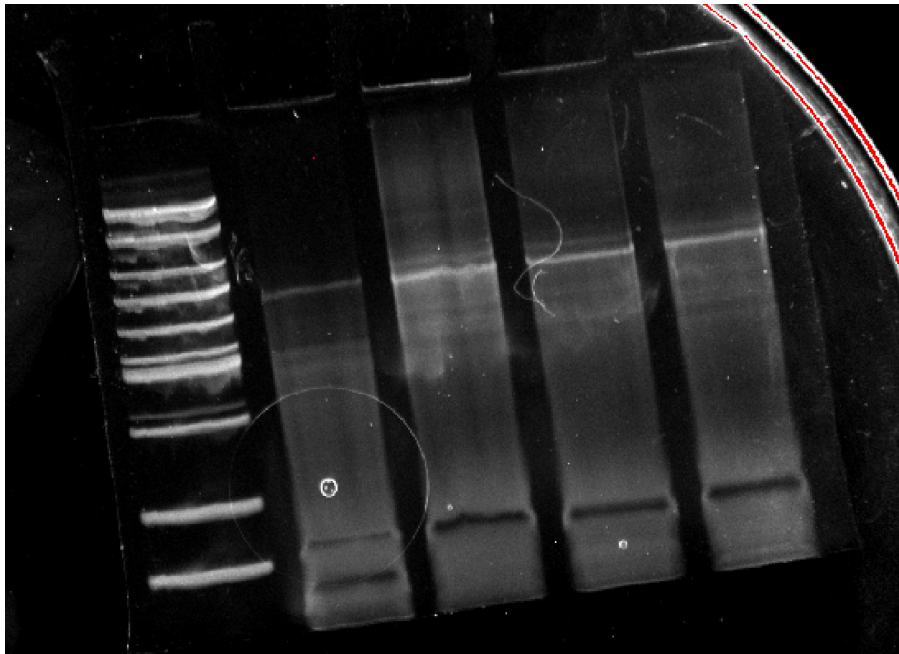
Supplementary Information Figure 4. Example data from QC samples that did not pass inspection. (A) PAGE analysis of qPCR amplicons after installing the FOX primer did not indicate the presence of the expected 221-bp product (lane 1 = Low MW ladder, lane 2 = 221-bp standard, lane 3-10 = single-bead amplicons after Sanger DNA sequencing adapter installation). Bands both smaller and larger than the standard can be observed, indicating truncation and overligation products. (B) Examples of low- quality Sanger sequencing reads include those with ambiguous base calls, which result from sequence contamination or low concentration of template. Dye blobs may result from errors during sequencing sample preparation, and samples should be re-sequenced for proper results. (C) An example of a library member that did not match the expected m/z is shown. This compound contains a building block (cycle 3 carboxylic acid) which often appears in samples that did not pass MALDI inspection.

Supplementary Information Figure 5



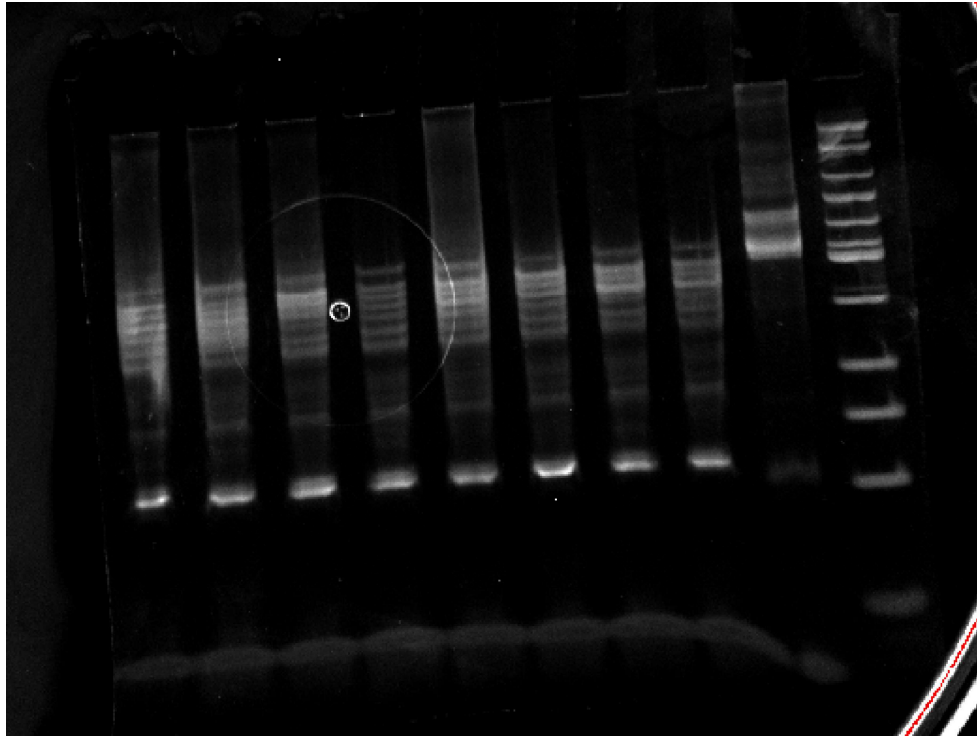
Supplementary Information Figure 5. Raw gel image (unannotated) used in (A) Figure 6A and (B) Figure 6B.

Supplementary Information Figure 6



Supplementary Information Figure 6. Raw gel image (unannotated) used in Supplementary Information Figure [3](#)

Supplementary Information Figure 7



Supplementary Information Figure 7. Raw gel image (unannotated) used in Supplementary Information Figure 4