What is a “DNA-Compatible” Reaction?

Marie L. Malone†,‡ and Brian M. Paegel§,†

†Department of Chemistry and ‡Doctoral Program in Chemical and Biological Sciences, The Scripps Research Institute 130 Scripps Way Jupiter, Florida 33458, United States

Supporting Information

ABSTRACT: DNA-encoded synthesis can generate vastly diverse screening libraries of arbitrarily complex molecules as long as chemical reaction conditions do not compromise DNA’s informational integrity, a fundamental constraint that “DNA-compatible” reaction development does not presently address. We devised DNA-encoded reaction rehearsal, an integrated analysis of reaction yield and impact on DNA, to acquire these key missing data. Magnetic DNA-functionalized sensor beads quantitatively report the % DNA template molecules remaining viable for PCR amplification after exposure to test reaction conditions. Analysis of solid-phase bond forming (e.g., Suzuki–Miyaura cross-coupling, reductive amination) and deprotection reactions (e.g., allyl esters, silyl ethers) guided the definition and optimization of DNA-compatible reaction conditions (>90% yield, >30% viable DNA molecules), most notably in cases that involved known (H+, Pd) and more obscure (Δ, DMF) hazards to DNA integrity. The data provide an empirical yet mechanistically consistent and predictive framework for designing successful DNA-encoded reaction sequences for combinatorial library synthesis.

KEYWORDS: split-and-pool, one-bead-one-compound (OBOC), encoded synthesis

At its inception, combinatorial synthesis represented a revolutionary departure from the throughput limitations of serial compound library synthesis. Split-and-pool diversification1,2 readily yields one-bead-one-compound (OBOC) collections3 that contain in excess of one million members, though library chemotypes that facilitate post-screening mass spectrometric (MS) hit structure elucidation (e.g., α-peptides, peptoids) dominate the landscape. Encoded synthesis circumvents this MS structure elucidation problem by storing structural information in co-synthesized tags4,5 such as nucleic acids6−13 electrophores,14 α-peptides,5,15−18 and peptoids18.

DNA-encoded synthesis in particular has recently garnered intense interest. The widespread availability of advanced high-throughput DNA sequencing technology for decoding19 molecular libraries that can eclipse OBOC library diversity by 2 orders of magnitude, and selection-mode hit identification collectively provide a small molecule discovery platform that is highly complementary to conventional high-throughput screening.11,20−25 Further DNA-encoded reaction development has yielded increasingly “drug-like” libraries20−23,26−28 and adaptation to DNA-encoded solid-phase synthesis (DESPS) promises a broader scope of chemistry for library preparation by liberating synthesis from the constraints of DNA’s limited solubility profile.9,29

The pursuit of expanded DNA-encoded chemical reaction scope has renewed interest in the concept itself: what constitutes “DNA-compatible” chemistry? Unlike other encoding strategies that utilize inert encoding chemotypes (e.g., PNA, hydrocarbons),14,30 most DNA-encoded library synthesis protocols involve chemical synthesis on an unprotected DNA substrate,11,20−28,31−33 thus compatible reactions generally must proceed under conditions that solubilize the DNA substrate and do not modify or destroy the information encoded therein. HPLC-based analysis of reaction product yield on a DNA substrate is now a routine method for evaluating a reaction’s synthesis compatibility with unprotected DNA11 and as a result, recent reaction development using this approach focuses solely on solubility and synthesis yield. The impact of synthesis conditions on DNA’s informational integrity, and consequently its viability in PCR to generate templates for sequencing-based decoding, is currently not a consideration in the development of new DNA-encoded reactions.32−35

The issue of DNA’s informational integrity became a pressing issue during DNA-encoded solid-phase reaction development.9,29 Most solid-phase synthesis conditions, which tend to be highly aggressive, resulted in significant loss of PCR-amplifiable encoding DNA, often below the limit of robust PCR detection (~100 molecules/resin particle).29 However, further investigation revealed trends in reactivity, guiding the development of optimal acylation conditions that generated product in high yield while preserving sufficient PCR-amplifiable encoding DNA for post-synthesis structure elucidation. DNA-encoded syntheses routinely employ up to 3 such transformations. Therefore, a DESPS resin particle harboring ~10 000 DNA-encoding tags could undergo a sequence of 3 chemical transformations, each leaving at least 30% of PCR amplifiable DNA
DNA-encoding tags, and maintain a sufficient quantity of DNA for robust PCR amplification and decoding. These considerations highlighted the need for a general method to evaluate candidate reaction conditions for both high product yield and compatibility with the DNA-encoding tag to guide the design of successful DNA-encoded libraries.

**RESULTS AND DISCUSSION**

Early work in combinatorial synthesis that highlighted the importance of testing each monomer and reaction condition of a library synthesis, dubbed “rehearsal,” inspired us to explore an analogous strategy for encoded synthesis. DNA-encoded reaction rehearsal (Figure 1), a quantitative measurement of reaction product yield and the impact of synthesis conditions on the amplifiability of DNA, uses a mixture of magnetic beads functionalized with a dsDNA encoding tag (“sensor beads”) and synthesis resin functionalized with a short peptide linker. Aliquots of mixed sensor beads and synthesis resin then undergo test reaction conditions, followed by magnetic separation of sensor beads from synthesis resin and analysis. Cleavage and HPLC/MALDI-TOF MS analysis of resin-bound material provides synthesis product yield (henceforth “yield”). Quantitative PCR (qPCR) analysis of sensor beads measures the amplifiable DNA remaining after exposure to the test reaction conditions (normalized to the unexposed sensor bead stock, henceforth “% DNA remaining”), which corresponds to the amplifiable DNA tags remaining for post-synthesis PCR detection and subsequent sequencing. Using this strategy, we investigated and optimized (where needed) a variety of solid-phase reactions.

Protecting group manipulations figure prominently in solid-phase synthesis, however, numerous deprotection strategies involve conditions known to compromise DNA structure. We examined a suite of deprotection reactions (Table 1), optimizing them for DESPS compatibility by identifying conditions that minimally impact DNA amplifiability while generating high product yield. Exposure to Pd(PPh₃)₄ during O-Allyl deprotection results in 0.1% DNA remaining, however, changing the order of addition (phenylsilane addition prior to Pd(PPh₃)₄) significantly preserves DNA (40% DNA remaining) while still affording >95% yield. Pd likely coordinates the N7 position of adenine, causing depurination and concomitant phosphodiester bond cleavage. Deprotection of silyl ethers to reveal alcohols using the oligonucleotide synthesis reagent TBAF is poor (3% DNA remaining), but TEA:3HF treatment gives satisfactory yield of alcohol and 70% DNA remaining. In addition to the previously examined DESPS Fmoc deprotection, Mtt-deprotection in the presence of 1% TFA proceeds with 50% amplifiable DNA remaining. An alternate route to primary amines, the Staudinger reduction of an azide, was optimized to 93% yield and negligible impact on DNA after overnight incubation in buffered solution. Amine product yield is equally high after TCEP treatment in DMF, but TCEP-HCl acidifies the DMF, degrading the DNA (<0.2% DNA remaining).

Bond-forming reactions are a second important class of transformations in solid-phase synthesis, and like deprotection, can compromise DNA’s viability as a template for PCR amplification. We evaluated and optimized common bond forming reaction conditions for DESPS compatibility. Amide bond formation is the quintessential high-yielding solid-phase reaction. Conditions for acylation of either primary or secondary amines at >95% yield are different due to the reduced reactivity of secondary amines. Harsher secondary amine acylation conditions leave only 17% DNA remaining compared to 51% DNA remaining after exposure to milder conditions of primary amine acylation. Formation of urea linkages following nucleophilic addition to an isocyanate proceeds with high yield and 61% DNA remaining. Reductive amination of an aldehyde with a resin-bound amine proved to be an ideal reaction, proceeding with 90% yield and no detectable loss in DNA amplifiability. Suzuki-Miyaura cross-coupling, a carbon–carbon bond forming reaction of a boronic acid and resin-bound aryl iodide, was optimized to 93% yield of biphenyl product and 47% DNA remaining.

DNA-encoded reaction rehearsal is a quantitative and predictive tool for selecting library synthesis reaction conditions. Optimal reaction conditions balance high single-step yield and preserve sufficient DNA remaining for robust post-synthesis detection by PCR amplification. High yield (ideally >90%) ensures that the DNA-encoded synthesis history predicts the major product. Meanwhile, the synthesis...
resin DNA-encoding tag loading capacity and PCR detection limit jointly define the sustainable severity and frequency of DNA-damaging reaction conditions. For example, a 10-μm resin particle with an encoding tag loading capacity of 10,000 molecules can undergo a sequence of up to 2 cycles of secondary amine acylation with a haloacid followed by nucleophilic displacement with an amine, and maintain at least ~100 DNA molecules necessary for detection using quantitative PCR and subsequent sequencing. The same resin could support ~10 iterative reductive amination and deprotection cycles.

While DNA-encoded rehearsal data directly inform reaction selection for solid-phase library synthesis, solution-phase libraries are much more prevalent in the literature and introduce potentially new considerations for defining chemistry compatibility. Whereas DESPS resin particle products exhibit quantifiable encoding tag redundancy, in that each resin particle displays multiple copies of the same encoding tag, solution-phase DNA-encoded synthesis products are single compounds each displaying a single encoding tag. A handful of studies now exist describing the preparation and screening of solution-phase libraries that incorporate diverse reactions—including those rehearsed above—though only with conversion yield data. We sought to determine whether our assay could be extensible to the investigation of solution-phase DNA-encoded reaction development, and whether and how those data could apply to solution-phase library planning and synthesis.

DNA-encoded reaction rehearsal can simulate solution-phase DNA-encoded library synthesis by combining magnetic sensor beads with a S′-amino-modified oligonucleotide to evaluate synthesis product yield on a soluble DNA substrate. Following a test reaction, sensor beads are analyzed as before, and the solution is analyzed by HPLC and MALDI-TOF MS to determine the conversion of S′-amino oligonucleotide starting material to DNA-conjugated product (Table 2). Condensation of S′-amino oligonucleotide substrate with Fmoc-Lys(Alloc)-OH in the presence of the activator DMT-MM, a staple amide bond forming reaction in DNA-encoded library synthesis, proceeds with 65% yield and 69% DNA remaining. Subsequent Alloc deprotection yields a mix of Alloc-deprotected (70%) and fully deprotected product (30%), and 50% DNA remaining. Published Suzuki-Miyaura cross-coupling conditions leave 30% of amplifiable DNA remaining. Conventional solution-phase library preparation includes HPLC purification, likely eliminating synthesis products attached to Pd-catalyzed DNA cleavage products (see Supporting Information). The particularly aggressive pH and temperature of published DNA-compatible quinazolinone formation unsurprisingly proved to be quite detrimental to encoding tag integrity (<0.2% DNA remaining).

Employing reactions that significantly degrade DNA-encoding tag PCR amplifiability during library synthesis may have serious consequences for later sequencing and decoding. Consider, for example, a 1-pmol aliquot of a 25-million member solution-phase library, which would contain an average of 24,000 copies of each library member if all library reaction conditions cause an equal loss of amplifiable DNA. However, if a single reaction condition results in ~1000-fold greater loss of DNA (e.g., quinazolinone formation) than average conditions, only 24 detectable copies of any library member product of that reaction will remain. Whether such inadvertent bias in encoding tag amplifiability will negatively impact solution-phase library selections is unclear since published libraries do not yet incorporate such aggressive chemistry. Rehearsal data suggest, though, that chemistry incompatibility can silence single library members or whole structural families within the library if the loss in DNA due to any single reaction or synthesis product proceeds with 65% yield and 69% DNA remaining. Subsequent Alloc deprotection yields a mix of Alloc-deprotected (70%) and fully deprotected product (30%), and 50% DNA remaining. Published Suzuki-Miyaura cross-coupling conditions leave 30% of amplifiable DNA remaining. Conventional solution-phase library preparation includes HPLC purification, likely eliminating synthesis products attached to Pd-catalyzed DNA cleavage products (see Supporting Information). The particularly aggressive pH and temperature of published DNA-compatible quinazolinone formation unsurprisingly proved to be quite detrimental to encoding tag integrity (<0.2% DNA remaining).

Employing reactions that significantly degrade DNA-encoding tag PCR amplifiability during library synthesis may have serious consequences for later sequencing and decoding. Consider, for example, a 1-pmol aliquot of a 25-million member solution-phase library, which would contain an average of 24,000 copies of each library member if all library reaction conditions cause an equal loss of amplifiable DNA. However, if a single reaction condition results in ~1000-fold greater loss of DNA (e.g., quinazolinone formation) than average conditions, only 24 detectable copies of any library member product of that reaction will remain. Whether such inadvertent bias in encoding tag amplifiability will negatively impact solution-phase library selections is unclear since published libraries do not yet incorporate such aggressive chemistry. Rehearsal data suggest, though, that chemistry incompatibility can silence single library members or whole structural families within the library if the loss in DNA due to any single reaction or synthesis product proceeds with 65% yield and 69% DNA remaining. Subsequent Alloc deprotection yields a mix of Alloc-deprotected (70%) and fully deprotected product (30%), and 50% DNA remaining. Published Suzuki-Miyaura cross-coupling conditions leave 30% of amplifiable DNA remaining. Conventional solution-phase library preparation includes HPLC purification, likely eliminating synthesis products attached to Pd-catalyzed DNA cleavage products (see Supporting Information). The particularly aggressive pH and temperature of published DNA-compatible quinazolinone formation unsurprisingly proved to be quite detrimental to encoding tag integrity (<0.2% DNA remaining).

Employing reactions that significantly degrade DNA-encoding tag PCR amplifiability during library synthesis may have serious consequences for later sequencing and decoding. Consider, for example, a 1-pmol aliquot of a 25-million member solution-phase library, which would contain an average of 24,000 copies of each library member if all library reaction conditions cause an equal loss of amplifiable DNA. However, if a single reaction condition results in ~1000-fold greater loss of DNA (e.g., quinazolinone formation) than average conditions, only 24 detectable copies of any library member product of that reaction will remain. Whether such inadvertent bias in encoding tag amplifiability will negatively impact solution-phase library selections is unclear since published libraries do not yet incorporate such aggressive chemistry. Rehearsal data suggest, though, that chemistry incompatibility can silence single library members or whole structural families within the library if the loss in DNA due to any single reaction or synthesis product proceeds with 65% yield and 69% DNA remaining. Subsequent Alloc deprotection yields a mix of Alloc-deprotected (70%) and fully deprotected product (30%), and 50% DNA remaining. Published Suzuki-Miyaura cross-coupling conditions leave 30% of amplifiable DNA remaining. Conventional solution-phase library preparation includes HPLC purification, likely eliminating synthesis products attached to Pd-catalyzed DNA cleavage products (see Supporting Information). The particularly aggressive pH and temperature of published DNA-compatible quinazolinone formation unsurprisingly proved to be quite detrimental to encoding tag integrity (<0.2% DNA remaining).

Employing reactions that significantly degrade DNA-encoding tag PCR amplifiability during library synthesis may have serious consequences for later sequencing and decoding. Consider, for example, a 1-pmol aliquot of a 25-million member solution-phase library, which would contain an average of 24,000 copies of each library member if all library reaction conditions cause an equal loss of amplifiable DNA. However, if a single reaction condition results in ~1000-fold greater loss of DNA (e.g., quinazolinone formation) than average conditions, only 24 detectable copies of any library member product of that reaction will remain. Whether such inadvertent bias in encoding tag amplifiability will negatively impact solution-phase library selections is unclear since published libraries do not yet incorporate such aggressive chemistry. Rehearsal data suggest, though, that chemistry incompatibility can silence single library members or whole structural families within the library if the loss in DNA due to any single reaction or synthesis product proceeds with 65% yield and 69% DNA remaining. Subsequent Alloc deprotection yields a mix of Alloc-deprotected (70%) and fully deprotected product (30%), and 50% DNA remaining. Published Suzuki-Miyaura cross-coupling conditions leave 30% of amplifiable DNA remaining. Conventional solution-phase library preparation includes HPLC purification, likely eliminating synthesis products attached to Pd-catalyzed DNA cleavage products (see Supporting Information). The particularly aggressive pH and temperature of published DNA-compatible quinazolinone formation unsurprisingly proved to be quite detrimental to encoding tag integrity (<0.2% DNA remaining).

Employing reactions that significantly degrade DNA-encoding tag PCR amplifiability during library synthesis may have serious consequences for later sequencing and decoding. Consider, for example, a 1-pmol aliquot of a 25-million member solution-phase library, which would contain an average of 24,000 copies of each library member if all library reaction conditions cause an equal loss of amplifiable DNA. However, if a single reaction condition results in ~1000-fold greater loss of DNA (e.g., quinazolinone formation) than average conditions, only 24 detectable copies of any library member product of that reaction will remain. Whether such inadvertent bias in encoding tag amplifiability will negatively impact solution-phase library selections is unclear since published libraries do not yet incorporate such aggressive chemistry. Rehearsal data suggest, though, that chemistry incompatibility can silence single library members or whole structural families within the library if the loss in DNA due to any single reaction or synthesis product proceeds with 65% yield and 69% DNA remaining. Subsequent Alloc deprotection yields a mix of Alloc-deprotected (70%) and fully deprotected product (30%), and 50% DNA remaining. Published Suzuki-Miyaura cross-coupling conditions leave 30% of amplifiable DNA remaining. Conventional solution-phase library preparation includes HPLC purification, likely eliminating synthesis products attached to Pd-catalyzed DNA cleavage products (see Supporting Information). The particularly aggressive pH and temperature of published DNA-compatible quinazolinone formation unsurprisingly proved to be quite detrimental to encoding tag integrity (<0.2% DNA remaining).

Employing reactions that significantly degrade DNA-encoding tag PCR amplifiability during library synthesis may have serious consequences for later sequencing and decoding. Consider, for example, a 1-pmol aliquot of a 25-million member solution-phase library, which would contain an average of 24,000 copies of each library member if all library reaction conditions cause an equal loss of amplifiable DNA. However, if a single reaction condition results in ~1000-fold greater loss of DNA (e.g., quinazolinone formation) than average conditions, only 24 detectable copies of any library member product of that reaction will remain. Whether such inadvertent bias in encoding tag amplifiability will negatively impact solution-phase library selections is unclear since published libraries do not yet incorporate such aggressive chemistry. Rehearsal data suggest, though, that chemistry incompatibility can silence single library members or whole structural families within the library if the loss in DNA due to any single reaction or synthesis product proceeds with 65% yield and 69% DNA remaining. Subsequent Alloc deprotection yields a mix of Alloc-deprotected (70%) and fully deprotected product (30%), and 50% DNA remaining. Published Suzuki-Miyaura cross-coupling conditions leave 30% of amplifiable DNA remaining. Conventional solution-phase library preparation includes HPLC purification, likely eliminating synthesis products attached to Pd-catalyzed DNA cleavage products (see Supporting Information). The particularly aggressive pH and temperature of published DNA-compatible quinazolinone formation unsurprisingly proved to be quite detrimental to encoding tag integrity (<0.2% DNA remaining).
sequence of reactions decreases the number of amplifiable DNA copies below the PCR limit of detection. These observations with respect to encoding fidelity draw parallels to recent theoretical studies, which predict that variable synthetic yield during DNA-encoded library synthesis may bias next-generation sequencing analysis of hits.

DNA-encoded reaction rehearsal provides a practical window into reaction compatibility with DNA, but is by no means comprehensive. First, % DNA remaining measured with sensor beads may not correlate exactly with the actual % DNA remaining after solution-phase library synthesis. The kinetic barrier associated with reactants diffusing to the sensor bead surface is not present in solution-phase encoded synthesis, thus the sensor beads provide a conservative basis for establishment of solution-phase chemistry compatibility. Second, the sensor beads are quantitative but phenotypic sensors in that they only store DNA, yet minimally impact amplification. The result expanded latitude in DNA-encoded synthesis will yield libraries encompassing more diverse connectivities and molecular content, translating broadly to the discovery of more "drug-like" leads or larger oligomers exhibiting self-assembly or catalytic properties.

### ASSOCIATED CONTENT

#### Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscombsci.5b00198.

Detailed experimental procedure, linker structures, compound characterization data, and qPCR sample data (PDF)

### AUTHOR INFORMATION

#### Corresponding Author
E-mail: brianna@scripps.edu.

#### Funding
A Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to M.L.M. and awards from the NIH Director’s New Innovator (OD008535) and DARPA Fold F(X) programs to B.M.P. (N66001-14-2-4057) supported this research.

#### Notes
The authors declare no competing financial interest.

### ACKNOWLEDGMENTS
We thank Mr. Vuong Dang for experimental assistance.

### ABBREVIATIONS

OBOC, one-bead-one-compound; DESPS, DNA-encoded solid-phase synthesis; HDNA, headpiece DNA; TCEP, tris(2-carboxyethyl)phosphine; DMT-MM, 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride; qPCR, quantitative polymerase chain reaction; DMF, N,N-dimethylformamide; NMP, N-methyl-2-pyrrolidone

---

Table 2. Solution-Phase DNA-Encoded Reaction Rehearsal

<table>
<thead>
<tr>
<th>Solution-phase reaction</th>
<th>% Yield</th>
<th>% Amplifiable DNA remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fmoc-H → NH₂</td>
<td>95</td>
<td>70</td>
</tr>
<tr>
<td>Alloc-H → NH₂</td>
<td>70</td>
<td>50</td>
</tr>
<tr>
<td>Alloc-N₂ → NH₂</td>
<td>N.D.</td>
<td>90/80</td>
</tr>
<tr>
<td>Alloc-N₂ → Alloc-N₂</td>
<td>65</td>
<td>69</td>
</tr>
<tr>
<td>Alloc-N₂ → Alloc-N₂</td>
<td>N.D.</td>
<td>0.2</td>
</tr>
<tr>
<td>Alloc-N₂ → Alloc-N₂</td>
<td>&gt; 95</td>
<td>30</td>
</tr>
</tbody>
</table>

*See Supporting Information for full structures. Reaction conditions: (a) piperidine (10%, H₂O, 4 h, RT); (b) Pd(PPh₃)₄/NaBH₄ (1.4/29 mM, 1:1:7 ACN/DMA/250 mM borate pH 9.4, 16 h, 70 °C); (c) NaOAc/MgCl₂ (30/200 mM, H₂O, 16 h, 18 °C); (d) 250 mM borate, pH 9.4 (16 h, 90 °C); (e) Fmoc-Lys(Alloc)-OH/DMT-MM (27/27 mM, 18/0.89/36 mM, 2:4:9:226 DCM/Toluene/ACN/H₂O, 1.5 h, 80 °C). The asterisk (*) indicates 70% yield of expected product, 30% yield of expected product -Fmoc.


