

**Supporting Information for**

**DNA-Encoded Solid-Phase Synthesis:**

**Encoding Language Design and Complex Oligomer Library Synthesis**

Andrew B. MacConnell,<sup>†</sup> Patrick J. McEnaney, Valerie J. Cavett, and Brian M. Paegel<sup>\*</sup>

Department of Chemistry and <sup>†</sup>Doctoral Program in Chemical and Biological Sciences  
The Scripps Research Institute  
130 Scripps Way  
Jupiter, FL 33458

<sup>\*</sup>Correspondence: [briandna@scripps.edu](mailto:briandna@scripps.edu)

## Supporting Experimental Procedures

**Coding sequence thermodynamic and distance optimization.** All operations were performed using the R programming language version 3.1.0<sup>1</sup> in RStudio with packages ‘Biostrings’ version 2.32.1,<sup>2</sup> ‘seqinr’ version 1.0–2,<sup>3</sup> ‘RCurl’ version 1.95–4.3,<sup>4</sup> ‘XML’ version 3.98–1.1,<sup>5</sup> and ‘httr’ version 0.5.<sup>4</sup> All possible sequences matching the degeneracy patterns 5′-NNRRRRNN-3′ and 5′-NNYYYYNN-3′ were generated, yielding the R-SET[+] and Y-SET[+] (each 4,096 members) of “parent” sequences. The reverse complements of these sets, R-SET[-] and Y-SET[-], respectively, were generated. Overhang  $\approx X1XX[+]$  was appended to the 5′ terminus of all R-SET[+] parent sequence members to yield OH1-R-SET[+], overhang  $\approx X3XX[+]$  was appended to the 5′ terminus of all R-SET[+] parent sequence members to yield OH3-R-SET[+], and overhang  $\approx X5XX[+]$  was appended to the 5′ terminus of all R-SET[+] parent sequence members to yield OH5-R-SET[+] (4,096 members each). Similarly, overhangs  $\approx X2XX[+]$ ,  $\approx X4XX[+]$ , and  $\approx X6XX[+]$  were each appended to the 5′ terminus of all Y-SET[+] parent sequence members to yield OH2-Y-SET[+], OH4-Y-SET[+], and OH6-Y-SET[+] (4,096 members each). Overhangs  $\approx X1XX[-]$ ,  $\approx X3XX[-]$ , and  $\approx X5XX[-]$  were each appended to the 5′ terminus of all R-SET[-] members to yield OH1-R-SET[-], OH3-R-SET[-], and OH5-R-SET[-] (4,096 members each OHX-R-SET[-]), and overhangs  $\approx X2XX[-]$ ,  $\approx X4XX[-]$ , and  $\approx X6XX[-]$  were each appended to the 5′ terminus of all Y-SET[-] members to yield OH2-Y-SET[-], OH4-Y-SET[-], and OH6-Y-SET[-] (4,096 members in each of the six [-] sets). Complementary pair sets, OH1-R-SET[±], OH3-R-SET[±], OH5-R-SET[±], OH2-Y-SET[±], OH4-Y-SET[±], and OH6-Y-SET[±], were generated by pairing each member of one overhang set [+] strand (e.g. OH1-R-

SET[+]) with its complement coding sequence and corresponding overhang on the [-] strand (e.g. OH1-R-SET[-]).

Thermodynamic parameters of the oligonucleotides and oligonucleotide pairs were calculated for all members of the various sets via HTTP POST requests. Requests were sent to an mfold<sup>6</sup> server (via an Integrated DNA Technologies, Inc. Oligo AnalyzerService interface) using encoding ligation conditions of Na<sup>+</sup> (50 mM), Mg<sup>2+</sup> (10 mM), dNTP (1 mM), and oligonucleotide (10 μM). The melting temperature of the target heteroduplex,  $T_{M,h\text{et}}$ , and free energy of target heteroduplex formation,  $\Delta G_{h\text{et}}$ , was calculated for each element of R-SET[+] and Y-SET[+]. The melting temperature of the most stable hairpin,  $T_{M,h\text{p}}$ , and the free energy of forming the most stable homoduplex,  $\Delta G_{h\text{omo}}$ , were calculated for each element of OH1-R-SET[+], OH3-R-SET[+], OH5-R-SET[+], OH2-Y-SET[+], OH4-Y-SET[+], OH6-Y-SET[+], OH1-R-SET[-], OH3-R-SET[-], OH5-R-SET[-], OH2-Y-SET[-], OH4-Y-SET[-], and OH6-Y-SET[-]. The free energy of forming the most stable off-target heteroduplex,  $\Delta G_{h\text{et},2^\circ}$ , was calculated for each of the complementary pair sets OH1-R-SET[±], OH3-R-SET[±], OH5-R-SET[±], OH2-Y-SET[±], OH4-Y-SET[±], and OH6-Y-SET[±]. Thus, for each parent sequence member of R-SET[+] and Y-SET[+], three  $\Delta\Delta G_{h\text{et}/h\text{omo}}$  values were calculated:

$$(1) \quad \Delta\Delta G_{h\text{et}/h\text{omo}} = \Delta G_{h\text{et}} - \Delta G_{h\text{omo}}$$

where  $\Delta G_{h\text{omo}}$  derives from the member of OH1-R-SET[+], OH3-R-SET[+], or OH5-R-SET[+]

for the corresponding parent sequence  $\Delta G_{het}$  in R-SET[+] and similarly  $\Delta G_{homo}$  derives from OH2-Y-SET[+], OH4-Y-SET[+], or OH6-Y-SET[+] for its corresponding parent sequence  $\Delta G_{het}$  in Y-SET[+]. For each member of R-SET[+] and Y-SET[+], three  $\Delta\Delta G_{het/het,2^\circ}$  values were calculated:

$$(2) \quad \Delta\Delta G_{het/het,2^\circ} = \Delta G_{het} - \Delta G_{het,2^\circ}$$

where  $\Delta G_{het,2^\circ}$  derives from the member of OH1-R-SET[+], OH3-R-SET[+], or OH5-R-SET[+] for the corresponding parent sequence  $\Delta G_{het}$  in R-SET[+] and similarly  $\Delta G_{het,2^\circ}$  derives from the member of OH2-Y-SET[+], OH4-Y-SET[+], or OH6-Y-SET[+] for the corresponding parent sequence  $\Delta G_{het}$  in Y-SET[+]. The  $\Delta\Delta G_{het/homo}$  was analogously calculated for R-SET[-] and Y-SET[-] using  $\Delta G_{homo}$  derived from OH1-R-SET[-], OH3-R-SET[-], OH5-R-SET[-], and OH2-Y-SET[-], OH4-Y-SET[-], OH6-Y-SET[-], respectively. For each member of R-SET[+] and Y-SET[+], three  $\Delta T_M$  values were calculated:

$$(3) \quad \Delta T_M = T_{M,hets} - T_{M,hps}$$

where  $T_{M,hps}$  derives from the member of OH1-R-SET[+], OH3-R-SET[+], or OH5-R-SET[+] for the corresponding parent sequence  $T_{M,hets}$  in R-SET[+], and similarly OH2-Y-SET[+], OH4-Y-SET[+], or OH6-Y-SET[+] for the corresponding parent sequence  $T_{M,hets}$  in Y-SET[+]. The  $\Delta T_M$  was analogously calculated for R-SET[-] and Y-SET[-] using  $T_{M,hps}$  derived from OH1-R-

SET[-], OH3-R-SET[-], OH5-R-SET[-], and OH2-Y-SET[-], OH4-Y-SET[-], OH6-Y-SET[-], respectively.

R-SET[+] and Y-SET[+] were pruned according to the calculated parameters. All sequences  $T_{M,het} < 30$  °C were discarded. Members of OH1-R-SET[+], OH3-R-SET[+], OH5-R-SET[+], OH2-Y-SET[+], OH4-Y-SET[+], or OH6-Y-SET[+] overhang sets exhibiting  $\Delta\Delta G_{het/homo}$  or  $\Delta\Delta G_{het/het,2^\circ} > -5.5$  kcal/mol or  $\Delta T_M < 15$  °C were stripped of their 5'-overhang sequence to reveal the identity of the parent sequence member, which was discarded from either R-SET[+] or Y-SET[+]. Members of OH1-R-SET[-], OH3-R-SET[-], OH5-R-SET[-], OH2-Y-SET[-], OH4-Y-SET[-], or OH6-Y-SET[-] overhang sets exhibiting  $\Delta\Delta G_{het/homo} > -5.5$  kcal/mol or  $\Delta T_M < 15$  °C were stripped of their 5'-overhang sequence and reverse complemented to reveal the identity of the parent sequence member, which was discarded from either R-SET[+] or Y-SET[+].

For both the pruned R-SET[+] and Y-SET[+], the Hamming string distance<sup>7</sup> between the first element in the set and all other members was calculated. Set members of distance  $< 3$  were discarded. The first element was removed and stored in R-HAM-SET[+] and Y-HAM-SET[+]. This process was executed recursively until no elements remained in either R-SET[+] or Y-SET[+].

**DNA oligonucleotide pair ligation yield determination.** Streptavidin-coated magnetic resin (1 mg, 200 pmol binding sites) was washed (BWBT, 4 x 200  $\mu$ L), combined with one of six

different 5'-phosphorylated-3'-biotinylated [-] strand oligonucleotide templates (270 pmol, 100  $\mu$ L BWB). The six different template [-] strand sequences were 5'-  
**CCAT**GTAGCGAAGCGAGCAGGACTGGGCGGAAAA-3' (overhang 1 template, OH1-TEM[-]),  
5'-**TGAG**TAGCGAAGCGAGCAGGACTGGGCGGAAAA-3' (OH2-TEM[-]), 5'-  
**AACG**TAGCGAAGCGAGCAGGACTGGGCGGAAAA-3' (OH3-TEM[-]). 5'-  
**TAGG**TAGCGAAGCGAGCAGGACTGGGCGGAAAA-3' (OH4-TEM[-]), 5'-  
**GAAG**TAGCGAAGCGAGCAGGACTGGGCGGAAAA-3' (OH5-TEM[-]), and 5'-  
**GCGG**TAGCGAAGCGAGCAGGACTGGGCGGAAAA-3' (OH6-TEM[-]). Overhangs are bold and underlined. Streptavidin resin and template [-] strand were incubated (15 min, RT), washed (BWB, 4 x 200  $\mu$ L; SSC 3 x 200  $\mu$ L), and resuspended (SSC, 100  $\mu$ L). 5'-fluorescein-labeled [+] strand 5'-CCGCCAGTCCTGCTCGCTTCGCTAC-3' probe (270 pmol, 4.5  $\mu$ L) was added to the template-immobilized bead solution and hybridized (5 min, RT). The hybridized resin was then washed (SSC, 3 x 200  $\mu$ L; BTPWB, 3 x 200  $\mu$ L) and resuspended (BTPWB, 200  $\mu$ L). For example, the OH1-TEM[-]/probe-hybridized resin tests ligation to an OP1 coding module (e.g.  $\approx$ **1101**[ $\pm$ ]). To determine the ligation yield of  $\approx$ **1101**[ $\pm$ ] (200 pmol, prepared as described above) and T4 DNA ligase (300 U) were combined (BTPLB, 20  $\mu$ L) and aliquoted to OH1-TEM[-]/probe-hybridized resin (0.1 mg, 20 pmol sites) and incubated with rotation (1 h, RT, 8 rpm). Ligation assays were performed in triplicate. The resin was then washed (BTPBB, 1 x 200  $\mu$ L; BTPWB, 3 x 200  $\mu$ L; DI H<sub>2</sub>O, 100  $\mu$ L), resuspended (GLB, 15  $\mu$ L), denatured (5 min, 90 °C) and electrophoretically resolved (12% polyacrylamide gel). Ligation yield was calculated by dividing the product band intensity by the sum of the product and starting material band

intensities. Replicate yields were averaged and the standard deviation calculated.

**Example side product synthesis and characterization.** Bifunctional-linker resin (1 mg) was aliquoted into fritted-spin column and washed (DMF, 2 x 150  $\mu$ L). Fmoc deprotection solution was added (20% piperidine in DMF, 2 x 150  $\mu$ L), and resin incubated with rotation (5 min first aliquot, 15 min second aliquot, RT, 8 rpm) then washed (DMF, 3 x 400  $\mu$ L; DCM, 3 x 400  $\mu$ L; DMF, 3 x 400  $\mu$ L). Sample was acylated with (2R, 3E)-5-chloro-2,4-dimethyl-3-pentenoic acid using same conditions used for DESPS compound samples to produce an intermediate terminal allylic chloride product (top panel, left frame). The resin was washed (DMF, 3 x 400  $\mu$ L; H<sub>2</sub>O, 2 x 400  $\mu$ L) then terminal allylic chloride was hydrolyzed (top panel, middle frame) in a solution of aqueous base (1 mM NaOH with 0.02% Tween-20, 500  $\mu$ L, 37 °C, 16 hr). Resin was washed (H<sub>2</sub>O, 3 x 400  $\mu$ L; DMF, 3 x 400  $\mu$ L; DCM, 2 x 400  $\mu$ L) then incubated while rotating (DCM, 400  $\mu$ L, 30 min, RT, 8 rpm) and dried *in vacuo*. Cleavage cocktail (90% TFA, 5% DCM, 5% TIPS, 300  $\mu$ L) was added to the dried resin, incubated (1 h, RT), and eluted sample was evaporated to dryness *in vacuo*. Residue was resuspended (10% ACN, 0.1% formic acid in H<sub>2</sub>O, 400  $\mu$ L) and an aliquot injected (20  $\mu$ L) for LC-MS analysis (Zorbax SB-C18, 4.6 x 100 mm, 80 Å, 3.5  $\mu$ m, Agilent) with gradient elution (mobile phase A: 0.1% formic acid in 5% ACN, 95% H<sub>2</sub>O; mobile phase B: 0.1% formic acid in 95% ACN, 5% H<sub>2</sub>O; 0% – 55% B, 25 min), and absorbance detection ( $\lambda$  = 330 nm).

## Supporting Information References

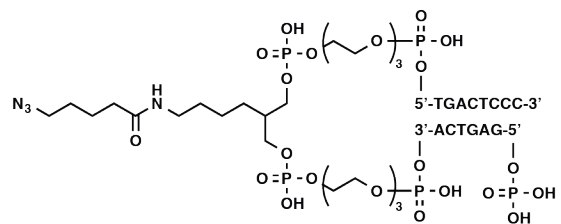
- (1) Venables, W. N.; Smith, D. M. R: A Language and Environment for Statistical Computing, 2014.
- (2) Pages, H.; Aboyoun, P.; Gentleman, R.; DebRoy, S. Biostrings: String Objects Representing Biological Sequences, and Matching Algorithms, 2014, 1–164.
- (3) Charif, D.; Lobry, J. R. Package “Seqinr,” 2007.
- (4) Lang, D. T. RCurl: General Network (HTTP/FTP/...) Client Interface for R, 2014.
- (5) Lang, D. T. XML: Tools for Parsing and Generating XML Within R and S-Plus, 2014, 1–169.
- (6) Zuker, M. Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res* 2003, *31*, 3406–3415.
- (7) Hamming, R. W. Error Detecting and Error Correcting Codes. *Bell System Technical Journal* 1950, *29*, 147–160.



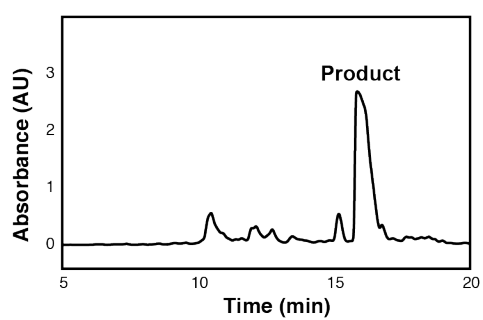
## Supporting Information Figures & Legends

**Figure S1. Structure and characterization of N<sub>3</sub>-DNA.** The target oligonucleotide features two partially complementary DNA strands joined by PEG<sub>3</sub> linkers and elaborated with 5-azidopentanoic acid. (A) Reaction products of coupling of 5-azidopentanoic acid to NH<sub>2</sub>-HDNA were separated using reverse-phase HPLC, and a predominant peak ( $\lambda = 260$  nm, product) was observed. (B) Negative polarity MALDI-TOF MS analysis of the product fraction yielded [M-H]<sup>1-</sup> signal at  $m/z$  of 5062.93,  $\Delta$ ppm of 591 from the theoretical exact mass of 5059.94 Da. The doubly charged [M-2H]<sup>2-</sup> species was also observed at  $m/z$  2532.88.

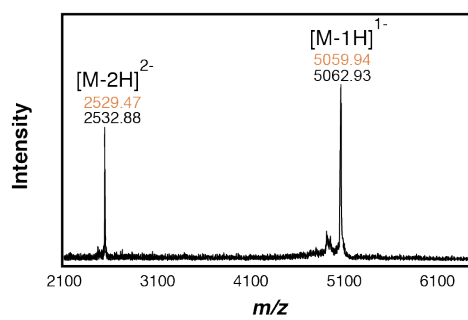
## Supporting Figure S1



**A**

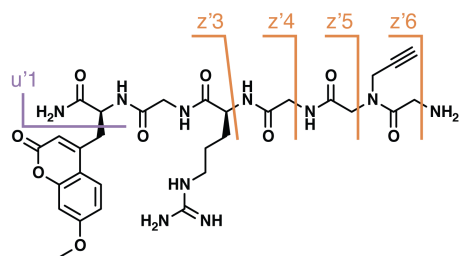


**B**

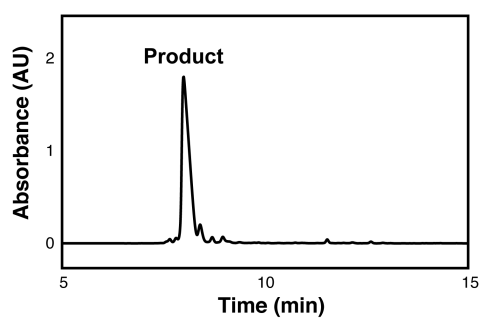


**Figure S2. Structure and characterization of synthesis resin linker.** The target linker features a coumarin chromophore for conversion yield quantitation, an arginine residue to enhance ionization efficiency during mass spectrometric analysis, and alkyne for coupling to N<sub>3</sub>-HDNA. ETD fragmentation yielded expected z ion products (orange) in addition to cleavage of the highly conjugated coumarin system (violet). (A) Reversed-phase HPLC analysis revealed a single predominant peak ( $\lambda = 330$  nm) corresponding to the desired product. (B) MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), z ion products (orange), and coumarin cleavage product (violet). Theoretical exact masses are shown in their respective color, observed  $m/z$  signals are shown in black.

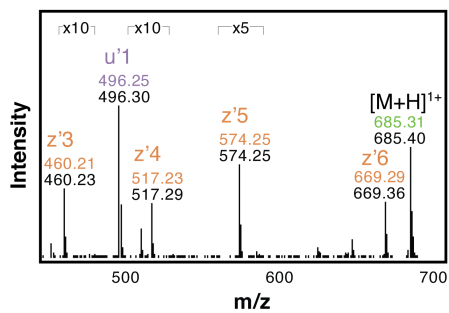
## Supporting Figure S2



**A**

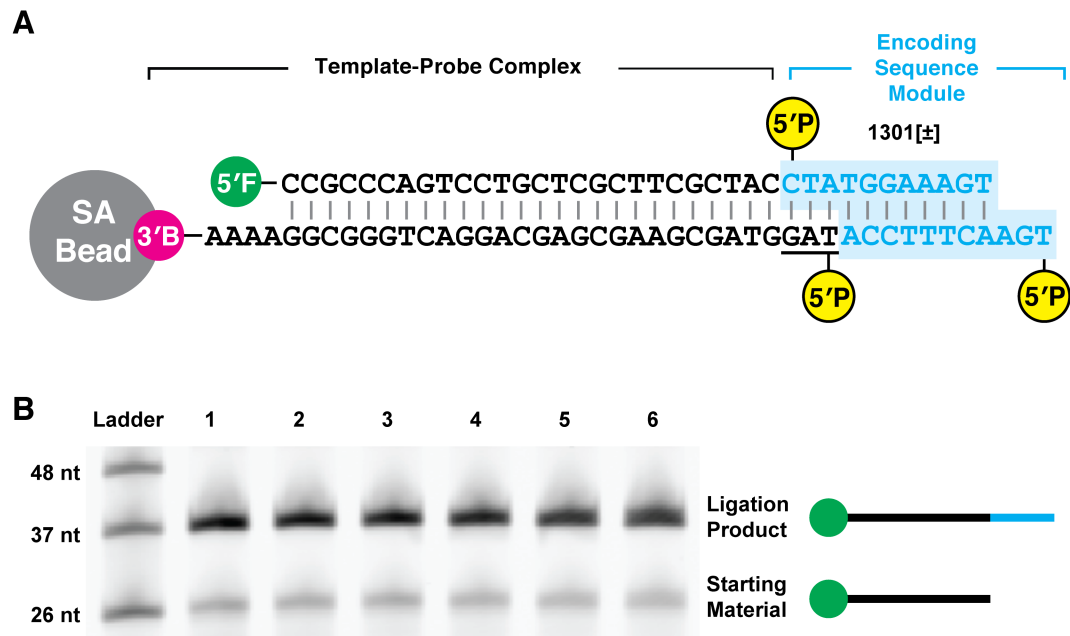


**B**



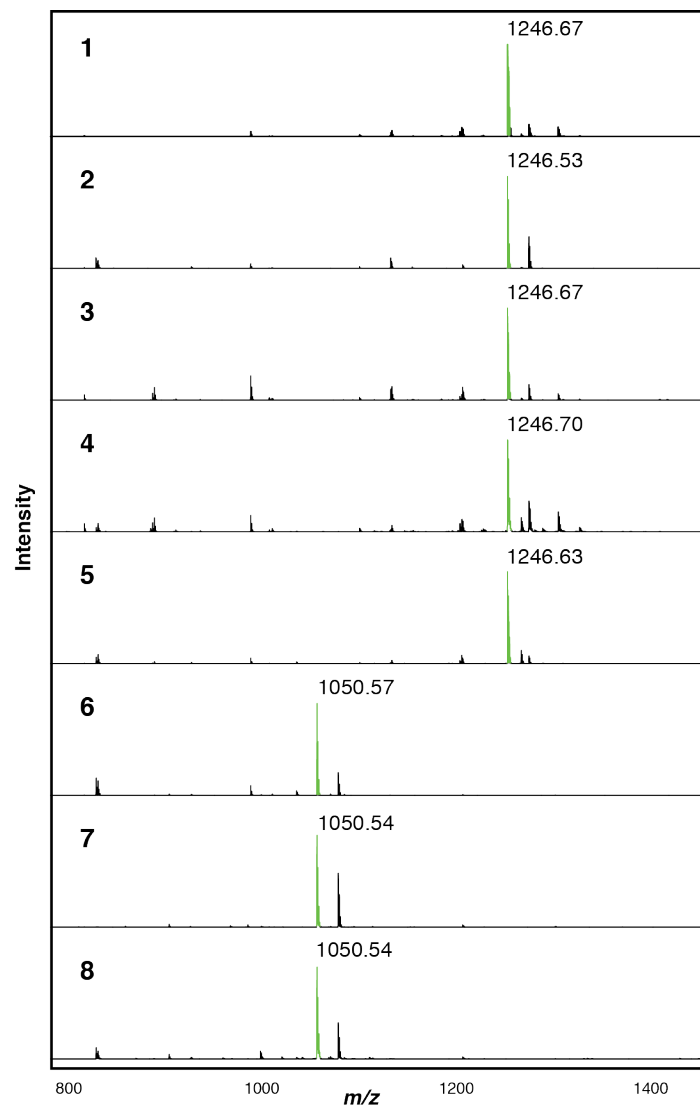
**Figure S3. Ligation yield assay schematic.** (A) Streptavidin-coated magnetic beads (SA Bead, gray) are first incubated with 3'-biotin- (3'B, magenta) labeled [-] strand template oligonucleotide that displays a 5'-phosphate (5'P, yellow). A 5'-fluorescein- (5'F, green) labeled [+] strand probe oligonucleotide is hybridized to the bead-immobilized [-] strand, yielding beads displaying [±] template-probe complex. The complex presents a 5'-phosphoryl-TAG-3' overhang (underlined) on the [-] strand for enzymatic ligation. A separately-prepared coding module, composed of two 5'-phosphorylated DNA oligonucleotides (here  $\approx$ **1301[+]** and  $\approx$ **1301[-]**) that are prehybridized to form a double-stranded coding module ( $\approx$ **1301[±]**, cyan), is added to the template-probe complex in the presence of T4 DNA ligase and ATP. (B) Fluorescent starting material (probe) and ligation product (probe covalently ligated with  $\approx$ **1301[+]**) are eluted from beads, electrophoretically resolved, and conversion of starting material to ligation product quantitated. Six replicate ligation reactions are shown (lanes 1–6) in the sample data. A ladder of fluorescent oligonucleotide standards is shown in the left-most lane.

## Supporting Figure S3



**Figure S4. Single-bead MALDI-TOF MS analysis.** Single-bead samples of the DNA-encoded synthesis of **1 – 8** were cleaved and subjected to MALDI-TOF MS analysis. The  $[M+H]^+$  is highlighted (green) in each spectrum. Stereo- and regioisomers **1 – 5** theoretical  $[M+H]^+$  of 1246.63 Da was observed for DESPS **1 – 5**, with  $< 80 \Delta\text{ppm}$ . Regioisomers **6 – 8** theoretical  $[M+H]^+$  of 1050.54 was observed for DESPS **6 – 8**, with  $< 29 \Delta\text{ppm}$ .

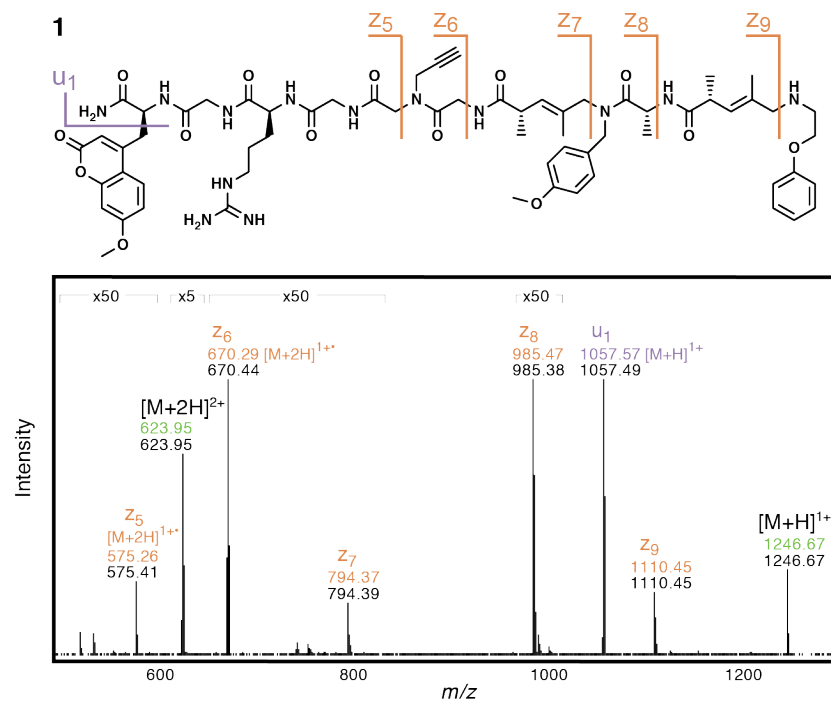
## Supporting Figure S4





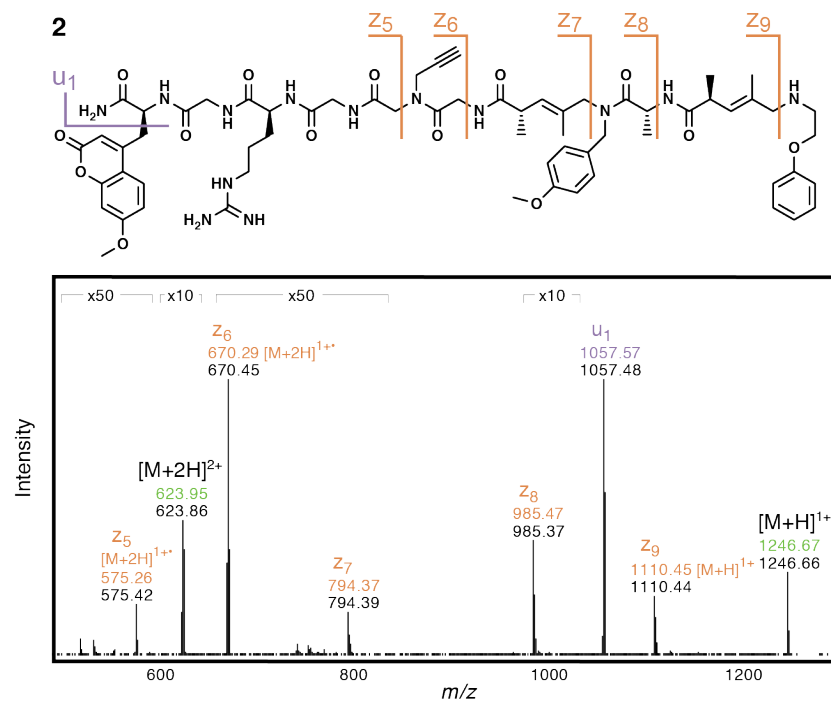
**Figure S5 - MS/MS-ETD fragmentation analysis of 1.** The DNA-encoded oligomer **1** structure displays expected ETD fragmentation to generate z ion products (orange) and coumarin fragmentation product (violet). MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), the expected z ion products (orange), and coumarin fragmentation product (violet). Theoretical exact masses are shown in their respective color, observed m/z signals are shown in black. Charge-reduced radical species that are the predominant ion product ions are indicated accordingly.

## Supporting Figure S5



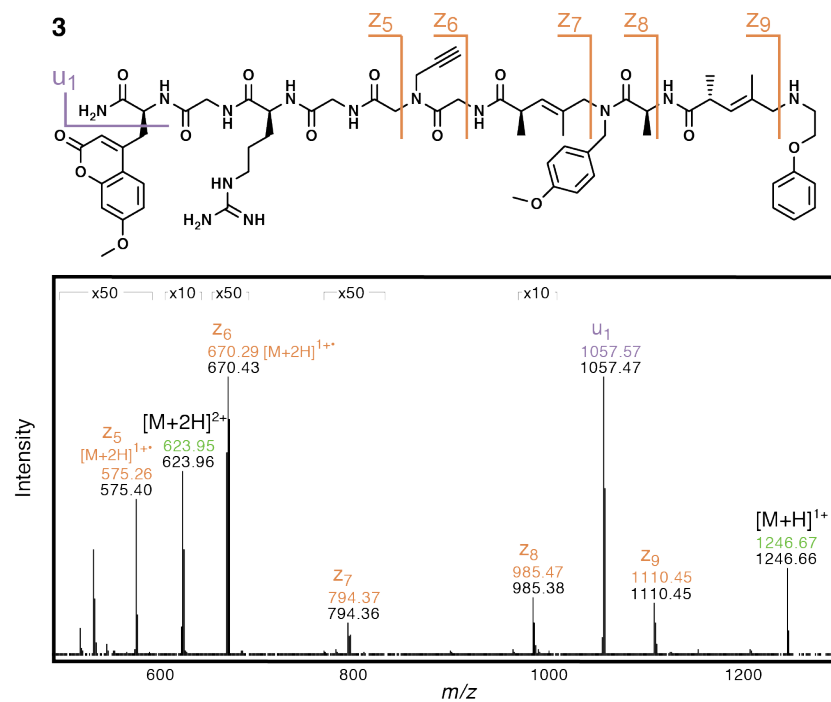
**Figure S6 - MS/MS-ETD fragmentation analysis of 2.** The DNA-encoded oligomer **2** structure displays expected ETD fragmentation to generate z ion products (orange) and coumarin fragmentation product (violet). MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), the expected z ion products (orange), and coumarin fragmentation product (violet). Theoretical exact masses are shown in their respective color, observed m/z signals are shown in black. Charge-reduced radical species that are the predominant ion product are indicated accordingly.

## Supporting Figure S6



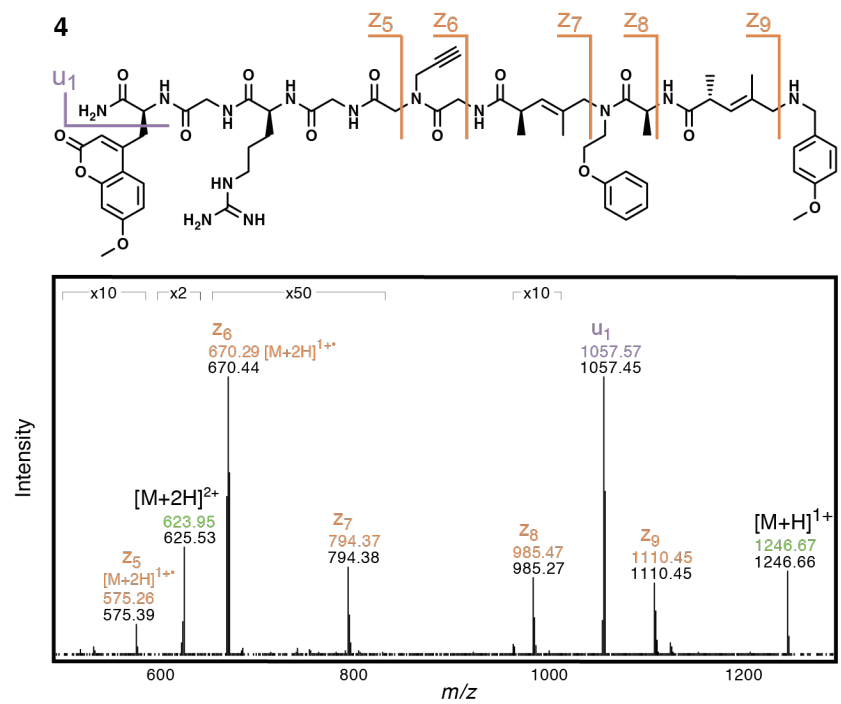
**Figure S7 - MS/MS-ETD fragmentation analysis of 3.** The DNA-encoded oligomer **3** structure displays expected ETD fragmentation to generate z ion products (orange) and coumarin fragmentation product (violet). MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), the expected z ion products (orange), and coumarin fragmentation product (violet). Theoretical exact masses are shown in their respective color, observed m/z signals are shown in black. Charge-reduced radical species that are the predominant ion product are indicated accordingly.

## Supporting Figure S7



**Figure S8 - MS/MS-ETD fragmentation analysis of 4.** The DNA-encoded oligomer **4** structure displays expected ETD fragmentation to generate z ion products (orange) and coumarin fragmentation product (violet). MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), the expected z ion products (orange), and coumarin fragmentation product (violet). Theoretical exact masses are shown in their respective color, observed m/z signals are shown in black. Charge-reduced radical species that are the predominant ion product are indicated accordingly.

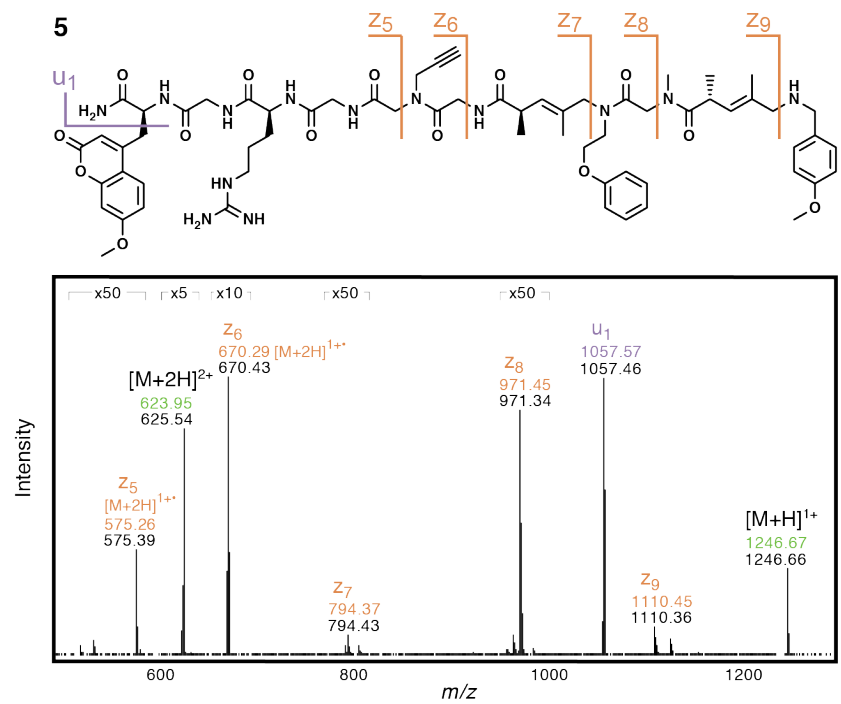
## Supporting Figure S8





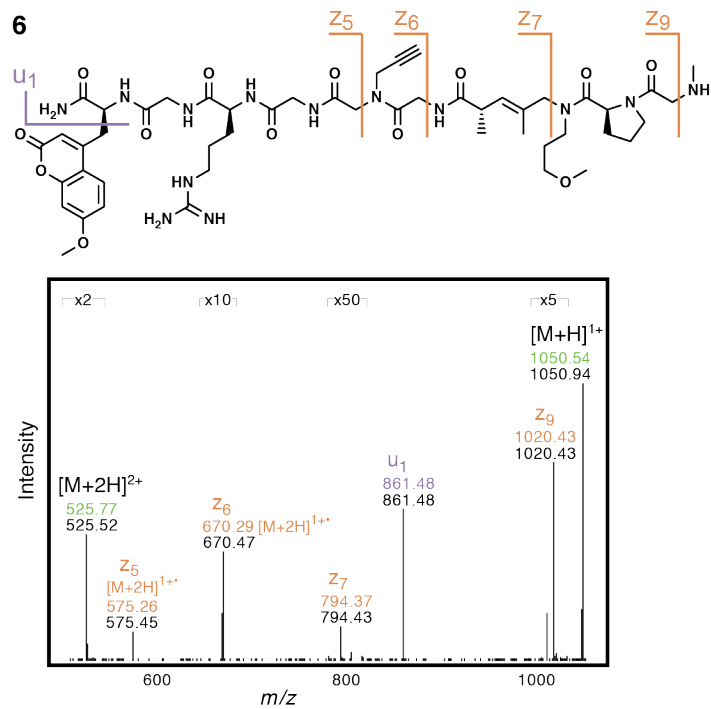
**Figure S9 - MS/MS-ETD fragmentation analysis of 5.** The DNA-encoded oligomer **5** structure displays expected ETD fragmentation to generate z ion products (orange) and coumarin fragmentation product (violet). MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), the expected z ion products (orange), and coumarin fragmentation product (violet). Theoretical exact masses are shown in their respective color, observed m/z signals are shown in black. Charge-reduced radical species that are the predominant ion product are indicated accordingly.

## Supporting Figure S9



**Figure S10 - MS/MS-ETD fragmentation analysis of 6.** The DNA-encoded oligomer **6** structure displays expected ETD fragmentation to generate z ion products (orange) and coumarin fragmentation product (violet). MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), the expected z ion products (orange), and coumarin fragmentation product (violet). Theoretical exact masses are shown in their respective color, observed m/z signals are shown in black. Charge-reduced radical species that are the predominant ion product are indicated accordingly.

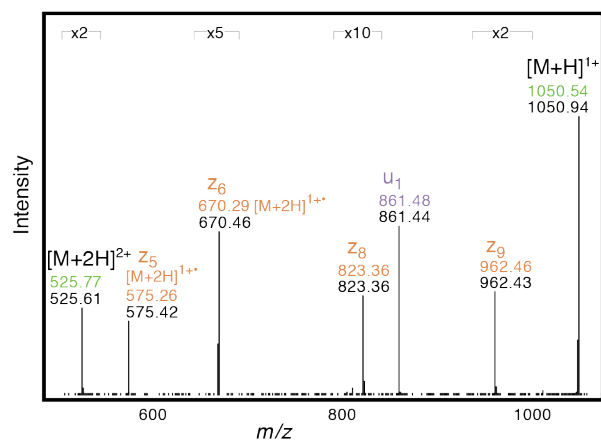
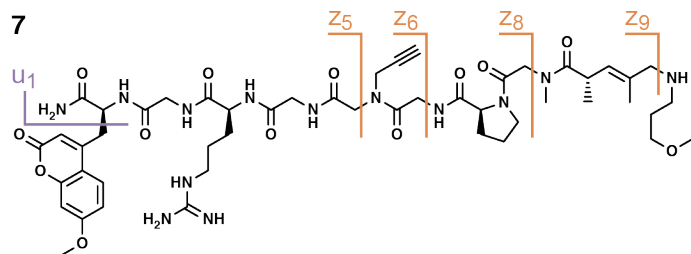
## Supporting Figure S10



**Figure S11 - MS/MS-ETD fragmentation analysis of 7.** The DNA-encoded oligomer 7 structure displays expected ETD fragmentation to generate z ion products (orange) and coumarin fragmentation product (violet). MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), the expected z ion products (orange), and coumarin fragmentation product (violet). Theoretical exact masses are shown in their respective color, observed m/z signals are shown in black. Charge-reduced radical species that are the predominant ion product are indicated accordingly.

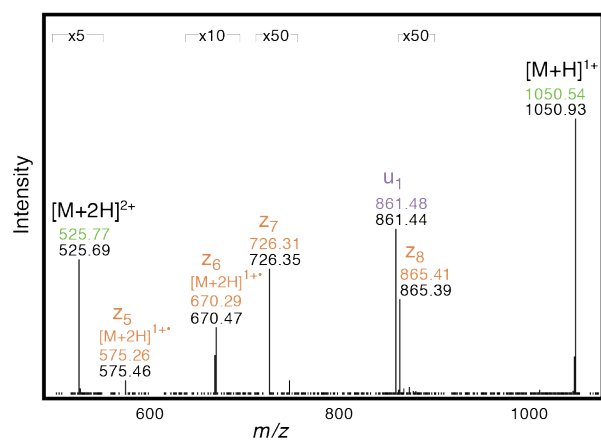
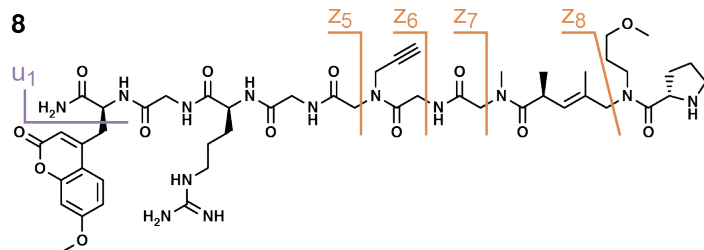
## Supporting Figure S11

7



**Figure S12 - MS/MS-ETD fragmentation analysis of 8.** The DNA-encoded oligomer **8** structure displays expected ETD fragmentation to generate z ion products (orange) and coumarin fragmentation product (violet). MS/MS analysis with ETD fragmentation of the parent  $[M+2H]^{2+}$  ion (green) yielded singly charged parent ion (green), the expected z ion products (orange), and coumarin fragmentation product (violet). Theoretical exact masses are shown in their respective color, observed m/z signals are shown in black. Charge-reduced radical species that are the predominant ion product are indicated accordingly.

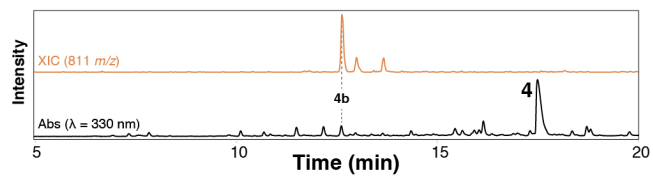
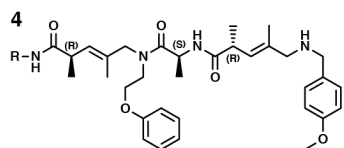
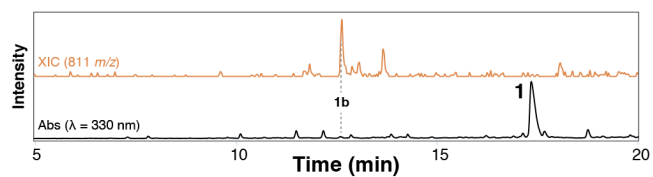
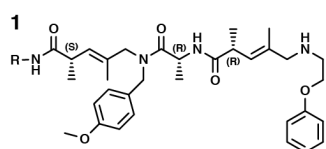
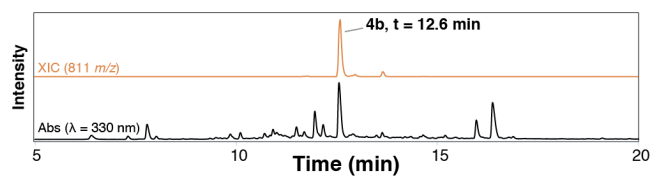
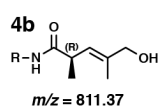
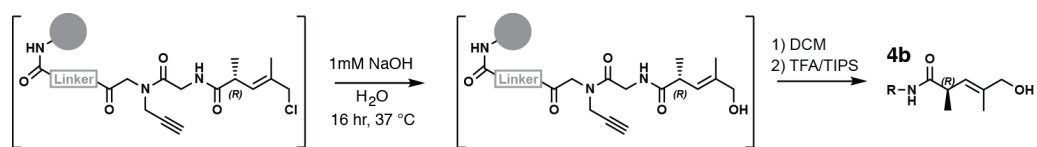
## Supporting Figure S12





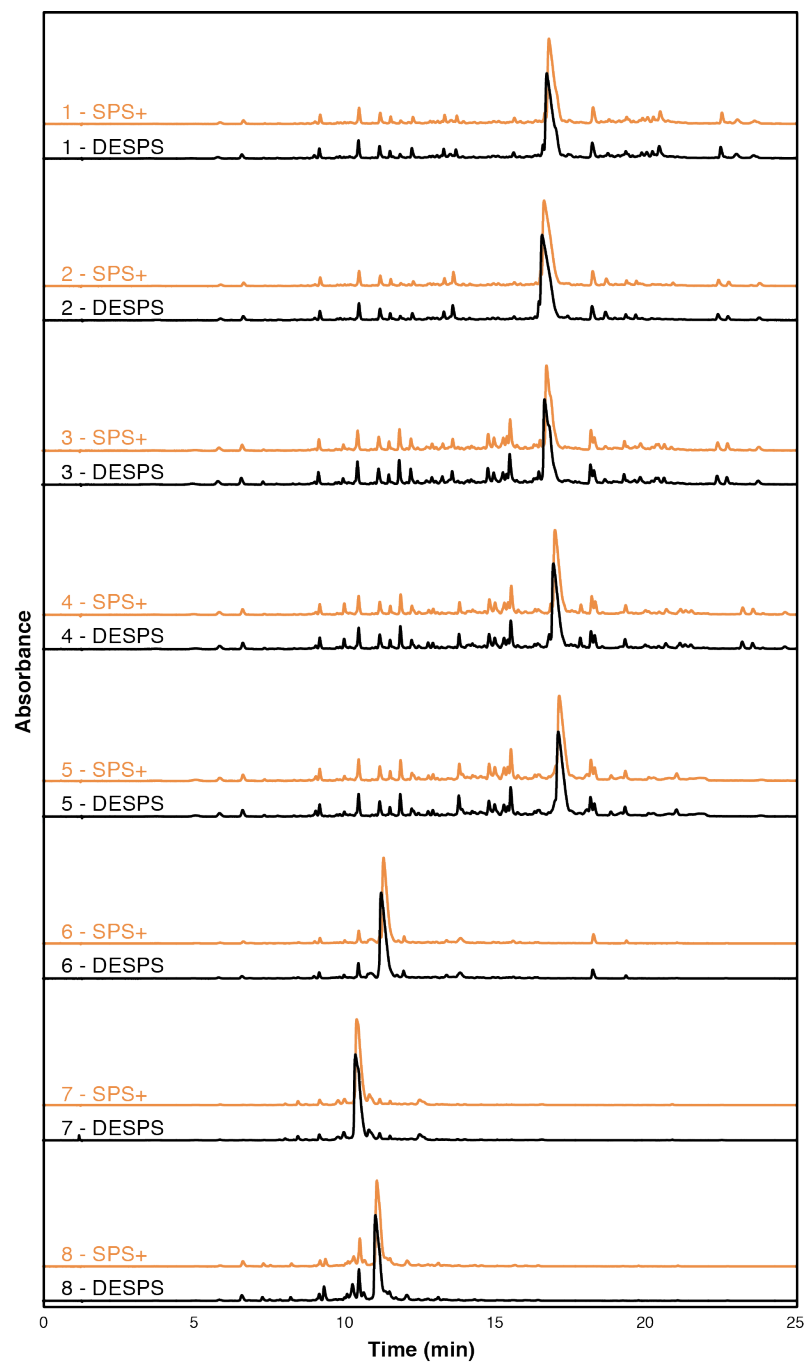
**Figure S13 - Example side product synthesis and LC-MS analysis.** Truncation side-product **4b** was prepared as a LC-MS characterization standard (top panel). The extracted ion chromatogram (XIC, 811  $m/z$ ) provided identification of side product **4b**. LC-MS analysis traces for SPS+ samples Co1 (middle chromatogram) and Co4 (bottom chromatogram) are shown with accompanying XIC (811  $m/z$ ) for comparison of truncation assignment.

## Supporting Figure S13



**Figure S14 - LC analysis of SPS+ and DESPS samples.** Reversed-phase HPLC analysis with absorbance detection ( $\lambda = 330$  nm) of both SPS+ and DESPS samples of **1 – 8** syntheses revealed that DNA encoding (with alternating exposure to buffered aqueous conditions) does not nominally change the profile of generated side products or the conversion of linker to product.

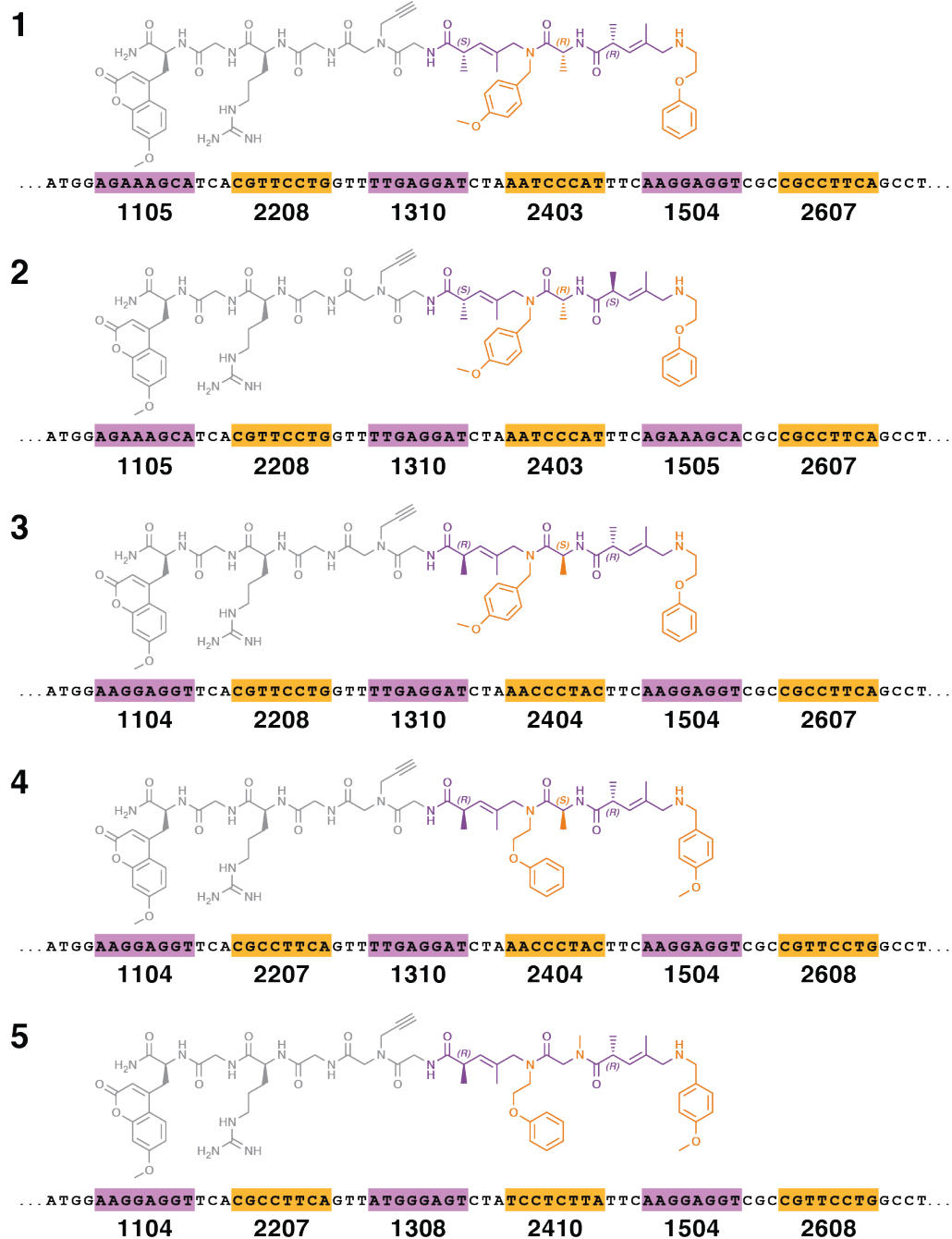
## Supporting Figure S14



**Figure S15 - Single-bead DNA sequence data and compound structure decoding.**

DNA-encoded compound structures **1** – **5** are presented above each compound's aligned sequence data. Structure color-coding indicates linker (gray), scaffold diversity elements (purple), and side chain diversity elements (orange). Encoding sequences are shown with corresponding color-coded highlighting. Overhang sequences are invariant (no highlighting). The order of synthesis steps correlates with 5'-to-3' DNA sequence. Numeric identifiers below each sequence correspond to encoded diversity elements found in the structure-identifier lookup table. Sequencing data were generated from PCR amplification products of single bead analyses. The average base call quality scores of **1** – **5** single-bead sequencing data are 45, 47, 49, 48, and 51, respectively.

## Supporting Figure S15

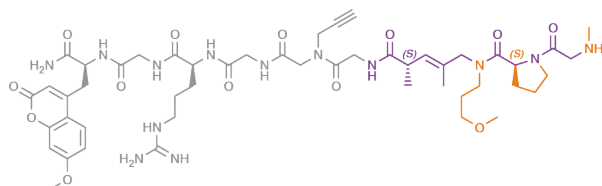


**Figure S16 - Single-bead DNA sequence data and compound structure decoding.**

DNA-encoded compound structures **6** – **8** are presented above each compound's aligned sequence data. Structure color-coding indicates linker (gray), scaffold diversity elements (purple), and side chain diversity elements (orange). Encoding sequences are shown with corresponding color-coded highlighting. Overhang sequences are invariant (no highlighting). The order of synthesis steps correlates with 5'-to-3' DNA sequence. Numeric identifiers below each sequence correspond to encoded diversity elements found in the structure-identifier lookup table. Sequencing data were generated from PCR amplification products of single bead analyses. The average base call quality scores of **6** – **8** single-bead sequencing data are 49, 51 and 49, respectively.

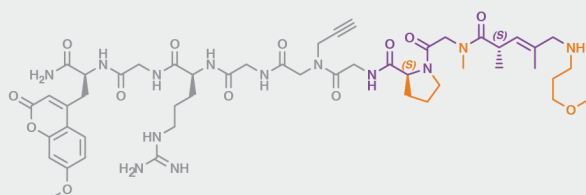
## Supporting Figure S16

6



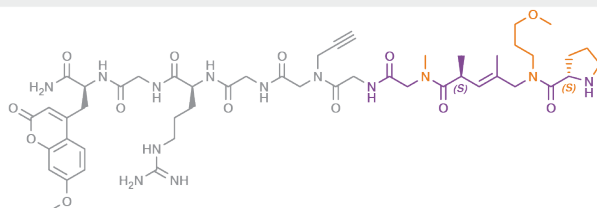
... ATGGAGAAAGCATCAAAACCTCAAGTTACGGAGCACTAAACCCTACTTCATGGGAGTCGCTCCTCTTAGCCT...  
1105 2202 1302 2404 1508 2610

7



... ATGGACGGAGCATCAAAACCTACGTTATGGGAGTCTATCCTCTTATTCAGAAAGCACGC AACCTCAAGCCT...  
1102 2204 1308 2410 1505 2602

8

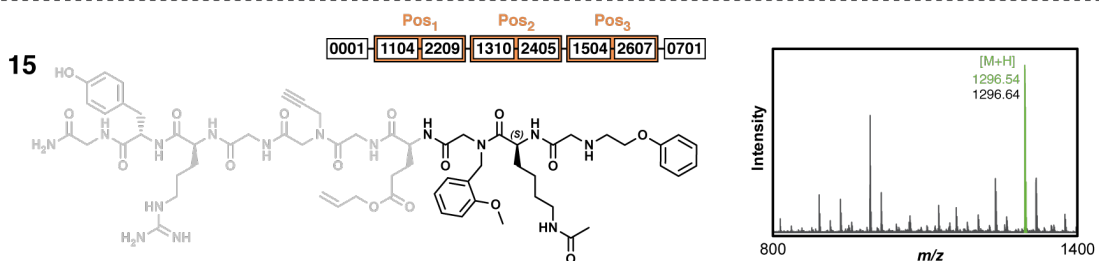
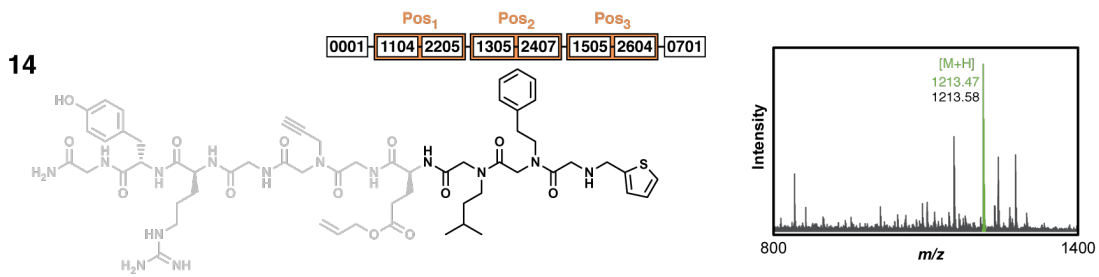
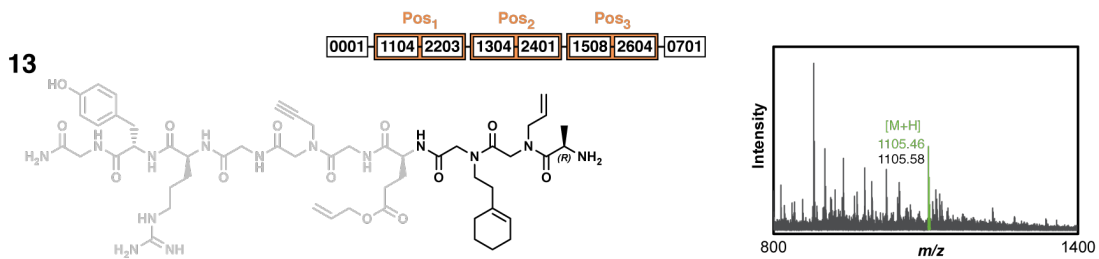
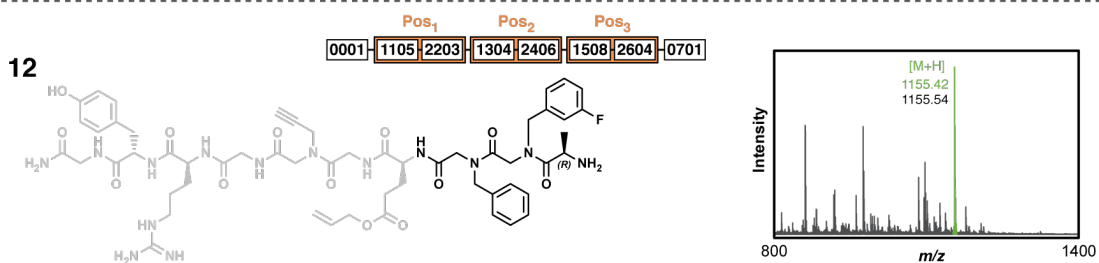
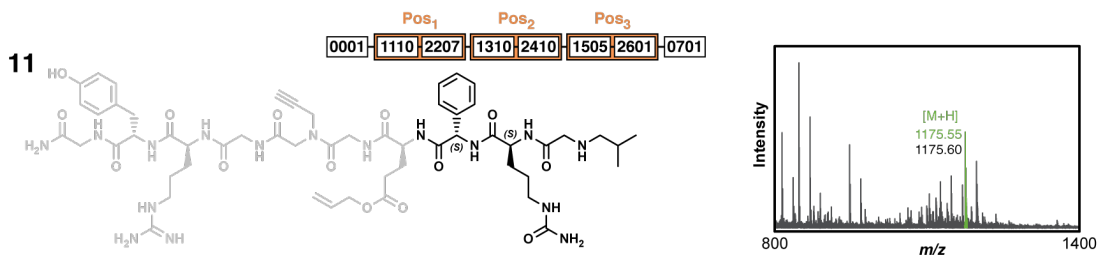


... ATGGATGGGAGTTCACTCCTCTTAGTTAGAAAGCACTAAACCCTCAATTCACGGAGCACGC AACCTACGCCT...  
1108 2210 1305 2402 1502 2604



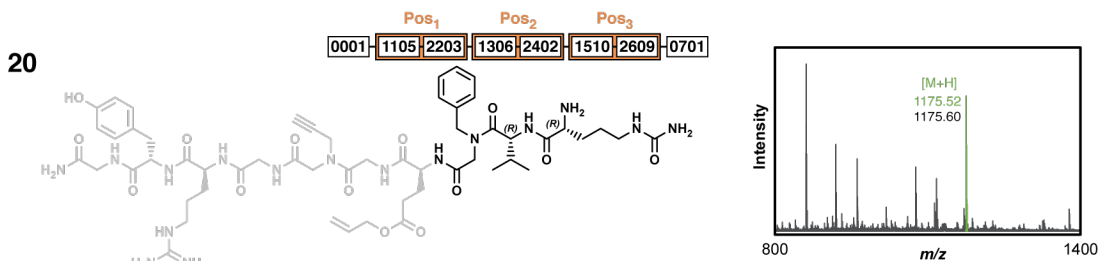
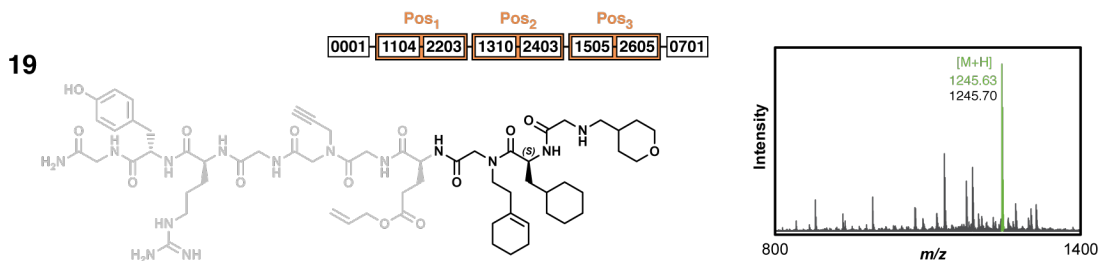
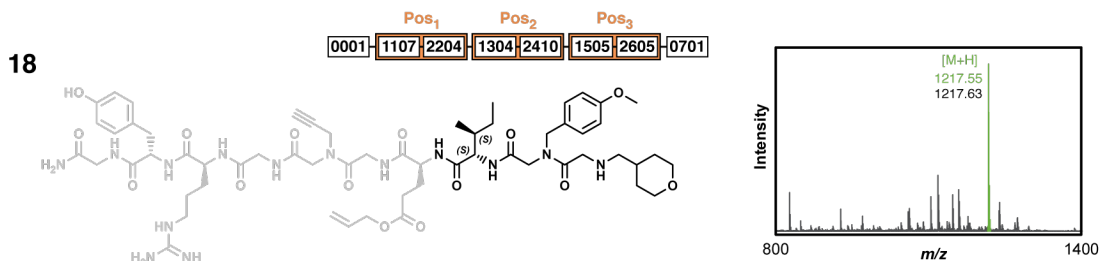
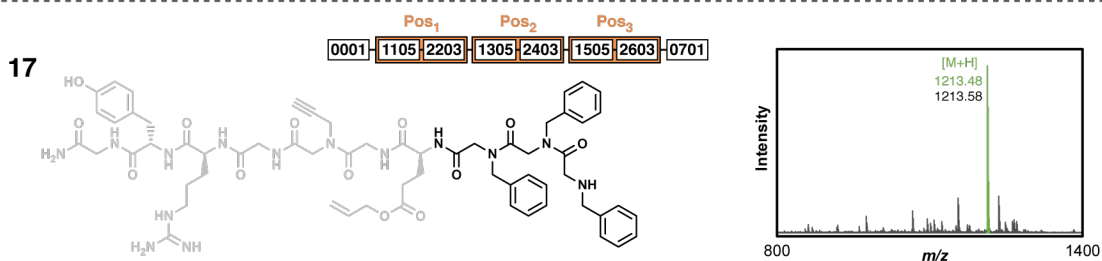
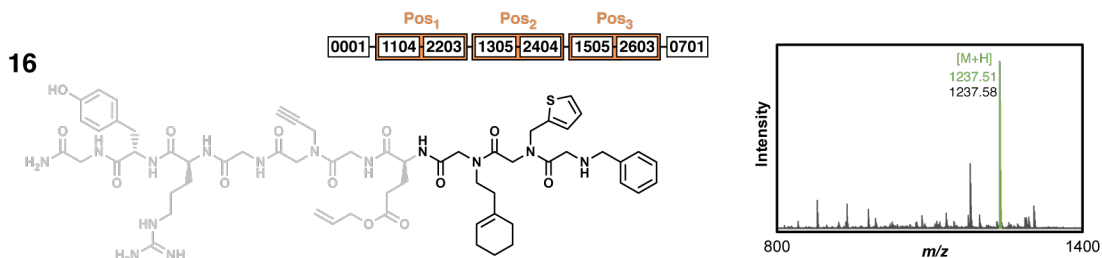
**Figure S17 - Combinatorial library synthesis quality control.** The DNA sequence amplified from each DNA-encoded 160- $\mu\text{m}$  library QC bead ( $N = 26$ , 19 shown) is displayed as a string of 4-digit identifiers, grouped by monomer position (orange, Pos<sub>1</sub> – Pos<sub>3</sub>), and flanked by the PCR primer identifiers (**0001**, **0701**). Decoding the sequence using the structure-identifier lookup table (Supporting Information Table T4) yields the predicted library compound structure for comparison to the corresponding 160- $\mu\text{m}$  library QC bead MALDI-TOF mass spectrum. The quality of the library can be evaluated based on the frequency of accurately predicting the observed compound mass based on sequence decoding ( $N = 26$  of 26).

# Supporting Figure S17



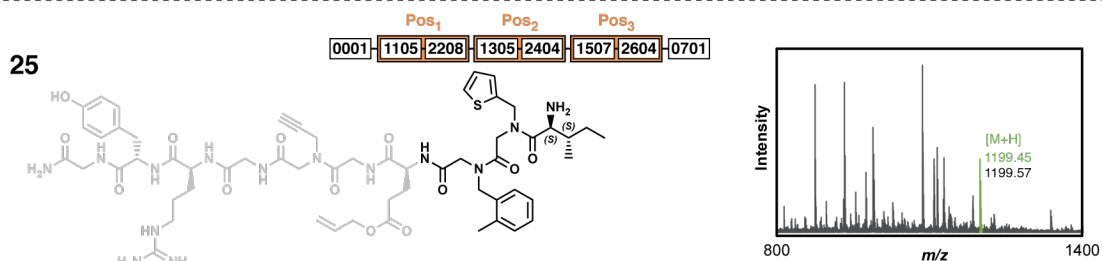
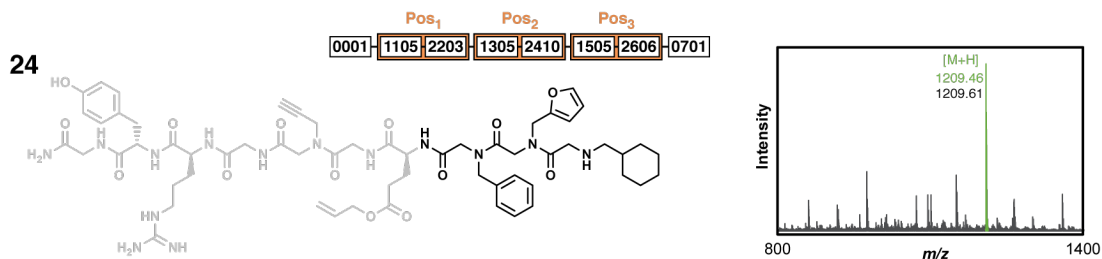
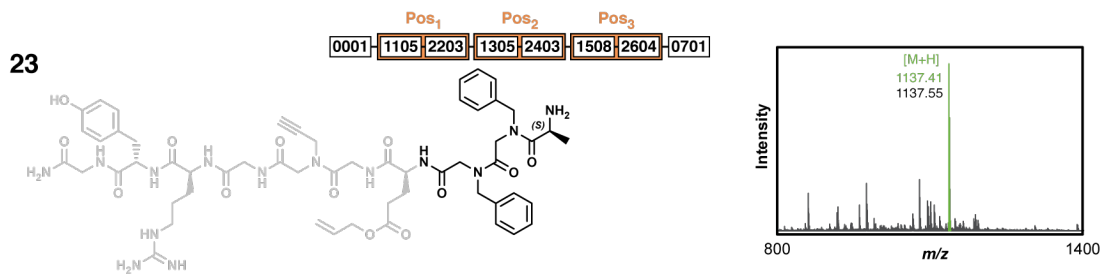
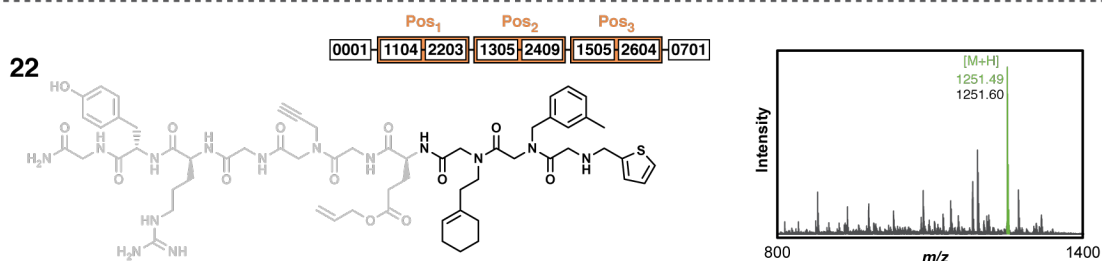
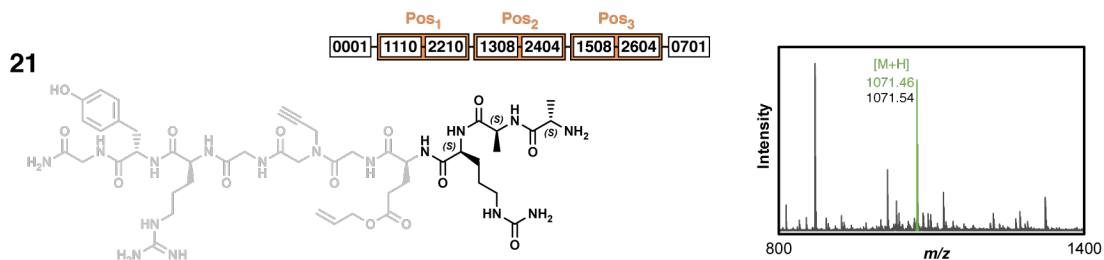
**Figure S18 - Combinatorial library synthesis quality control.** The DNA sequence amplified from each DNA-encoded 160- $\mu\text{m}$  library QC bead ( $N = 26$ , 19 shown) is displayed as a string of 4-digit identifiers, grouped by monomer position (orange, Pos<sub>1</sub> – Pos<sub>3</sub>), and flanked by the PCR primer identifiers (**0001**, **0701**). Decoding the sequence using the structure-identifier lookup table (Supporting Information Table T4) yields the predicted library compound structure for comparison to the corresponding 160- $\mu\text{m}$  library QC bead MALDI-TOF mass spectrum. The quality of the library can be evaluated based on the frequency of accurately predicting the observed compound mass based on sequence decoding ( $N = 26$  of 26).

# Supporting Figure S18



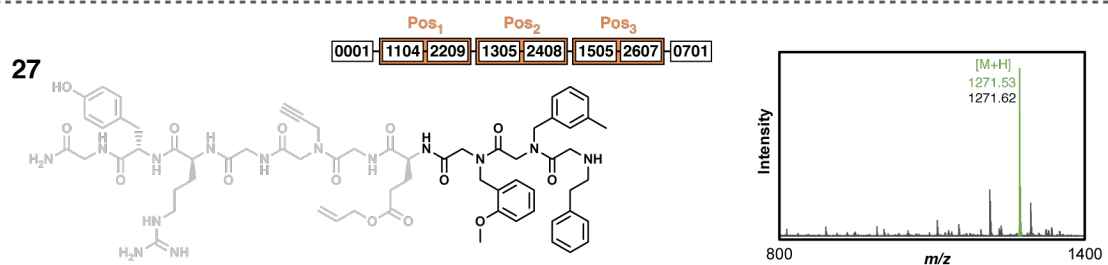
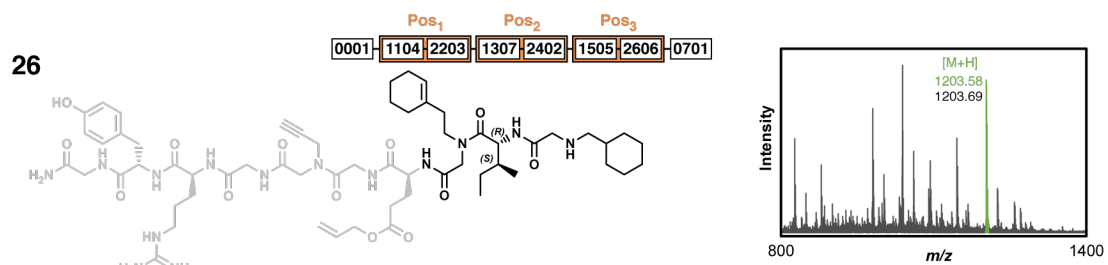
**Figure S19 - Combinatorial library synthesis quality control.** The DNA sequence amplified from each DNA-encoded 160- $\mu\text{m}$  library QC bead ( $N = 26$ , 19 shown) is displayed as a string of 4-digit identifiers, grouped by monomer position (orange, Pos<sub>1</sub> – Pos<sub>3</sub>), and flanked by the PCR primer identifiers (**0001**, **0701**). Decoding the sequence using the structure-identifier lookup table (Supporting Information Table T4) yields the predicted library compound structure for comparison to the corresponding 160- $\mu\text{m}$  library QC bead MALDI-TOF mass spectrum. The quality of the library can be evaluated based on the frequency of accurately predicting the observed compound mass based on sequence decoding ( $N = 26$  of 26).

## Supporting Figure S19



**Figure S20 - Combinatorial library synthesis quality control.** The DNA sequence amplified from each DNA-encoded 160- $\mu\text{m}$  library QC bead ( $N = 26$ , 19 shown) is displayed as a string of 4-digit identifiers, grouped by monomer position (orange, Pos<sub>1</sub> – Pos<sub>3</sub>), and flanked by the PCR primer identifiers (**0001**, **0701**). Decoding the sequence using the structure-identifier lookup table (Supporting Information Table T4) yields the predicted library compound structure for comparison to the corresponding 160- $\mu\text{m}$  library QC bead MALDI-TOF mass spectrum. The quality of the library can be evaluated based on the frequency of accurately predicting the observed compound mass based on sequence decoding ( $N = 26$  of 26).

## Supporting Figure S20





**Supporting Table T1. Oligonucleotide sequence lookup table.**

## Supporting Table T1

Overhangs		Coding Sequences			
Overhang #	Sequence	Identifier #	Sequence	Identifier #	Sequence
≈X1XX[+]	/5Phos/ATGG	≈1X01[+]	AAGAGAGG	≈2X01[+]	AGTTTCAG
≈X1XX[-]	/5Phos/TGA	≈1X01[-]	CCTCTCTT	≈2X01[-]	CTGAAACT
≈X2XX[+]	/5Phos/TCA	≈1X02[+]	ACGGAGCA	≈2X02[+]	AACCTCAA
≈X2XX[-]	/5Phos/AAC	≈1X02[-]	TGCTCCGT	≈2X02[-]	TTGAGGTT
≈X3XX[+]	/5Phos/GTT	≈1X03[+]	ACAAAGAG	≈2X03[+]	AATCCCAT
≈X3XX[-]	/5Phos/TAG	≈1X03[-]	CTCTTTGT	≈2X03[-]	ATGGGATT
≈X4XX[+]	/5Phos/CTA	≈1X04[+]	AAGGAGGT	≈2X04[+]	AACCCTAC
≈X4XX[-]	/5Phos/GAA	≈1X04[-]	ACCTCCTT	≈2X04[-]	GTAGGGTT
≈X5XX[+]	/5Phos/TTC	≈1X05[+]	AGAAAGCA	≈2X05[+]	ATCCTCTC
≈X5XX[-]	/5Phos/GCG	≈1X05[-]	TGCTTTCT	≈2X05[-]	GAGAGGAT
≈X6XX[+]	/5Phos/CGC	≈1X06[+]	ATAAAGGT	≈2X06[+]	ATTCTCCG
≈X6XX[-]	/5Phos/AGGC	≈1X06[-]	ACCTTTAT	≈2X06[-]	CGGAGAAT
		≈1X07[+]	ATAGAAGG	≈2X07[+]	CGCCTTCA
		≈1X07[-]	CCTTCTAT	≈2X07[-]	TGAAGGGG
		≈1X08[+]	ATGGGAGT	≈2X08[+]	CGTTCCTG
		≈1X08[-]	ACTCCCAT	≈2X08[-]	CAGGAACG
		≈1X09[+]	GCAAAGGA	≈2X09[+]	CTCTCCAC
		≈1X09[-]	TCCTTTGC	≈2X09[-]	GTGGAGAG
		≈1X10[+]	TTGAGGAT	≈2X10[+]	TCCTCTTA
		≈1X10[-]	ATCCTCAA	≈2X10[-]	TAAGAGGA

Primers	
PCR Primer #	Sequence
≈0001[+]	/5Phos/GCCGCCAGTCTGCTCGCTTCGCTAC
≈0001[-]	/5Phos/CCATGTAGCGAAGCGAGCAGGACTGGGCGGCGG
≈0701[+]	/5Phos/GCCTGTTGCCCCCAGTTGTTGTGCCAC
≈0701[-]	/5AmMC6/GTGGCACAACAACACTGGCGGGCAAAC

**Supporting Table T2. Microplate format and oligonucleotide usage in control compound encoded syntheses.**


## Supporting Table T2


		Microplate Column / Compound							
		1	2	3	4	5	6	7	8
<b>A</b>	OP1	≈1105[±]	≈1105[±]	≈1104[±]	≈1104[±]	≈1104[±]	≈1105[±]	≈1102[±]	≈1108[±]
	OP2	≈2208[±]	≈2208[±]	≈2208[±]	≈2207[±]	≈2207[±]	≈2202[±]	≈2204[±]	≈2210[±]
	OP3	≈1310[±]	≈1310[±]	≈1310[±]	≈1310[±]	≈1308[±]	≈1302[±]	≈1308[±]	≈1305[±]
	OP4	≈2403[±]	≈2403[±]	≈2404[±]	≈2404[±]	≈2410[±]	≈2404[±]	≈2410[±]	≈2402[±]
	OP5	≈1504[±]	≈1505[±]	≈1504[±]	≈1504[±]	≈1504[±]	≈1508[±]	≈1505[±]	≈1502[±]
	OP6	≈2607[±]	≈2607[±]	≈2607[±]	≈2608[±]	≈2608[±]	≈2610[±]	≈2602[±]	≈2604[±]
<b>B</b>	OP1	≈1105[±]	≈1105[±]	≈1104[±]	≈1104[±]	≈1104[±]	≈1105[±]	≈1102[±]	≈1108[±]
	OP2	≈2208[±]	≈2208[±]	≈2208[±]	≈2207[±]	≈2207[±]	≈2202[±]	≈2204[±]	≈2210[±]
	OP3	≈1310[±]	≈1310[±]	≈1310[±]	≈1310[±]	≈1308[±]	≈1302[±]	≈1308[±]	≈1305[±]
	OP4	≈2403[±]	≈2403[±]	≈2404[±]	≈2404[±]	≈2410[±]	≈2404[±]	≈2410[±]	≈2402[±]
	OP5	≈1504[±]	≈1505[±]	≈1504[±]	≈1504[±]	≈1504[±]	≈1508[±]	≈1505[±]	≈1502[±]
	OP6	≈2607[±]	≈2607[±]	≈2607[±]	≈2608[±]	≈2608[±]	≈2610[±]	≈2602[±]	≈2604[±]
<b>C</b>	SPS+								


**Supporting Table T3. Control compound structure-identifier lookup table.**


# Supporting Table T3


**SCAFFOLD ENCODING**











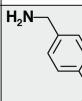
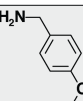
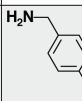











SIDE CHAIN ENCODING	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010
2001										
2002										
2003										
2004										
2005										
2006										
2007										
2008										
2009										
2010				H <sub>2</sub> N-Me	H <sub>2</sub> N-Me			H <sub>2</sub> N-Me		

**Supporting Table T4. Combinatorial library structure-identifier lookup table.**

# Supporting Table T4

**SCAFFOLD ENCODING**



	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010
2001										
2002										
2003										
2004										
2005										
2006										
2007										
2008										
2009										
2010										

Scaffold elements **1001/20xx** are only utilized in Pos<sub>2</sub>. Scaffold elements **1002/2008**, **1003/2008**, and **1009/2008** are not used in Pos<sub>2</sub>.