Supplementary Information for

Sequence-directed Dynamic Covalent Assembly of Base-4-encoded Oligomers

Samuel C. Leguizamon,[a†] Megan F. Dunn,[a†] and Timothy F. Scott*[,b,c]

[a] Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48105, USA.

[b] Department of Chemical Engineering, Monash University, Clayton, VIC 3800, Australia.

[c] Department of Materials Science and Engineering, Monash University, Clayton, VIC 3800, Australia.

† These authors contributed equally to this work

General experimental procedures

^{1}H NMR and ^{11}B NMR spectra of the monomers were collected using a Varian MR400 spectrometer. Chemical shifts were measured in δ (ppm) relative to residual solvent (CD_3CN=1.94). Electrospray ionization (ESI) mass spectra were recorded using an Agilent Q-TOF 1200 series spectrometer in positive ion mode. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectra were collected by utilizing a Bruker Autoflex mass spectrometer used in reflectron mode with both positive and negative ionizations as indicated. Reverse phase high performance liquid chromatography (RP-HPLC) was performed using both a preparative reversed phase Phenomenex Luna C18(2) columns with a linear gradient of water and acetonitrile as the eluent at 30°C as well as an analytical scale column. The RP-HPLC system was equipped with dual Shimadzu LC-6AD HPLC pump, Shimadzu FRC 70A fraction collector, and monitored using Shimadzu Prominence detector at 214 nm. Unless otherwise noted all reagents and materials were purchased from Sigma Aldrich, AK Scientific, Oakwood Products, and TCI America.

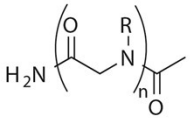
Oligomer synthesis

The peptoid-based oligomers were prepared using a microwave-assisted Liberty Blue automated peptide synthesizer (CEM Corporation) that had been modified to synthesize peptoids. The peptoids of the imine-only system were synthesized on a photolabile solid support resin to maintain the acid-labile ethylene acetal protecting group. The remaining peptoids were synthesized on acid labile rink amide resin. Both of the resins contain a fluorenylmethyloxycarbonyl (Fmoc)-protected amine that is initially deprotected prior to synthesis by treatment in 4-methylpiperidine:dimethylformamide (DMF) (20:80, volume ratio) to yield a terminal amine on the solid support. The synthesis then proceeds by a sequential addition reaction whereby a terminal amine is acetylated with 1 M bromoacetic acid using 1.2 M diisopropylcarbodiimide (DIC) as an activator for 5 minutes at 75°C, to afford a terminal bromide which is subsequently displaced via nucleophilic substitution with a 0.5 M primary amine for 5 minutes at 75°C. This two-step process is followed to synthesize the different predefined sequences. The *N*-terminal of the complementary oligomers was capped with 1 M acetic anhydride activated with DIC to prevent further chain elongation. The primary amines fall into two categories, dynamic covalent functional groups and inert spacer monomers (Table S1). The dynamic covalent functional group consisted of the prepared dopamine (acetonide) (Nace), 4-(2-aminoethyl)-*N*-(*tert*-butoxycarbonyl)phenylamine (Nbam),[1] 4-(2-aminoethyl)-N-(allylcarbonyloxy)phenylamine (Npam),[2] 4-(1,3-dioxacyclopent-2-yl)benzylamine (Npal),[1] and commercially available 4-aminomethylphenylboronic acid, pinacol ester (Npbe) purchased from AccelaChem. The inert spacer monomers were commercially available 2-methoxyethylamine (Nme) and 2-ethoxyethoxyethylamine (Neee) prepared from the established protocol of Wei et al.[1] All of the reagents were prepared in dimethylformamide (DMF) with the exception of the

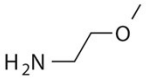
monomers used in the imine only system that were prepared in *N*-methyl-2-pyrrolidone (NMP) for increased solubility.

Table S1. Primary amine monomers used in this study.

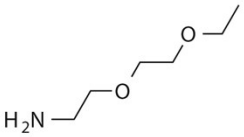
Peptoid Oligomer Backbone



Inert Spacer Monomers

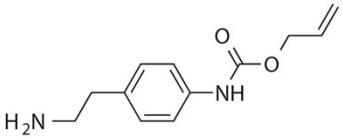


Nme = 2-methoxyethylamine

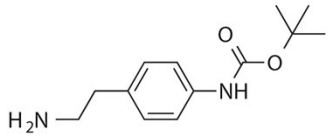


Neee = 2-(2-ethoxyethoxy)ethylamine

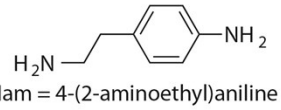
Dynamic Covalent Monomers



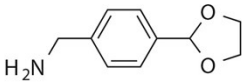
Npam = 4-(2-aminoethyl)-N-(allylcarbonyloxy)phenylamine



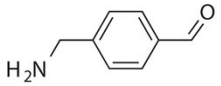
Nbam = tert-butyl-(4-(2-aminoethyl)phenyl)carbamate



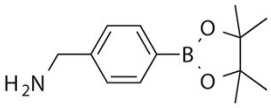
Nam = 4-(2-aminoethyl)aniline



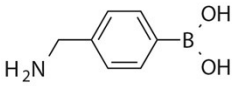
Npal = 4-(1,3-dioxacyclopent-2-yl)benzylamine



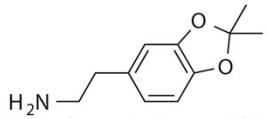
Nal = 4-(aminomethyl)benzaldehyde



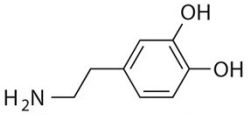
Npbe = (4-(aminomethyl)phenyl)boronic acid, pincacol ester



Npba = (4-(aminomethyl)phenyl)boronic acid



Nace = dopamine(acetonide)



Ndop = dopamine

Oligomer deprotection and purification

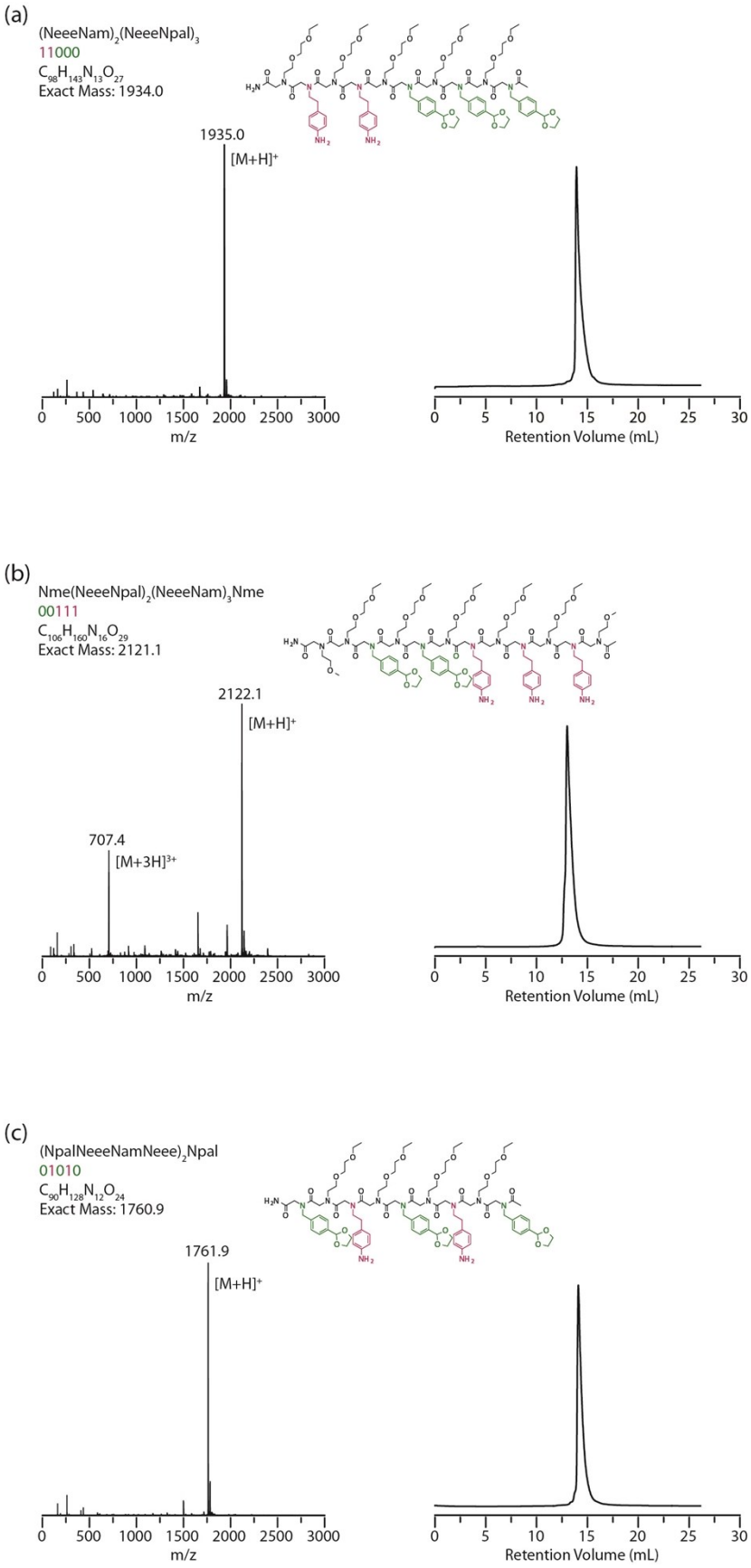
Peptoids bearing Alloc-protected amines (imine only system) were deprotected via an adaptation of a previously reported approach.[2] On-resin peptoids were suspended in dry DCM and treated with 0.1 equivalents tetrakis(triphenylphosphine)palladium(0) and 25 equivalents of phenylsilane per Alloc group for one hour. After filtration, deprotection was repeated and the photo-labile resin was subsequently cleaved in DMF for 36 hours under irradiation at approximately 25 mW/cm^2 with 405 nm. The cleavage solution was filtered and evaporated to dryness under vacuum.

All other peptoids were cleaved from acid-labile Rink amide resin by a 5 minute incubation with a cleavage cocktail containing 95% trifluoroacetic acid (TFA) and 5% water in a glass fritted reaction vessel. The resin was then rinsed with dichloromethane (DCM) to remove any residual peptoid on the resin. The solvents (DCM and trace water) and TFA were removed by blowing with a N_2 stream. The acid-labile protecting groups acetonide, acetal, and Boc were removed by TFA during the cleavage. The pinacol protecting group is not compressively removed during the cleavage step, but the remaining pinacol can be removed during purification with reverse phase high performance chromatography (RP-HPLC).[3] Once only the peptoid residue remained, the peptoids were put into a 50:50 solution of 0.1% TFA in acetonitrile and 0.1% TFA in water (approximate pH of 3) this was to influence the equilibrium of the dynamic covalent reactions towards the initial functional groups.

The oligopeptoids were purified by preparative scale RP-HPLC using a linear gradient of acetonitrile and water. Major peaks were collected and fractions were combined before utilizing electrospray ionization (ESI) to confirm the identity of each strand. Fractions of the peptoids were combined and lyophilized to a white powder.

Table S2. Nomenclature of peptoids used in this study, the associated sequence and exact mass.

Peptoid	Sequence	Exact mass (g/mol)
	Imine system	
11000	$(NeeeNam)_2(NeeeNpal)_3$	1934.0
00111	$Nme(NeeeNpal)_2(NeeeNam)_3Nme$	2121.1
01010	$(NpalNeeeNamNee)_2Npal$	1760.9
10101	$(NamNeeeNpalNee)_2NamNme$	1832.9
	Boronate ester system	
222333	$(NmeNdop)_2(NmeNpba)_2$	2016.9
223233	NmeNdopNmeNdopNmeNpbaNmeNdopNmeNpbaNpba	1901.9
	Base-4 system	
303030	$(NdopNal)_3$	1163.5
222	$(NmeNpba)_3Nme$	1092.5
111	$(NeeeNam)_3$	1106.6
213112	$NmeNpbaNmeNamNmeNdop(NmeNam)_2NmeNpbaNme$	1968.0
300203	NmeNdopNmeNalNmeNalNmeNpbaNmeNalNmeNdop	1851.8
023	NmeNalNmeNpbaNmeNdop	963.4
123321	$NamNpbaNdop_2NpbaNam$	1179.5


(a)
(NeeeNam)2(NeeeNpal)3
11000
C98H143N13O27
Exact Mass: 1934.0
1935.0
[M+H]+
0 500 1000 1500 2000 2500 3000
m/z
0 5 10 15 20 25 30
Retention Volume (mL)
(b)
Nme(NeeeNpal)2(NeeeNam)3Nme
00111
C106H160N16O29
Exact Mass: 2121.1
2122.1
[M+H]+
707.4
[M+3H]3+
0 500 1000 1500 2000 2500 3000
m/z
0 5 10 15 20 25 30
Retention Volume (mL)
(c)
(NpalNeeeNamNeee)2Npal
01010
C90H128N12O24
Exact Mass: 1760.9
1761.9
[M+H]+
0 500 1000 1500 2000 2500 3000
m/z
0 5 10 15 20 25 30
Retention Volume (mL)

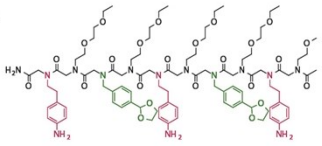
(d)

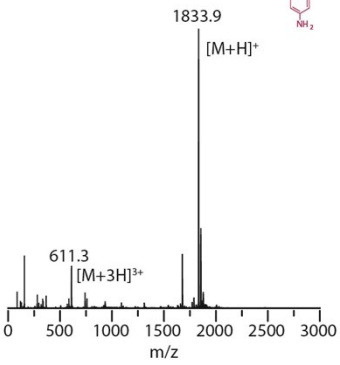
$(NamNeeeNpalNeee)_2NamNme$

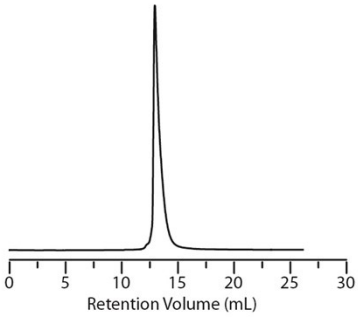
10101

$C_{93}H_{136}N_{14}O_{24}$

Exact Mass: 1832.9







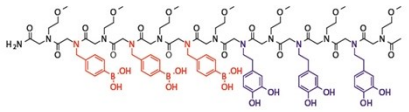
(e)

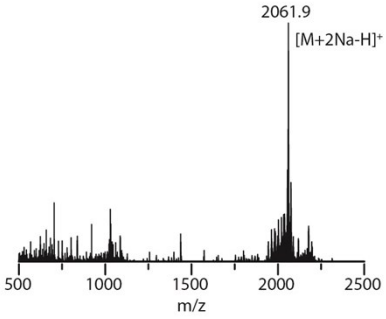
$(NmeNpba)_3(NmeNdop)_3Nme$

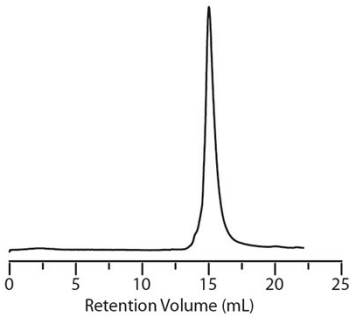
222333

$C_{94}H_{131}B_3N_{14}O_{33}$

Exact Mass: 2016.9







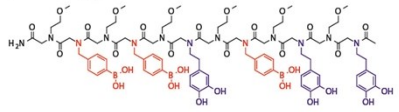
(f)

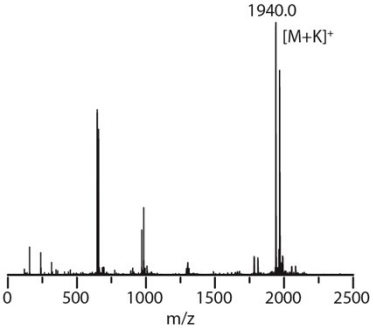
$(NmeNpba)_2NmeNdopNmeNpba(NmeNdop)_2$

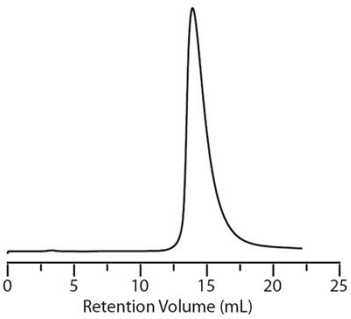
223233

$C_{89}H_{122}B_3N_{13}O_{31}$

Exact Mass: 1901.9







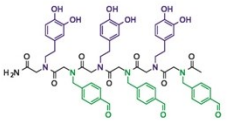
(g)

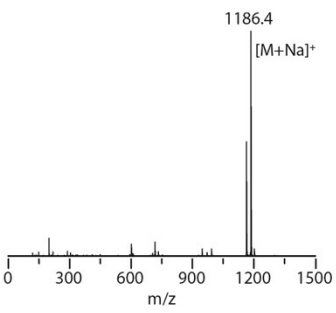
$(NdopNal)_3$

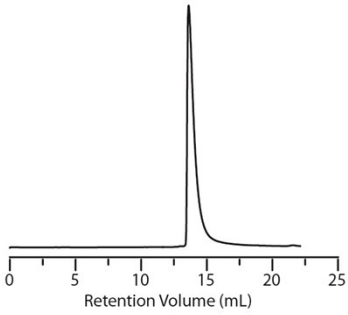
303030

$C_{62}H_{65}N_7O_{16}$

Exact Mass: 1163.5







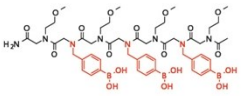
(h)

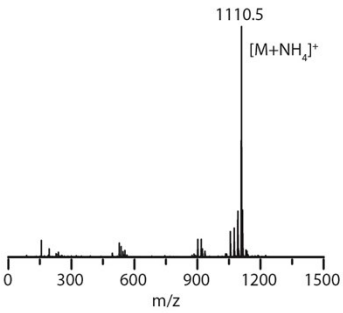
$(NmeNpba)_3Nme$

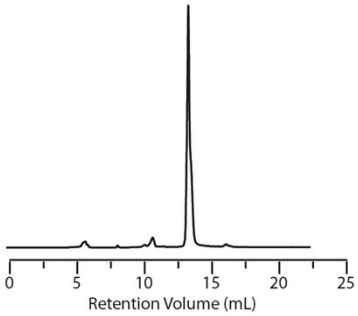
222

$C_{49}H_{71}B_3N_8O_{18}$

Exact Mass: 1092.5







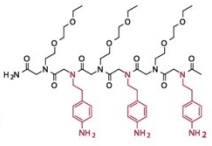
(i)

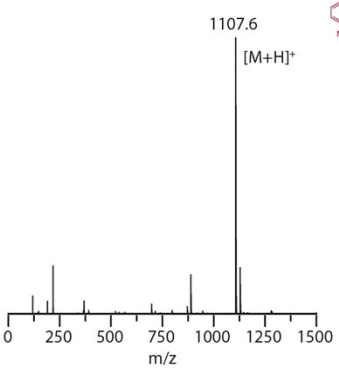
$(NeeeNam)_3$

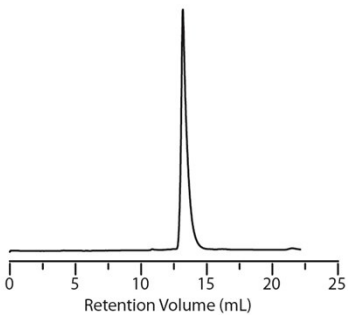
111

$C_{56}H_{86}N_{10}O_{13}$

Exact Mass: 1106.6







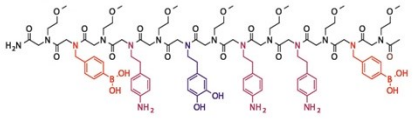
(j)

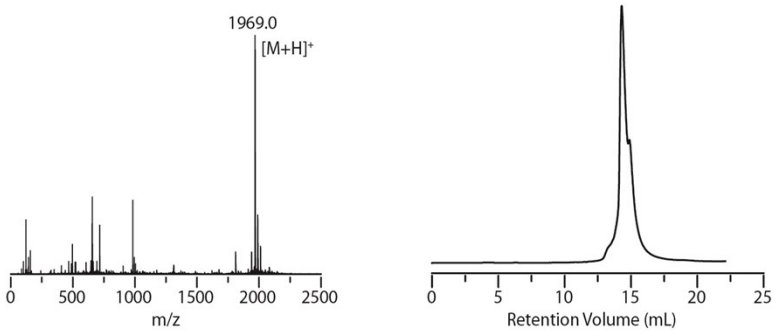
NmeNpbaNmeNamNmeNdop(NmeNam)$_2$NmeNpbaNme

213112

$C_{95}H_{135}B_2N_{17}O_{27}$

Exact Mass: 1968.0





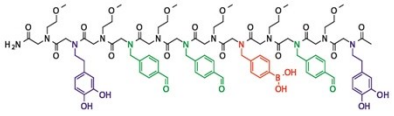
(k)

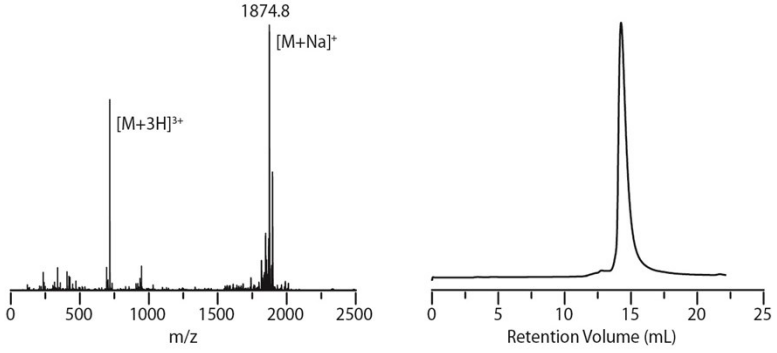
NmeNdop(NmeNal)$_2$NmeNpbaNmeNalNmeNdop

300203

$C_{96}H_{118}BN_{13}O_{28}$

Exact Mass: 1851.8





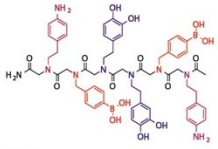
(l)

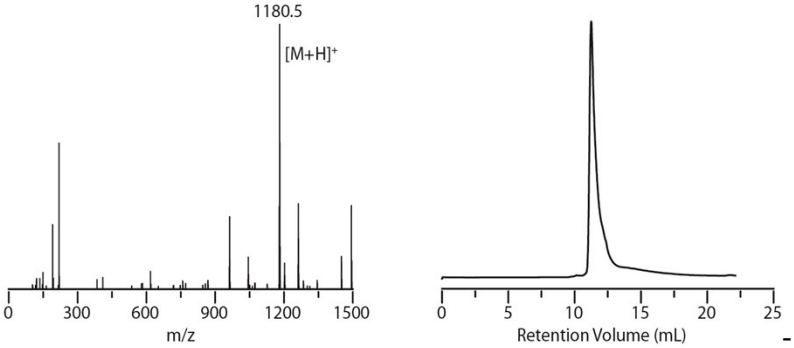
NamNpbaNdop$_2$NpbaNam

123321

$C_{60}H_{71}B_2N_9O_{15}$

Exact Mass: 1179.5





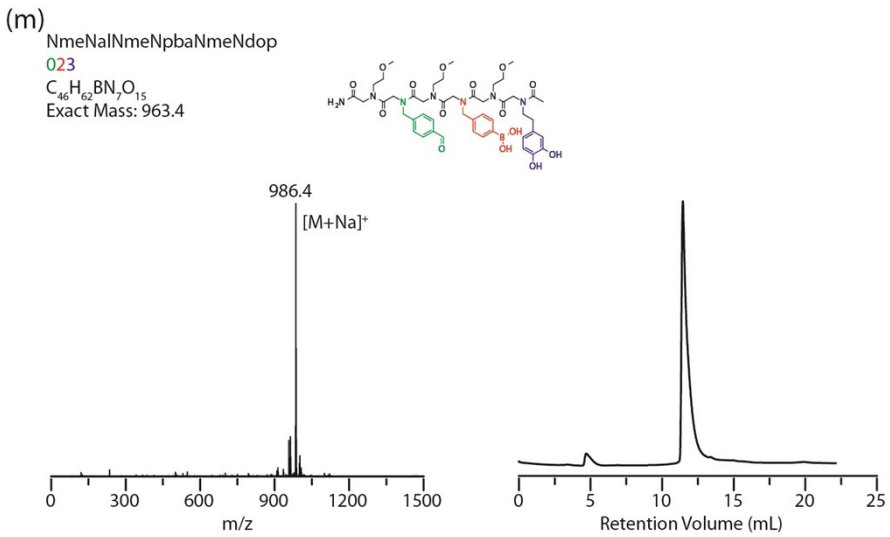
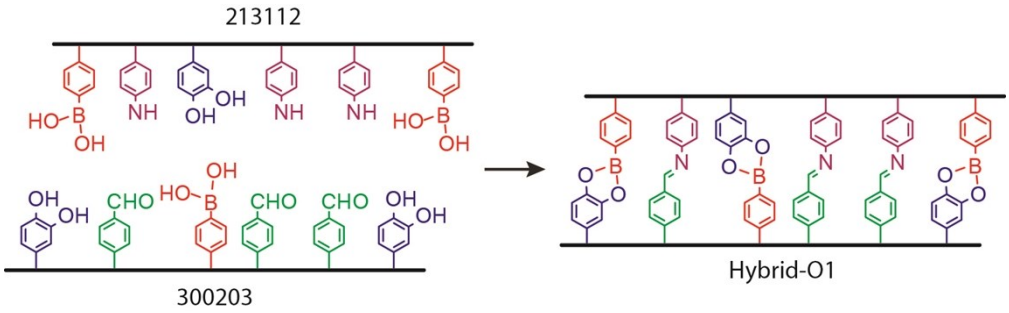


Fig. S1. ESI mass spectra and corresponding analytical HPLC traces of peptoids synthesized for this study.



Scheme S1. The assembly of a base-4 molecular ladder.

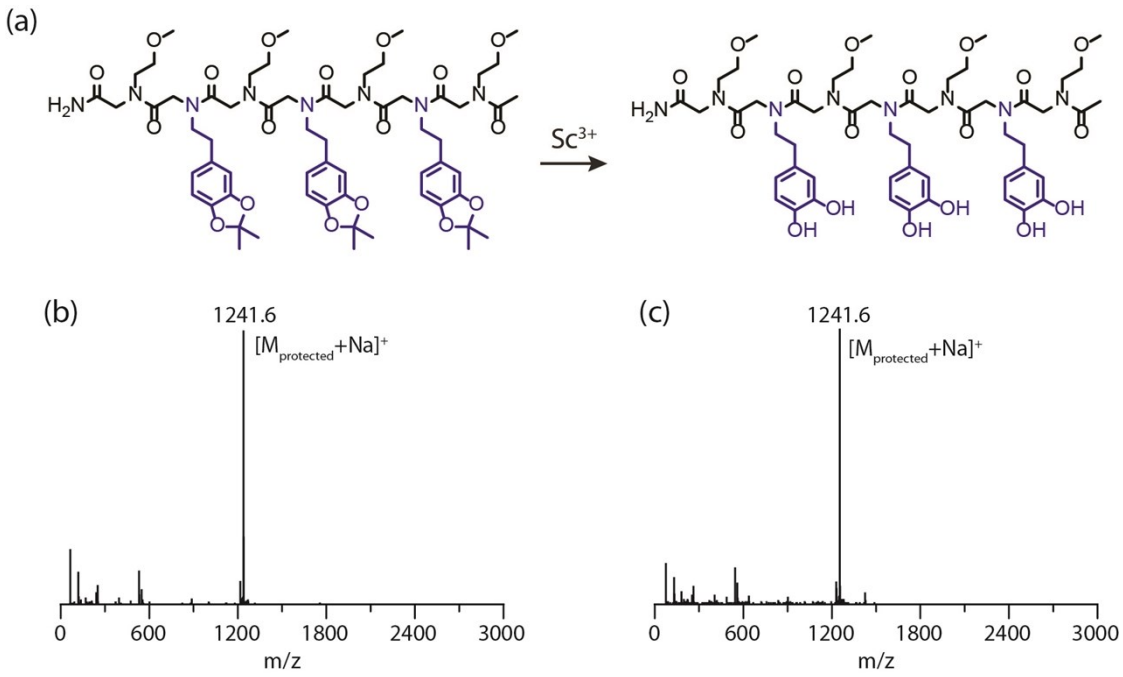


Fig. S2. Attempted deprotection of an oligomer bearing acetonide-protected diol groups. **(a)** Hypothesized reaction scheme depicting the deprotection of the acetonide-protected catechol pendant groups on a peptoid by Sc^{3+}. Up to 0.33 equivalents of scandium triflate proved unsuccessful, as shown by the mass spectra **(b)** before and **(c)** after addition of scandium triflate. Expected exact masses: $[M_{protected}+Na]^+$ = 1241.6; $[M_{deprotected}+Na]^+$ = 1121.5.

Imine molecular ladders

A vial was charged with 20 µL of 10 mM imine-bearing peptoids and 100 µL of trifluoroacetic acid. The mixture was gently stirred for 20 minutes. 400 µL of chloroform were added before adjusting the pH to 14 with 1 M NaOH. The solution was allowed to stand until complete phase separation was observed. The organic layer was removed using a pipette. Residual NaOH was removed via subsequent extractions with brine and water. The mixture was stirred overnight and characterized by positive mode MALDI-TOF mass spectrometry using hydroxyazobenzene-2-carboxylic acid (HABA) as the matrix. HABA samples were prepared by mixing 3:1 ratio of a saturated HABA solution in acetonitrile to the reaction mixture.

Boronate ester molecular ladders

10 mM stock solutions of the self-hybridizing boronic acid and catechol-bearing peptoids were prepared in a 50:50 mixture of acetonitrile and water. Initially, a vial was charged with 20 µL of each peptoid and 100 µL of trifluoroacetic acid and the mixture was gently stirred for 20 minutes. The pH of the solution was adjusted to 14 with 1 M NaOH before stirring overnight. As the presence of boronic acid- and catechol-functionalities afford water soluble peptoid species, addition of chloroform, and subsequent extraction, were unnecessary. However, upon mixing aliquots of the final reaction mixture with MALDI-TOF mass spectrometry matrix, the matrix underwent an immediate color change and no signal could be detected in the resulting mass spectrum. This effect was observed for several matrices, including hydroxyazobenzene-2-carboxylic acid, α-cyano-4-hydroxycinnamic acid, and *trans*-2-[3-(4-*tert*-butylphenyl)-2-methyl-2-propenylidene]malononitrile, and is a consequence of the relatively high concentration of NaOH and TFA in solution. As such, the self-assembly procedure was modified to omit dissociation by TFA and reduction of the NaOH amount added to adjust the reaction mixture pH. Thus, 20 µL of

the boronate ester peptoids were put in a vial of 180 µL basic aqueous solution where the pH of the solution was previously adjusted to be approximately 9 with sodium hydroxide. The vial was stirred overnight and MALDI-TOF mass spectrometry in negative ion mode was used to confirm the formation of a molecular ladder. The negative mode MALDI-TOF samples were prepared by mixing 4 µL of matrix (5 mg of α-cyano-4-hydroxycinnamic acid in 1 mL of 50:50 MeCN:water) with 2 µL of the reaction mixture.

Table S3. Nomenclature and corresponding exact mass for assembled structures used in this study.

Molecular Ladder	Exact Mass (g/mol)	Sequences
Hybrid-I1	3745.0	00111 × 11000
Hybrid-I2	3283.7	10101 × 01010
Hybrid-BE1	3817.9	222333 × 222333
Hybrid-BE2	3587.7	223233 × 223233
Hybrid-G	3200.5	303030 × 111 × 222
Hybrid-O1	3657.7	213112 × 300203
Hybrid-G1	2926.4	123321 × 023 × 023

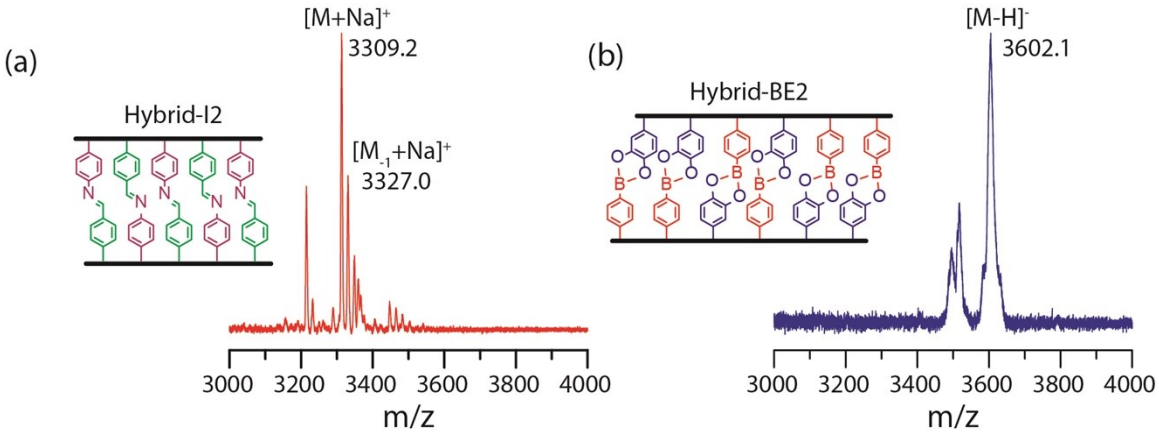


Fig. S3. Sequence-selective hybridization of molecular ladders bearing **(a)** imine and **(b)** boronate ester rungs. Expected exact masses: $[M_{Hybrid\text{-}I2}+Na]^+$ = 3306.7; $[M_{Hybrid\text{-}BE2}-H]^-$ = 3586.7. For imine-bearing ladders, peaks at multiples of +18 m/z values (e.g., M_{-1}) are attributable to ladders species with fewer rungs.

Orthogonality examination with model compounds

100 mM stock solutions of commercially purchased aniline, benzaldehyde, phenylboronic acid, and catechol were prepared in CD_3CN. Pairs of the compounds were reacted by adding equal volumes (300 µL) of each monomer to a vial and gently mixing for three hours before ^1H NMR and ^{11}B NMR spectra were collected. The ^{11}B NMR spectra were collected in a quartz NMR tube.

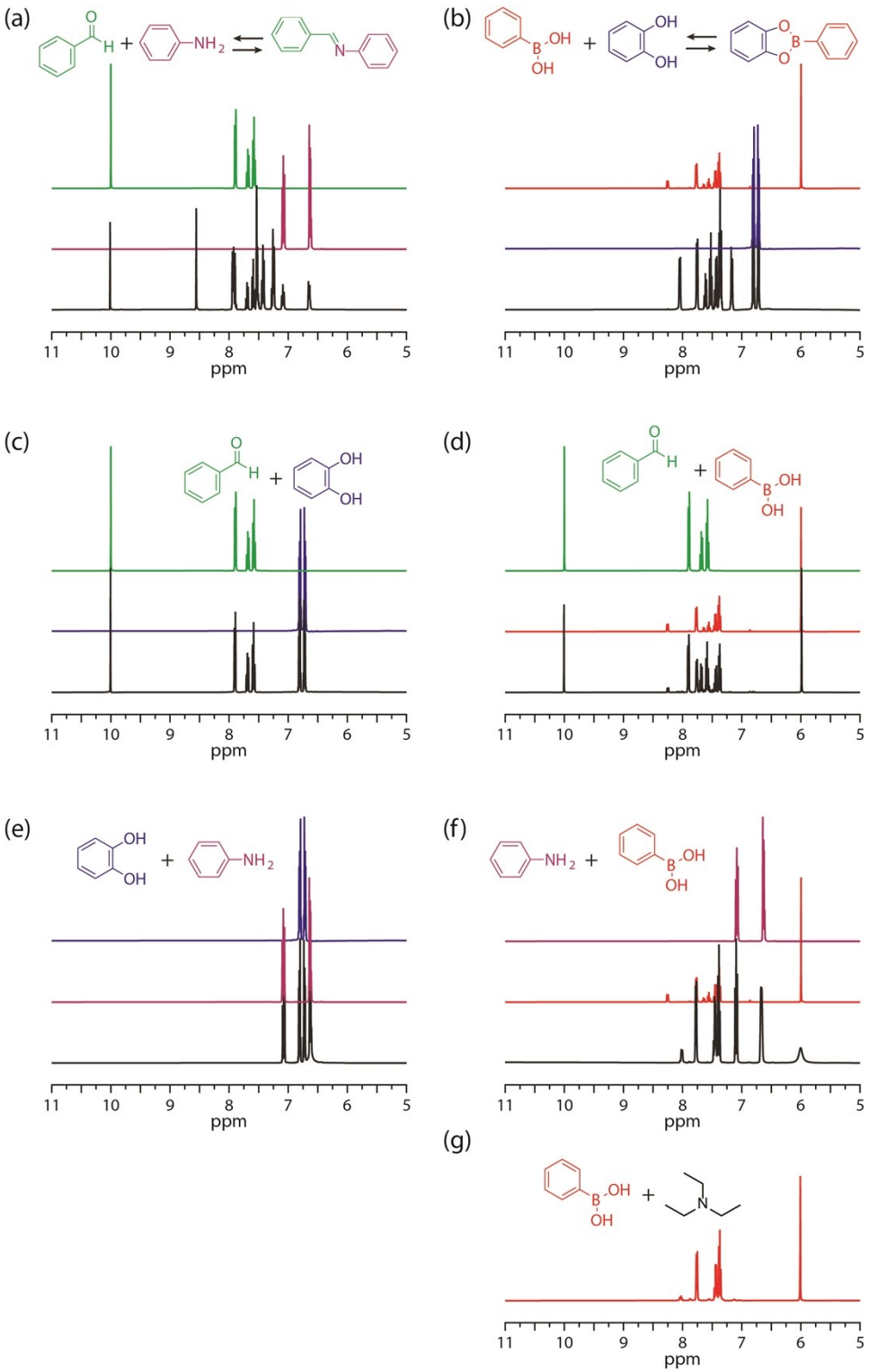


Fig. S4. (a-f) Proton nuclear magnetic resonance (H^1 NMR) spectra of cross reactions between model compounds for the dynamic covalent pendant groups. **(g)** ^{1}H NMR spectra of a phenylboronic acid and triethylamine solution demonstrating the peak shifts observed in the ^{1}H-NMR spectra of the reaction between phenylboronic acid and aniline are caused by the basicity of aniline rather than a coupling reaction.

Molecular ladder and grid assembly

Hybrid-G was formed by adding 20 µL from 10 mM stock solutions of the boronic acid bearing peptoid (222), the amine bearing peptoid (111), the peptoid with both aldehyde and catechol functional groups (303030), and lastly, 2 µL of a 10 mM solution of scandium (III) triflate into an aqueous solution that was previously adjusted to a pH of 9. Hybrid-O1 was self-assembled by adding 20 µL of 10 mM stock solutions of each of the two complementary sequences 213112 and 300203 and 2 µL of a 10 mM solution of scandium (III) triflate in in a vial of 158 µL basic aqueous solution where the pH of the solution was previously adjusted to be approximately 9 with sodium hydroxide. The vial was stirred overnight in anaerobic conditions. The negative mode MALDI-TOF samples were prepared by mixing 4 µL of matrix (5 mg of α-cyano-4-hydroxycinnamic acid in 1 mL of 50:50 MeCN:water) with 2 µL of the reaction mixture. Likewise, the base-4 molecular grid (Hybrid-G1) was prepared by adding 10 µL of the core strand (123321) with 20 µL of the side strand (023) and 2 µL of the same 10 mM solution of scandium(III) triflate into a basic aqueous water solution with a final volume 200 µL. The reaction was allowed to stir overnight in anaerobic conditions.

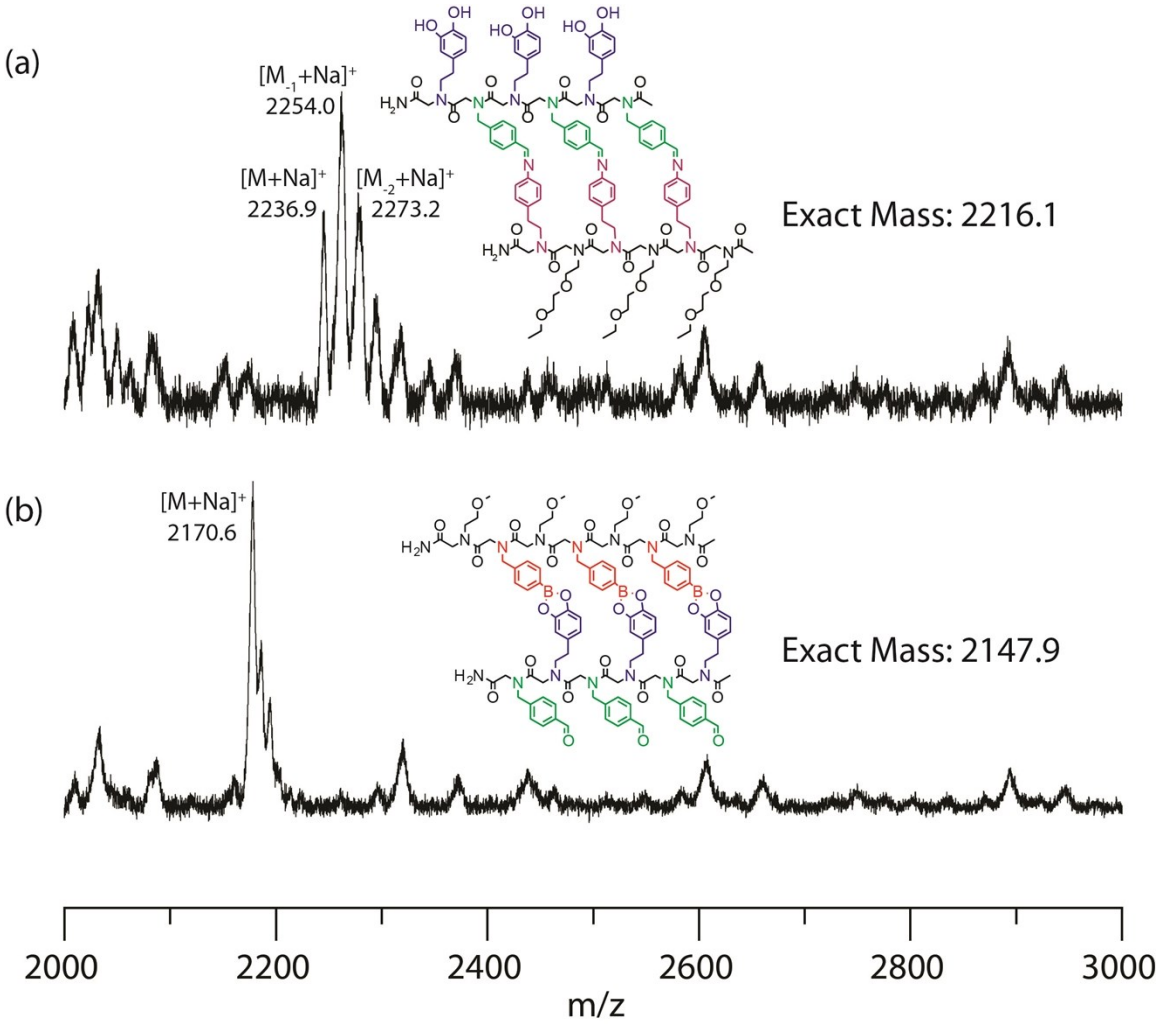


Fig. S5. Positive mode MALDI-TOF spectra of the two duplexes formed with the catechol and aldehyde core and either the **(a)** amine- or **(b)** boronic acid-bearing flanking strands.

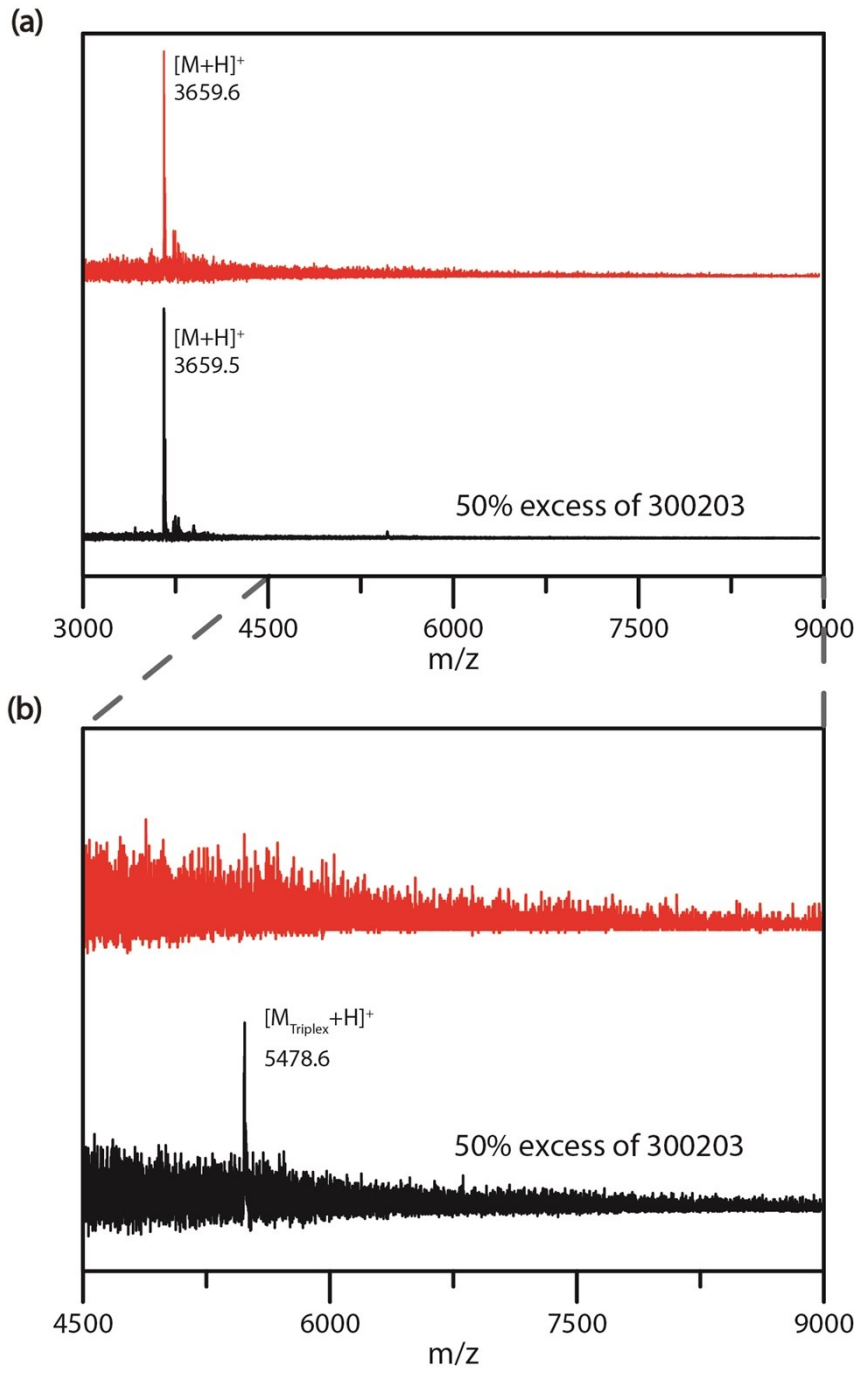


Fig. S6. (a,b) Positive mode MALDI-TOF spectra of Hybrid-O1 with a 1:1 stoichiometric ratio of sequences 213112 and 300203 (top red) and a 1:1.5 ratio. Expected exact masses: $[M_{Hybrid\text{-}O1}+H]^+$ = 3658.7; $[M_{Triplex}+H]^+$ = 5474.6. $M_{Triplex}$ refers to the mass of a multimeric specie comprised of three strands; one 213112 and two 300203 oligomers.

References

1 T. Wei, J. H. Jung and T. F. Scott, *J. Am. Chem. Soc.*, 2015, **137**, 16196–16202.

2 T. Wei, J. C. Furgal and T. F. Scott, *Chem. Commun.*, 2017, **53**, 3874–3877.

3 J. Xu, D. Duran and B. Mao, *J. Liq. Chromatogr. Relat. Technol.*, 2006, **29**, 2795–2806.