Information-Directed Hybridization of Abiotic, Sequence-Defined Oligomers

by

Samuel Leguizamon

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in The University of Michigan
2020

Doctoral Committee:

Associate Professor Timothy F. Scott, Chair
Professor Lola Eniola-Adefeso
Professor Jinsang Kim
Associate Professor Greg Thurber

Samuel Leguizamon

sleguiz@umich.edu

ORCID iD: 0000-0003-3395-6562

© Samuel Leguizamon 2020

Dedication

This dissertation is dedicated to the memory of

Julia Haines

Acknowledgements

I would like to thank my doctoral advisor Prof. Timothy F. Scott for his support and guidance throughout my studies at the University of Michigan and as a Visiting Scholar at Monash University. I would also like to thank my dissertation committee members, Professor Greg Thurber, Professor Jinsang Kim and Professor Lola Eniola-Adefeso for their advice and feedback.

I am fortunate to have worked with many talented and wonderful individuals in my tenure in the Scott Polymer Dojo (SPD). This included Dr. Scott Zavada, Dr. Joseph Furgal, Dr. Dowon Ahn, Dr. Tao Wei, Dr. Megan Dunn, Dr. Harry van der Laan, Dr. Max Ma, Abdulla Alqubati, Austin Bingham, Futianyi Wang, Alex Commisso, Dr. Gopal Sama, Davood Khoeini, Dr. Jae Hwan Jung, Jules van Dijck, Nathan Wood, and Dan Li. You'll never find a lab with as much camaraderie, and I am ever so grateful to have be a small part of it. I would like to specifically thank co-authors that directly contributed to the described work: Megan and Abdulla, as well as Tao, Megan and Joe for their immeasurable guidance in the early years.

I am also very grateful to Professor Greg Thurber and Thurber lab members for their allowing me to use instrumentation such as high performance liquid chromatography (HPLC) in their lab as well as the general exchange of techniques, methodologies, and ideas between the two labs. I would like to express gratitude to James Windak and Paul Lennon of the Chemistry Department Mass Spec Services Lab for their help with characterization along with Scott Blundell at the Monash Analytical Platform. I would also like to thank the Chemical Engineering Department faculty and staff, especially Susan Hamlin, Kelly Raickovich, Barbara Perry, Jennifer Downey, Pamela Bogdanski, and Shelley Fellers for their assistance. I would also like to thank the support from Rackham Graduate School.

Moreover, I would like to thank my friends and colleagues at Monash University; Diana Alves, Rodrigo Curvello, Edward Henderson, Vidhishri Kesarwani, Julia A. Walker, McLiesh, Vikram Singh Raghuwanshi, Hajar Samadian, Professor Mark M. Banaszak Holl, Professor Gil Garnier, and Professor Simon R. Corrie, for their support at Monash University as well as, Tracy Groves, Laura McManus, Kim Phu, Edna Tan, and Kris Wirthensohn for their assistance in the transition between the two Universities.

I would also like to acknowledge the funding that supported my graduate studies. This includes personal fellowships awarded from the Rackham Graduate School and the National Science Foundation GRFP. I would also like to thank grants specifically funding the research from U.S. Department of Energy and the National Science Foundation.

I am eternally grateful for all those that have helped me reach this point including the numerous teachers, mentors, and friends I have had since I was a child. I’d like to thank my mother and father for encouraging my scientific curiosity and creativity, and for their endless support and wisdom. They are a constant source of inspiration. To my sisters, this degree will irrefutably prove that I am the smartest child, but I am indebted to both of you for keeping me grounded and for providing a foundation that only siblings can. Most importantly, I’d like to thank my wife for her support throughout my doctoral studies. She has moved across the world with me, forgiven my absence as I labored seemingly endlessly in lab (and even provided me with lunch, dinner and baked treats on occasion), and never ceases to amaze me.

Table of Contents

Dedication .... **ii**

Acknowledgements .... **iii**

List of Tables .... **x**

List of Schemes .... **xi**

List of Figures .... **xii**

Abstract .... **xxxii**

Chapter 1 Introduction **1**

- **1.1 Overview of Research** **1**
- **1.2. Nucleic Acid Nanotechnology** **1**
- **1.3. Dynamic Covalent Interactions** **6**
  - 1.3.1. Imine Chemistry 7
  - 1.3.2. Boronate Ester Chemistry 8
  - 1.3.3. Diels-Alder Chemistry 10
- **1.4. Abiotic, Sequence-defined Polymers** **11**
  - 1.4.1. Solid-phase Synthesis 13
  - 1.4.2. Solution-phase Synthesis 14
  - 1.4.3. Living Polymerization 15
- **1.5. Nucleic Acid Mimicry** **16**
  - 1.5.1 Encoding and Decoding Abiotic Polymers 16

1.5.2. Self-Assembly of Ladder Species 28
1.5.3. Replication of Abiotic Polymers 35
1.6. Dynamic Covalent Assembly 42
1.6.1. Dynamic Covalent Micelles 42
1.6.2. Dynamic Covalent Hydrogels 46
1.6.3. Covalent Organic Frameworks 48
1.6.4. Dynamic Covalent Cages 51
1.7. Overview of Subsequent Chapters 54
1.8. References 55
Chapter 2 Sequence-selective Dynamic Covalent Assembly of Information-bearing Oligomers 65
2.1 Original Publications Information 65
2.2 Abstract 65
2.3 Introduction 66
2.4 Experimental 67
2.4.1 General Experimental Procedure 67
2.4.2 Monomer Synthesis 68
2.4.3 Oligo(peptoid) Synthesis 72
2.4.4 Peptoid Hybridization 80
2.4.5 Förster Resonance Energy Transfer (FRET) 81
2.4.5 Inductively Coupled Plasma Mass Spectrometry (ICP-MS) 81
2.5 Results and Discussion 82
2.5.1 Dynamic Covalent Assembly of Encoded Molecular Ladders 82
2.5.2 Competitive Dynamic Covalent Assembly 96
2.5.3 Sequence-selective Hybridization 100

2.6 Conclusions **111**

2.7 References **112**

Chapter 3 Sequence-directed Dynamic Covalent Assembly of Base-4-encoded Oligomers **115**

3.1 Original Publication Information **115**

3.2 Abstract **115**

3.3 Introduction **116**

3.4 Experimental **117**

3.4.1. General Experimental Procedure 117

3.4.2. Oligomer Synthesis 118

3.4.3. Peptoid Deprotection and Purification 123

3.4.4. Imine Molecular Ladders 129

3.4.5. Boronate Ester Molecular Ladders 129

3.4.6. Orthogonality Examination with Model Compounds 130

3.4.7. Molecular Ladder and Grid Assembly 130

3.5. Results and Discussion **132**

3.5.1. Base-2 Sequence Selective Assembly 132

3.5.2. Confirmation of Dynamic Covalent Interactions Orthogonality 135

3.5.3. Base-4 Grid Formation by Sequential Dynamic Covalent Interactions 138

3.5.4. Base-4 Self-Assembly 141

3.6. Conclusion **144**

3.7 References **144**

Chapter 4 Temperature-mediated Molecular Ladder Self-assembly Employing Diels-Alder Cycloaddition **146**

4.1 Abstract **146**

4.2 Introduction 147

4.3 Experimental 148

4.3.1. General Experimental Procedure 148

4.3.2. Monomer Synthesis 149

4.3.3. Oligomer Synthesis 153

4.3.4. Oligomer Cleavage and Purification 154

4.3.5. Oligomer Deprotection 156

4.3.6. Molecular Ladder Formation 157

4.3.7. Molecular Ladder Degree of Alignment Analysis 157

4.4 Results and Discussion 158

4.4.1. Monomer Design 158

4.4.2. Maleimide Deprotection 160

4.4.3. Initial Molecular Ladder Formation 161

4.4.4. Gel Permeation Chromatography 164

4.4.5. Degree of Alignment 167

4.4.6. Sequence-selective Ladder Assembly 171

4.5 Conclusion 175

4.6 References 175

Chapter 5 Rapid Gel-card Agglutination Assays for Serological Analysis Following SARS-CoV-2 Infection in Humans 178

5.1 Original Publication Information 178

5.2 Abstract 178

5.3 Introduction 179

5.4 Experimental 182

5.4.1. General Experimental Procedure 182

5.4.2. Peptide Synthesis 182

5.4.3. Anti-D-IgG Purification and Bioconjugation Reaction 185

5.4.4. Bioconjugate Reaction 185

5.4.5. Flow Cytometry 186

5.4.6. Column Agglutination Tests 186

5.4.7. Indirect IgG ELISA 187

5.5 Results and Discussion 188

5.5.1. Peptide Selection and Bioconjugation 188

5.5.2. Red Blood Cell Agglutination 193

5.5.3. Testing with Clinical Samples 196

5.6 Conclusion 198

5.6 References 200

Chapter 6 Summary and Future Directions 203

6.1 Summary of Research 203

6.2 Dynamic Covalent Assembly Future Directions 205

6.2.1. 2-D Grid Structures 206

6.2.2. Non-linear Branched Structures 209

6.2.3. Dynamic Covalent Scaffolding 210

6.3 Peptide-antibody Bioconjugates Future Directions 211

6.4 Importance 212

6.5 References 212

List of Tables

Table 2.1 Hybridization specificity of message strands with their library complements assessed by normalized MALDI mass spectrum intensities. 111

Table 3.1 Nomenclature of peptoids used in this chapter, the associated sequence and exact mass. 124

Table 3.2 Nomenclature and corresponding exact mass for assembled structures used in this study. 131

Table 5.1 List of peptides synthesized, Sars-Cov-2 protein sequence, and the start and end amino acid position on the protein sequence. 189

Table 5.2 List of peptides synthesized with modifiers. 190

Table 5.3 List of clinical samples used in the study 196

List of Schemes

Scheme 3.1 pH-mediated dynamic covalent condensation reactions. The addition of amines (0) to aldehydes (1) yields imines, while the reaction of boronic acids (2) with diols (3) affords boronate esters. 117

Scheme 3.2 The assembly of a base-4 molecular ladder. 117

Scheme 3.3 Reaction pathway for the formation of Nace. 122

Scheme 3.4 Alternate pathways for the assembly of a base-4 molecular grid *via* the sequential addition of oligomeric sequences bearing boronic acid or amine pendant groups. 139

Scheme 6.1 Macrocycle scaffolding to assemble shapes dependent on scaffold length. 211

List of Figures

Figure 1.1 A branched DNA molecule with four arms that is able to form multi-junction structures through the use of overhangs or 'sticky ends' [14,15]. 2

Figure 1.2 The generation of a cube entirely from DNA strands by Junghuei Chen and Nadrian Seeman[17]. 3

Figure 1.3 The principle design behind DNA origami[26] and resultant structures[20]. 4

Figure 1.4 DNA bricks used to form three-dimensional letter structures[28]. 5

Figure 1.5 Examples of dynamic covalent interactions[32]. 6

Figure 1.6 Literature values comparing the dissociation rate coefficients for a variety of reversible reactions (k_d(obs))[33].. 7

Figure 1.7 (a) Imine hydrolysis, (b) transimination, and (c) imine metathesis reactions[36]. .. 8

Figure 1.8 Key interactions of boron, with respect to self-assembly[44]. 9

Figure 1.9 Parent [4+2] Cycloaddition of unsubstituted furan with maleimide to provide the kinetic endo and thermodynamic exo diels-alder adducts[49]. 10

Figure 1.10 Diels-Alder and retro Diels-Alder reaction between furan derivatives and maleimide for the creation of new materials[49]. 11

Figure 1.11 Furan substitution patterns that influence reversibility[50]. 11

Figure 1.12 Furan substitution patterns that influence reversibility[70]. 12

Figure 1.13 Peptoid synthesis via Zuckerman's submonomer aporach to solid-phase synthesis[74,75]. 13

Figure 1.14 Schematic representation of solution-phase synthesis of polymers by linear, bidirectional, and iterative exponential growth mechanisms[82]... 15

Figure 1.15 Schematic representation of sequence-specific polymers by regioselective ROMP[83].. 15

Figure 1.16 Reported synthetic sequence-defined macromolecules employed for data storage and retrieval. 20

Figure 1.17 Sequencing of a 4-byte digital polymer that contains the ASCII-encoded word Byte. (A) High-resolution electrospray mass spectrum (MS^1) obtained in the negative ion mode for a 4-byte digital polymer. (b) MS^2 spectrum (0.56 eV, center of mass frame) obtained by collision-induced dissociation of the $[M-12H]^{12-}$ precursor ion. (c) Molecular sequencing (pseudo-MS^3) of a byte fragment obtained by collision-induced dissociation (0.56 eV, center of mass frame) of the precursor trianion $[M-3H]^{3-}$. For clarity, only the sequencing of byte 4 is shown as an example in this figure.[108].. 21

Figure 1.18 Abiotic macromolecule espionage. In method 1, a four-component molecular key is deposited on an envelope delivering an encoded message[121]. The molecular key is isolated and characterized, revealing the decryption rules for the encoded message. Method

2 relies on encoded the message in sequence-defined macromolecules. Decryption rules can be elucidated from individual strands by ion mobility spectrometry enabling the decoding of the binary data gathered from MS/MS sequencing[122]. 26

Figure 1.19 Photo-editable poly(phosphodiester)s where light triggered sequence mutations enabled the photo-erasing or photo-revealing of encoded-information[123]. 28

Figure 1.20 Noncovalent linkage mechanisms used to form abiotic molecular ladder; (a) hydrogen bonding, (b) metal coordination, (c) π-stacking, (d) conjugated radical cations, and (e) multivalent molecular recognition. 29

Figure 1.21 Reported approaches to forming covalently-linked molecular ladders bearing (a) imine or (b) boronate ester linkages. 33

Figure 1.22 Schematic representation of an imine-bearing molecular ladder formed by Vernier templating[146]. 35

Figure 1.23 Autocatalyzed self-replication (a) mechanism schematic and (b) application employing a salt bridge medium[159,160]. 36

Figure 1.24 (a and b) Cryo-TEM images, (c) circular dichroism spectra, and (d) schematic of the macrocycles formed from Canrall's work[162]. 37

Figure 1.25 Synthetic, sequence-defined polymer replication using a DNA template[163]. 38

Figure 1.26 Abiotic polymer replications using parent/daughter templating techniques in which the parent strands can be reused. Synthetic pathways employing (a) salt bridge

linkages and template-enhanced ring-opening metathesis polymerization[164], (b) imine chemistry and Sonogashira coupling[172], and (c) ester formation with CuACC[173]................ 40

Figure 1.27 Synthesis of nanorings by way of Vernier templating[171].................................. 42

Figure 1.28 Controlled formation of micellar aggregates from the aromatic aldehyde polar head group reversibly reacting with various amines to form an amphiphilic surfactant that can self-assemble into micelles.. 43

Figure 1.29 Schematic showing the formation of imine bonds from difunctionalized aldehydes and aliphatic amines to form macrocycles and oligomers leading to the vesicles, vesicle clusters, and finally vesicle networks capable of forming pH-reversible hydrogels. 44

Figure 1.30 The formation of wormlike micelles from Gemini surfactants comprised of a bisaldehyde head group and amine tail groups[179].. 45

Figure 1.31 The formation of PTX loaded micelles through disulfide bond formation[180]... 46

Figure 1.32 Dyanmic covalent nanogel a) schematic showing the formation of a nanogel from functionalized nanoparticles b) structures of the nanogel including the disulfide chemistry[181] 46

Figure 1.33 Dynamic covalent hydrogel a) Schematic showing the formation of a pattern into the photoadaptable hydrogel b) The patterned surface scale bar = 1.5mm c) The resulting patterned hydrogel scale bar = 1.5mm[182]. 48

Figure 1.34 Synthetic schemes for the functionalization of covalent organic frameworks; (a) COF-102 is modified by (b) thiol-ene or (c) boronic acid, chemistries. (d) Modification of 3D-OH-COF[185] 50

Figure 1.35 A timeline of porous organic cages[195]. 52

Figure 1.36 Synthetic pathways for the formation of porous organic cages by (a)imine, (b) boronate ester, and (c) alkyne metathesis chemistries[195]. 53

Figure 1.37 Formation of a molecular cage bearing imine- and boronate ester-linkages[42].53

Figure 2.1 Primary amine monomers used in this study for peptoid synthesis. 68

Figure 2.2 Monomer synthetic schemes and ^{1}H-NMR spectra. (A) Monomer synthetic schemes with reagents and conditions: (i) allyl chloroformate, 10% aqueous acetic acid, 1,4-dioxane, room temperature, overnight; (ii) ethylene glycol, toluene-p-sulfonic acid, toluene, reflux, overnight; (iii) $LiAlH_4$, anhydrous Et_2O, 0 °C for 4 h then room temperature for 12 h; (iv) tosyl chloride, THF, 0 °C; (v) NaN3, DMF, 60 °C, 36 h; (vi) triphenylphosphine, THF, overnight. (B) Monomer ^{1}H-NMR spectra (500 MHz, $CDCl_3$): (i) 4-(2-aminoethyl)-N-(allylcarbonyloxy)phenylamine (Npam); (ii) 4-(1,3-dioxacyclopent-2-yl)benzylamine (Npal); (iii) 2-(2-ethoxyethoxy)ethylamine (Neee). 72

Figure 2.3 ESI mass spectra of purified peptoid sequences; a, $Nme_2(NeeeNam)_5NeeNme$; b, $(NeeeNpal)_5Nee$; c, $Nme(NeeeNpal)_2(NeeeNam)_3Nme$; d, $(NeeeNam)_2(NeeeNpal)_3$; e, $(NamNeeeNpalNee)_2NamNme$; f, $(NpalNeeeNamNee)_2Npal$; g, $NmeNam(NeeeNpal)_3NeeeNamNme$; h, $(NpalNeeeNamNee)_2NpalFAM$; i, $(NamNeeeNpalNee)_2NamNmeNdab$; j, $NmeNam(NeeeNpal)_3NeeeNamNmeNdab$. 75

Figure 2.4 a-j, RP-HPLC traces of purified peptoid sequences. a, $Nme_2(NeeeNam)_5NeeNme$; b, $(NeeeNpal)_5Nee$; c, $Nme(NeeeNpal)_2(NeeeNam)_3Nme$; d, $(NeeeNam)_2(NeeeNpal)_3$; e, $(NamNeeeNpalNee)_2NamNme$; f, $(NpalNeeeNamNee)_2Npal$; g, $NmeNam(NeeeNpal)_3NeeeNamNme$; h, $(NpalNeeeNamNee)_2NpalFAM$; i, $(NamNeeeNpalNee)_2NamNmeNdab$; j, $NmeNam(NeeeNpal)_3NeeeNamNmeNdab$. 77

Figure 2.5 ^{1}H NMR spectrum (500 MHz, $CDCl_3$) of purified peptoid sequences. 79

Figure 2.6 Dynamic covalent assembly of encoded molecular ladders. a, Structures of complementary, mass-labeled precursor peptoids and desired, in-registry molecular ladder product. b, MALDI mass spectra of molecular ladder reaction mixtures generated either after a single-step, deprotection and direct assembly process (bottom, black), or *via* a stepwise, dissociation/extraction/annealing assembly process (top, red). Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs (e.g., *y* = in-registry, 5-rung 10101×01010 molecular ladder, *y*-1 = out-of-registry, 4-rung 10101×01010 molecular ladder, etc.). c, MALDI mass spectra of the reaction mixture initially containing 10101 and 01010 sequences and a high loading of $Sc(OTf)_3$ (bottom, black), immediately after aqueous extraction (middle, red), and after annealing at 70°C for 2 hours (top, blue). Expected exact masses: $[M_{01010}+Na]^+ = 1651.83$; $[M_{10101}+Na]^+ = 1767.92$; $[M_{01010\times01010}+Na]^+$ (x) = 3208.63; $[M_{10101\times01010}+Na]^+$ (y) = 3306.71; $[M_{10101\times10101}+Na]^+$ (z) = 3440.81. 83

Figure 2.7 MALDI mass spectra of a 10101×01010 molecular ladder reaction mixture annealed at room temperature over a seven-day period. Expected exact masses: $[M_{01010\times01010}+Na]^+$ (x) = 3208.63; $[M_{10101\times01010}+Na]^+$ (y) = 3306.71; $[M_{10101\times10101}+Na]^+$ (z) = 3440.81. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer

rungs (e.g., *y* = in-registry, 5-rung 10101×01010 molecular ladder, *y*-1 = out-of-registry, 4-rung 10101×01010 molecular ladder, etc.).. 89

Figure 2.8 a, Superimposed GPC traces of an Alloc-protected, inert single strand mixture, a 10101×01010 molecular ladder mixture after performing the dissociation/extraction/annealing method, and a post-purification 10101×01010 molecular ladder. Each trace was deconvoluted (b, c, and d) by fitting Gaussian functions to simulated peaks with OriginPro. Upon subtracting the baseline, each peak over 0.5% of the max intensity was utilized to fit the spectra. e, MALDI mass spectra of the different fractions (i, ii, and iii) collected from a 10101×01010 molecular ladder mixture after GPC purification. The peaks in the molecular weight region between 4750 and 5250 correspond to 01010×01010×01010, 01010×01010×10101, and 01010×10101×10101 triple ladder species. MALDI mass spectra expanded to highlight the m/z range expected for single-stranded oligomeric species of f, the crude 01010×10101 reaction mixture, and g, GPC fraction iii. Expected exact masses: $[M_{01010}+Na]^+$ (*y*) = 1651.83; $[M_{10101}+Na]^+$ (*z*) = 1767.92. Peaks at multiples of -18 m/z values are attributable to single-stranded species incorporating intramolecular imine bonds (e.g., *y* = 01010, *y*-1 = 01010 with one intramolecular imine bond, *y*-2 = 01010 with two intramolecular imine bonds, etc.)... 90

Figure 2.9 DOSY NMR spectra of a 10101×01010 molecular ladder mixture upon dissociation with 1.5 eq of scandium triflate (green) and subsequent binding upon extracting the scandium triflate and annealing at 70°C (purple)... 92

Figure 2.10 MALDI mass spectra of 10101×01010 molecular ladders treated with additional $Sc(OTf)_3$ post-annealing and allowed to equilibrate. Expected exact masses:

[$M_{01010×01010}$+Na]$^+$ (*x*) = 3208.63; [$M_{10101×01010}$+Na]$^+$ (*y*) = 3306.71; [$M_{10101×10101}$+Na]$^+$ (*z*) = 3440.81. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs (e.g., *y* = in-registry, 5-rung 10101×01010 molecular ladder, *y*-1 = out-of-registry, 4-rung 10101×01010 molecular ladder, etc.)... 94

Figure 2.11 MALDI mass spectra of 10101×01010 molecular ladders *via* the dissociation/extraction/ annealing process challenged against a library of rare-earth metal triflates. Expected exact masses: [$M_{01010×01010}$+Na]$^+$ (*x*) = 3208.63; [$M_{10101×01010}$+Na]$^+$ (*y*) = 3306.71. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs (e.g., *y* = in-registry, 5-rung 10101×01010 molecular ladder, *y*-1 = out-of-registry, 4-rung 10101×01010 molecular ladder, etc.)... 96

Figure 2.12 Competitive dynamic covalent assembly. a, MALDI mass spectra of reaction mixtures generated after applying the dissociation/extraction/annealing process to non-complementary 01010 and 10001 sequences (bottom, red) and to a ternary mixture of a competitive environment consisting of a complementary sequence and a non-complementary sequence *via* the dissociation/extraction/annealing process (top, blue). Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs. b, Sequence-selective hybridization assessed by FRET and normalized to the fluorescence intensity of a 01010-FAM sequence single strand. Assembly solutions were excited at 495 nm and fluorescence emission was measured at 425 nm. Error bars represent standard error. The horizontal solid and dashed lines 1 and 0.2, respectively, represent the unquenched 01010-FAM control and the reported FAM/DABCYL fluorescence quenching efficiency of 80%[40], respectively..... 97

Figure 2.13 Sequence-selective hybridization. a, Structures of mass-labeled, ethylene acetal-protected precursor peptoid sequences, deprotected, dissociated sequences upon treatment with $Sc(OTf)_3$, and in-registry molecular ladders upon [Sc^{3+}] extraction and annealing. b, MALDI mass spectra of individual encoded molecular ladders assembled *via* the dissociation/extraction/annealing process, including 10101×01010 (bottom, black), 00111×11000 (second from bottom, red), 11111×00000 (middle, red), and a single-pot solution of all six oligomers to yield three in-registry molecular ladders (second from top, green). A single-pot solution of the six oligomers after the single-step, deprotection and direct assembly process (top, black) is shown for comparison. Expected exact masses: $[M_{10101\times01010}+Na]^+$ = 3306.71; $[M_{00111\times11000}+Na]^+$ = 3767.93; $[M_{11111\times00000}+Na]^+$ = 4229.25. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs. 101

Figure 2.14 MALDI mass spectra of molecular ladder reaction mixtures utilizing either the previously-determined, single-pot, single-step approach (bottom, black), or *via* the dissociation/extraction/ annealing process (top, red) of a, 11111×00000, and b, 00111×11000. 102

Figure 2.15 MALDI mass spectra, with counts, of molecular ladder reaction mixtures utilizing the previously-determined, single-pot, single-step approach of a, 10101×01010, b, 00111×11000, c, 11111×00000, and d, all six single strands simultaneously....................... 104

Figure 2.16 MALDI mass spectrum of a single-pot solution of six oligomers, including 10101, 01010, 00111, 11000, 11111, and 00000, assembled *via* the dissociation/extraction/annealing process. a, m/z 1000-10000 (full range), b, m/z 1250-4750 (expected range for dimeric and

single-stranded species), and c, m/z 1550-2100 (expected range for single-stranded species). Expected exact masses: $[M_{01010}+Na]^+$ (*v*) = 1651.83; $[M_{10101}+Na]^+$ (*w*) = 1767.92; $[M_{11000}+Na]^+$ (*x*) = 1824.93; $[M_{00000}+Na]^+$ (*y*) = 1995.97; $[M_{00111}+Na]^+$ (*z*) = 2056.09. Peaks at multiples of -18 m/z values are attributable to single-stranded species incorporating intramolecular imine bonds (e.g., *v* = 01010, *v*-1 = 01010 with one intramolecular imine bond, *v*-2 = 01010 with two intramolecular imine bonds, etc.). 106

Figure 2.17 Dynamic covalent information storage and retrieval. a, Schematic of a known library composed of three unique, mass-labeled oligomeric sequences A (01010), B (11000), and C (11111), and individual messages sequences used to challenge the library. b, MALDI mass spectrum of the known sequence library treated to the dissociation/extraction/annealing assembly process in the absence of a message strand (black, bottom), resulting in non-specific binding among the library strands, yielding all five possible sets of dimeric ladder species with varying numbers of rungs. Introduction of a message sequence to the known library and treatment of the mixture to the multi-step assembly process yielded MALDI mass spectra with dominant peaks attributable to the in-registry hybridization product of the message strand with its library complement (10101 message, second from bottom, red; 00111 message, second from top, blue; 00000 message, top, green). c, Hybridization specificity of message strands with their library complements assessed by normalized MALDI mass spectrum intensities. Possible dimeric ladder species are shown; red boxes indicate expected hybrids between message and complementary library strands. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs. 108

Figure 2.18 Dynamic covalent information storage and retrieval with varying equivalents of the message strand. a, Schematic of a known library composed of three unique, mass-labeled oligomeric sequences A, B, and C, and individual message sequence used to challenge the library, D. b, MALDI mass spectrum of the known sequence library treated to increasing equivalents of the message strand. 111

Figure 3.1 Primary amine monomers used throughout this chapter. 121

Figure 3.2 ^{1}H NMR spectra for the Nace and each intermediate product. 122

Figure 3.3 ESI mass spectra and corresponding analytical HPLC traces of peptoids synthesized for this study. 129

Figure 3.4 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........ 132

Figure 3.5 Dynamic covalent assembly of information-bearing molecular ladders. Idealized schematics and MALDI-TOF mass spectra of molecular ladders bearing (a) imine and (b) boronate ester rungs. Expected exact masses: $[M_{Hybrid\text{-}I1}+Na]^+$ = 3768.0; $[M_{Hybrid\text{-}BE1}-H]^-$ = 3816.9. For imine-bearing ladders, peaks at multiples of +18 m/z values (e.g., M_{-1}) are attributable to ladders species with progressively fewer rungs. 134

Figure 3.6 Sequence-selective hybridization of molecular ladders bearing (a) imine and (b) boronate ester rungs. Expected exact masses: $[M_{Hybrid-I2}+Na]^+$ = 3306.7; $[M_{Hybrid-BE2}-H]^-$ = 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. 134

Figure 3.7 Orthogonality between imine and boronate ester systems. (a) ^{11}B-NMR spectra highlighting the specificity of boronic acids to diols. (b) Structure and MALDI-TOF spectra of a 3×3 molecular grid assembled from the concurrent addition of a core peptoid strand with aldehyde and catechol functionality and complementary, exterior sequences bearing amine- and boronic acid-based pendant groups. Expected exact mass: $[M+Cl]^-$ = 3235.5. Digital image capturing the alternative pathways for a 3×3 molecular grid *via* (c) the sequential addition of exterior strands and (d) the final 303030×111×222 molecular grid reaction solution by way of a 303030×111 intermediate. 136

Figure 3.8 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. 137

Figure 3.9 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. 140

Figure 3.10 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. 141

Figure 3.11 (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: $[M_{Hybrid-O1}+H]^+ = 3658.7$; $[M_{Triplex}+H]^+ = 5474.6$. $M_{Triplex}$ refers to the mass of a multimeric specie comprised of three strands; one 213112 and two 300203 oligomers. 143

Figure 4.1 ^{1}H NMR (400 MHz, $CDCl_3$) of compound 1. 150

Figure 4.2 ^{1}H NMR (400 MHz, $CDCl_3$) of compound 2. 150

Figure 4.3 ^{1}H NMR (400 MHz, D_2O) of compound 3. 152

Figure 4.4 ESI mass spectra, preparative HPLC traces, and corresponding analytical HPLC traces of peptoids synthesized for this study. 156

Figure 4.5 Dynamic covalent assembly of Diels-Alder-based molecular ladders. (a) Synthetic scheme for the furan-protected, maleimide-bearing peptoid monomer, 3 and (b) chemical structures depicting the deprotection of the Diels-Alder adduct-bearing peptoid, 3PM, to afford the maleimide-bearing oligomer 3M. (c) ESI mass spectra generated after exposing solutions of 3PM strands to 60°C, 80°C, 100°C, or 120°C for one hour. Here, 3M is the fully deprotected oligomer, while 2M1PM and 1M2PM represent partially deprotected oligomers bearing one and two furan-protected maleimide groups, respectively. (d) HPLC traces of a solution of 3PM single strands prior to (black) and after deprotection at 140°C for 30 minutes (red). (e) Idealized schematic diagram of the hybridization of complementary, trimeric

maleimide- and furan-bearing oligomers (3M and 3F, respectively) to afford a peptoid-based molecular ladder bearing three Diels-Alder adduct rungs (3M×3F). (f) ESI mass spectrum of a 3PM and 3F hybridization reaction mixture after heating at 140°C for 30 minutes, then at 60°C for five days. Expected exact masses: $[M_{3F}+TEA+H]^+ = 802.4$; $[M_{3M}+TEA+H]^+ = 931.4$; $[M_{3M\times 3F}+TEA+H]^+ = 1631.7$. Exact masses found: $[M_{3F}+TEA+H]^+ = 802.4$; $[M_{3M}+TEA+H]^+ = 931.5$; $[M_{3M\times 3F}+TEA+H]^+ = 1631.8$. (g) ^{1}H-NMR spectra of the maleimide-bearing peptoid 3M (bottom, orange), the furan-bearing peptoid 3F (middle, purple), and a 3M×3F molecular ladder reaction mixture (top, green). 158

Figure 4.6 ^{1}H NMR (400 MHz, CD_3CN) of a trimeric peptoid bearing maleimide species before (bottom, blue) and after deprotection (top, red). 160

Figure 4.7 Mass spectra of a 3M×3F hybridization solution after five days at 60°C and absent any catalyst. 161

Figure 4.8 DOSY NMR spectra of 3PM (bottom, blue), 3F (middle, purple) and a 3M×3F molecular ladder mixture (top, green) after reaction for two days at 60°C and absent any catalyst. As hybridization reactions in deuterated anisole generated significant amounts of precipitate, CD_3CN was used as the solvent for DOSY NMR studies................................. 163

Figure 4.9 GPC traces of sequence-defined oligomers using a UV-Vis detector at 313 nm. 164

Figure 4.10 Time-resolved characterization of molecular ladder formation. (a) Idealized schematic diagram of the self-assembly of tetramaleimide (4M) and tetrafuran (4F) precursor strands to afford a molecular ladder bearing four Diels-Alder adduct-based rungs

(4M×4F). GPC traces, normalized to a polystyrene internal standard, of aliquots periodically sampled from 4M×4F hybridization reaction mixtures reacted at (b) 60°C and (c) 100°C, where *i*, *ii*, and *iii* indicate single stranded, dimeric, and multimeric species, respectively, bearing residual maleimide groups. Normalized concentrations of single strand species in hybridization mixtures with and without catalysts and reacted at (d) 60°C and (e) 100°C, determined from deconvolution of overlapping GPC trace peaks by Gaussian curve fitting. 165

Figure 4.11 Normalized concentrations of single strand species in hybridization mixtures with and without added catalysts at room temperature found by deconvoluting GPC traces with fitted Gaussian functions and normalized to an internal polystyrene standard. 166

Figure 4.12 Molecular ladder registry. (a) Reaction scheme for the base catalyzed thiol-Michael addition between a maleimide and methyl 3-mercaptopropionate (S). (b) ESI mass spectra of thiol-Michael addition reaction mixtures between the peptoid 3M, bearing three maleimide pendant groups, and reacted for 30 minutes with various equivalents of S to maleimide groups. (c) ^{1}H-NMR spectra of a solution of 3M strands before (bottom, red) and after (top, blue) reaction with 3 equivalents of S to maleimide groups. ESI mass spectra of 3M×3F hybridization mixtures (d) after reaction at 60°C for various periods as shown, and (e) after reaction at 60°C for various periods as shown, then subjected to a thiol-Michael addition with a threefold excess of S. Expected exact masses: $[M_{3F}+TEA+H]^+$ = 802.4; $[M_{3M}+TEA+H]^+$ = 931.4; $[M_{3M\times 3F}+TEA+H]^+$ = 1631.7; $[M_{3M+S}+TEA+H]^+$ = 1051.4; $[M_{3M+2\bullet S}+TEA+H]^+$ = 1171.3; $[M_{3M+3\bullet S}+TEA+H]^+$ = 1291.3; $[M_{3M\times 3F+S}+TEA+H]^+$ = 1751.7; $[M_{3M\times 3F+2\bullet S}+TEA+H]^+$ = 1871.7;. Exact masses found: $[M_{3F}+TEA+H]^+$ = 802.4; $[M_{3M}+TEA+H]^+$ = 931.4; $[M_{3M\times 3F}+TEA+H]^+$ = 1631.7; $[M_{3M+S}+TEA+H]^+$ = 1051.5;

$[M_{3M+2\bullet S}+TEA+H]^+ = 1171.5$; $[M_{3M+3\bullet S}+TEA+H]^+ = 1291.5$; $[M_{3M\times 3F+S}+TEA+H]^+ = 1751.8$; $[M_{3M\times 3F+2\bullet S}+TEA+H]^+ = 1871.8$. 167

Figure 4.13 HPLC traces generated from thiol-Michael addition reaction mixtures between the peptoid, 3M, and various equivalents of S to maleimide groups. 169

Figure 4.14 ESI mass spectra of 3M×3F hybridization mixtures (a) after reaction at 100°C for 24 and 48 hours, and (b) after reaction at 100°C for 24 and 48 hours, then subjected to a thiol-Michael addition with a threefold excess of S. Expected exact masses: $[M_{3F}+TEA+H]^+ = 802.4$; $[M_{3M}+TEA+H]^+ = 931.4$; $[M_{3M\times 3F}+TEA+H]^+ = 1631.7$; $[M_{3M+S}+TEA+H]^+ = 1051.4$; $[M_{3M+2\bullet S}+TEA+H]^+ = 1171.3$; $[M_{3M+3\bullet S}+TEA+H]^+ = 1291.3$; $[M_{3M\times 3F+S}+TEA+H]^+ = 1751.7$; $[M_{3M\times 3F+2\bullet S}+TEA+H]^+ = 1871.7$;. Exact masses found: $[M_{3F}+TEA+H]^+ = 802.4$; $[M_{3M}+TEA+H]^+ = 931.4$; $[M_{3M\times 3F}+TEA+H]^+ = 1631.7$; $[M_{3M+S}+TEA+H]^+ = 1051.5$; $[M_{3M+2\bullet S}+TEA+H]^+ = 1171.5$; $[M_{3M+3\bullet S}+TEA+H]^+ = 1291.5$; $[M_{3M\times 3F+S}+TEA+H]^+ = 1751.8$; $[M_{3M\times 3F+2\bullet S}+TEA+H]^+ = 1871.8$. 170

Figure 4.15 Dynamic covalent assembly of sequence-defined molecular ladders (a) Idealized schematic diagram of the concurrent self-assembly of 4PM, 4F, and 2PM2F precursor strands to afford the four-rung molecular ladders 4M×4F and 2M2F×2M2F. ESI mass spectra of molecular ladder reaction mixture aliquots removed from hybridization solutions heated at 140°C for 30 minutes then cooled and maintained at (b) 60°C, (c) 100°C, or (d) 100°C for 24 hours, allowed to slowly cool, then maintained at 60°C. Expected exact masses: $w_1 = [M_{4M\times 4F}+Na]^+ = 2099.8$; $w_2 = [M_{4M\times 4F}+2Na-H]^+ = 2121.8$; $x_1 = [M_{4F\times 2M2F}+H]^+ = 2280.0$; $x_2 = [M_{4F\times 2M2F}+Na]^+ = 2303.0$; $x_3 = [M_{4F\times 2M2F}+2Na-H]^+ = 2324.0$; $y_1 = [M_{4M\times 2M2F}+Na]^+ = 2474.0$; $y_2 = [M_{4M\times 2M2F}+2Na-H]^+ = 2496.0$; $y_3 = [M_{4M\times 2M2F}+MeCN+Na]^+ = 2515.1$; $z_1 =$

$[M_{2M2F\times2M2F}+Na]^+ = 2676.2$; $z_2 = [M_{2M2F\times2M2F}+MeCN+Na]^+ = 2717.2$. Exact masses found: $w_1 = [M_{4M\times4F}+Na]^+ = 2099.8$; $w_2 = [M_{4M\times4F}+2Na-H]^+ = 2121.8$; $x_1 = [M_{4F\times2M2F}+H]^+ = 2280.0$; $x_2 = [M_{4F\times2M2F}+Na]^+ = 2303.0$; $x_3 = [M_{4F\times2M2F}+2Na-H]^+ = 2324.0$; $y_1 = [M_{4M\times2M2F}+Na]^+ = 2474.0$; $y_2 = [M_{4M\times2M2F}+2Na-H]^+ = 2496.0$; $y_3 = [M_{4M\times2M2F}+MeCN+Na]^+ = 2515.0$; $z_1 = [M_{2M2F\times2M2F}+Na]^+ = 2676.2$; $z_2 = [M_{2M2F\times2M2F}+MeCN+Na]^+ = 2717.1$. 171

Figure 4.16 Dynamic covalent assembly of sequence-defined molecular ladders. ESI mass spectra of molecular ladder reaction mixture aliquots removed from hybridization solutions maintained at 60°C. Each time point shown indicates the period at 60°C. Expected exact masses: $w_1 = [M_{4M\times4F}+Na]^+ = 2099.8$; $w_2 = [M_{4M\times4F}+2Na-H]^+ = 2121.8$; $x_1 = [M_{4F\times2M2F}+H]^+ = 2280.0$; $x_2 = [M_{4F\times2M2F}+Na]^+ = 2303.0$; $x_3 = [M_{4F\times2M2F}+2Na-H]^+ = 2324.0$; $y_1 = [M_{4M\times2M2F}+Na]^+ = 2474.0$; $y_2 = [M_{4M\times2M2F}+2Na-H]^+ = 2496.0$; $y_3 = [M_{4M\times2M2F}+MeCN+Na]^+ = 2515.1$; $z_1 = [M_{2M2F\times2M2F}+Na]^+ = 2676.2$; $z_2 = [M_{2M2F\times2M2F}+MeCN+Na]^+ = 2717.2$. Exact masses found: $w_1 = [M_{4M\times4F}+Na]^+ = 2099.8$; $w_2 = [M_{4M\times4F}+2Na-H]^+ = 2121.8$; $x_1 = [M_{4F\times2M2F}+H]^+ = 2280.0$; $x_2 = [M_{4F\times2M2F}+Na]^+ = 2303.0$; $x_3 = [M_{4F\times2M2F}+2Na-H]^+ = 2324.0$; $y_1 = [M_{4M\times2M2F}+Na]^+ = 2474.0$; $y_2 = [M_{4M\times2M2F}+2Na-H]^+ = 2496.0$; $y_3 = [M_{4M\times2M2F}+MeCN+Na]^+ = 2515.0$; $z_1 = [M_{2M2F\times2M2F}+Na]^+ = 2676.2$; $z_2 = [M_{2M2F\times2M2F}+MeCN+Na]^+ = 2717.1$. 173

Figure 4.17 (a) ESI mass spectra were generated of 2M2F×2M2F hybridization reaction mixtures at 60°C. The 3M×3F hybridization reaction mixtures were subjected to a thiol-Michael addition with S and subsequently characterized by additional ESI mass spectra (b) and (c). Expected exact masses: $[M_{2M2F}+TEA+H]^+ = 1428.7$; $[M_{2M2F+S}+TEA+H]^+ = 1548.8$; $[M_{2M2F\times2M2F}+H]^+ = 2654.2$; $[M_{2M2F\times2M2F}+TEA+H]^+ = 2755.3$; $[M_{2M2F\times2M2F+S}+TEA+H]^+ = 2875.3$. Exact masses found: $[M_{2M2F}+TEA+H]^+ = 1428.7$; $[M_{2M2F+S}+TEA+H]^+ = 1548.7$;

$[M_{2M2F\times2M2F}+H]^+$ = 2654.2; $[M_{2M2F\times2M2F}+TEA+H]^+$ = 2755.3; $[M_{2M2F\times2M2F+S}+TEA+H]^+$ = 2875.3.. 174

Figure 5.1 Peptide sequences, expected structures and LC/MS results. 184

Figure 5.2 Schematic of blood-typing CAT assay and the introduction of antibody-peptide bioconjugates to produce SARS-CoV-2 serology assay. (a) In a typical blood typing assay, RRBCs are incubated with patient samples on a gel-card prior to centrifugation to generate a pattern of agglutination results to determine a blood type. (b) Reaction scheme employed to produce the antibody-peptide bioconjugate in a two-step process. 188

Figure 5.3 SDS-PAGE gel showing results of Protein A-purification of anti-D-IgG FFMU supernatant to purified and desalted anti-D-IgG. “Protein A FT” = flow through from Protein A column; “Protein A Elution” = eluant from Protein A column, “Desalted” = concentrated product after buffer exchange into PBS. 191

Figure 5.4 Anti-D-IgG/peptide bioconjugate characterisation. (a) Fluorescence scan of protein gel under FAM filter. (b) Brightfield image of the same gel following Coomassie staining. (c) Graph showing the bioconjugate binding to D+RRBCs using flow cytometry, and the effect of bioconjugate titration. Dotted line indicates equimolar bioconjugate and D-antigen in the incubation reaction. 192

Figure 5.5 Optimisation of gel-card assays for SARS-CoV-2 serology. (a) Testing the ability of bioconjugates to crosslink RRBCs independent of attached peptide, using anti-IgG in solution in PBS. “P*n*” indicates the peptide used in the reactions (*n* = 1, 2 or 5), and ratios indicate the bioconjugate to D-antigen (on cells). (b) Demonstrating the selective

agglutination of SARS-CoV-2 antibodies present in a clinical sample in comparison to negative controls, using bioconjugate-saturated RRBCs. Reactions involving SARS-CoV-2-positive samples indicated by “+” and SARS-CoV-2-negative samples indicated by “-”. Reactions labelled “P1/2/5” indicate that bioconjugates were pre-mixed to provide the same total peptide concentration as used for other reactions. 193

Figure 5.6 Light microscopy images of RRBCs collected from the top layer of gel cards. (a) False positive gel cards revealed cells that were not agglutinated, and showed some apparent interactions with the gel beads. While retained on top of the gel card, the lack of agglutination confirms the false positive status. This was most often associated with the presence of glycerol in the RRBC stocks, carried over from the bioconjugate solutions. (b) True positive gel cards revealed clear regions of strong agglutination, a 3D network of cells aggregated together in a continuous pattern. 195

Figure 5.7 Clinical sample analysis comparing indirect IgG ELISA against agglutination approach using RRBCs coated with P1, P2 and P5 bioconjugates prior to mixing. (A) Indirect IgG ELISA results comparing 5 PCR-confirmed SARS-CoV-2-positive samples (filled circles) against 5 samples collected from healthy individuals prior to SARS-CoV-2 pandemic (empty circles). The dotted line indicates the limit of quantification (LOQ) for the assay, determined to be three standard deviations above wells containing PBST instead of clinical sample. (B) Digital images of gel card assays comparing 5 PCR-confirmed SARS-CoV-2-positive samples against 5 samples collected from healthy individuals prior to SARS-CoV-2 pandemic. Negative controls (“N”) samples were performed using RRBCs and clinical sample (sample 5 for positives; sample 10 for negatives) without bioconjugates. 198

Figure 6.1 Iterative grid assembly. (a) A bi-directional amine bearing core-strand is (b) reacted with two equivalents of a strand bearing free-aldehydes on one face and alloc-protected amines on another. (c) The assembly is reduced and (d) the alloc-protecting groups are removed for further reactions. (e) Similarly, a bi-directional aldehyde bearing core-strand is (b) reacted with two equivalents of a strand bearing free-amines on one face and acetal-protected aldehydess on another. (c) The assembly is reduced and (d) the acetal-protecting groups are removed for further reactions. 206

Figure 6.2 MALDI- mass spectra of a grid assembly from five strands. 208

Figure 6.3 Self-assembly of a molecular grid from information-bearing peptoid strands. 209

Figure 6.4 (a) Schematic of a molecular junction from the self-assembly of three peptoids and (b) the MALDI mass spectrum of a reaction mixture attempting the assembly. 210

Abstract

The capacity for sequence-specific polymer strands to selectively assemble into intricate, folded structures and multimeric complexes relies upon the information borne by their residue sequences. Particularly suitable for the formation of multi-dimensional structures, nucleic acids have emerged as sophisticated nanoconstruction media where encoded sequences self-assemble in a designed manner through the gradual cooling of denatured and dissociated strands from raised temperatures. Unfortunately, the weakness of the hydrogen bonds holding the strands together affords nanoconstructs with thermal and mechanical instabilities. In contrast, molecular self-assembly employing dynamic covalent interactions has contributed to the improved mechanical and chemical stabilities of resultant structures. Nevertheless, compared with supramolecular chemistries, dynamic covalent interactions suffer from low dissociation rates, impeding rearrangement amongst the assembled components and often result in the kinetic trapping of non-equilibrium species. To overcome this limitation, molecular architectures are generally restricted to homo-functionalized constituents bearing few reactive sites or utilize harsh self-assembly conditions.

This dissertation examines the deliberate equilibrium shifting of dynamic covalent interactions to fabricate sequence-selective molecular architectures with high degrees of functionalization. First, we explored the use of a Lewis acidic catalyst, scandium triflate, $Sc(OTf)_3$, to affect the equilibrium of imine formation, a well-characterized dynamic covalent interaction. Here, high concentrations of scandium triflate, dissociated oligomeric-strands encoded with amine- and aldehyde-pendant group species. Upon removal of excess scandium triflate with a liquid-liquid

extraction, the equilibrium was shifted as to promote imine-formation between complementary strands. Subsequent annealing of the self-assembly solutions at 70°C, enabling rearrangement and error-correction of out-of-registry or non-complementary sequences, afforded the simultaneous formation of three distinct information-bearing ladder species and a mechanism for information storage and retrieval of data by abiotic polymers. The information-directed self-assembly of encoded molecular ladders was further developed by incorporating an orthogonal reaction into the oligomeric strands to mimic the information dense, sequence-selective hybridization of DNA. Thus, the base-4 information-directed assembly of molecular ladders and grids bearing covalent bond-based rungs was demonstrated from encoded precursor strands using dual concurrent, orthogonal dynamic covalent interactions (i.e., amine/aldehyde and boronic acid/catechol condensation reactions).

Additionally, the self-assembly of well-characterized ladder species employing the thermally-reversible Diels-Alder cycloaddition reaction was explored to establish a self-assembly mechanism requiring an external stimulus to alleviate or eliminate kinetic trapping. By utilizing furan-protected maleimide and furfurylamine residues, sequence-defined strands were synthesized simultaneously bearing both furan and maleimide species while precluding premature hybridization and self-assembled in an information-directed manner to form distinct ladder species using a temperature-mediated process.

Finally, given the large-scale efforts underway to develop rapid SARS-CoV-2 (Severe Acute Respiratory Syndrome - coronavirus – 2) diagnostic tests, the fundamental principles of sequence-selective hybridization were applied to transform blood-typing tests into SARS-CoV-2 serology tests using robust gel card agglutination reactions in combination with easily prepared antibody-peptide bioconjugates.

Chapter 1 Introduction

1.1 Overview of Research

As progress towards improved control over macro- and supra-molecular nanostructures and materials proceeds, the challenge of designing and fabricating regular, but modifiable, assemblies arises. Numerous complex supramolecular assemblies occur throughout biology, from lipid bilayer membranes and folded proteins to ribozymes and double helical DNA.[1,2] Many of these biological examples result from self-assembly process, in which small sub-units generate larger architectures through thermodynamically-driven pathways by employing intermolecular interactions such as π-stacking and hydrogen bonding.[1] While the inherently weak nature of these intermolecular bonds allows for sub-unit rearrangement and error-checking, the resultant structures are often susceptible to thermal and mechanical degradation.[3,4] Consequently, dynamic covalent interactions, a class of covalent bonds that are reversible or rearrangeable under mild conditions, have recently been employed to yield intricate macromolecules such as ladders, cages, and stacks offering increased bond strengths and robust structures.[5–7] However, these assembled structures typically rely only on a single dynamic covalent reaction, resulting in relatively simple structures[8]; by utilizing other chemistries, it is not only possible to create more complex structures, but permits greater design control on the molecular level[9].

1.2. Nucleic Acid Nanotechnology

Given their remarkable fidelity and high information-density, nucleic acids (DNA and RNA), ubiquitous information-bearing molecules, have been readily employed as nanoconstruction medium for the self-assembly of complex architectures[10–12]. In particular, DNA, in which nucleobase residues encode information within a single deoxyribophosphate strand, has spurred interest with its anti-parallel hybridization to generate double helixes via hydrogen bonding[13]. Importantly, this hydrogen-bonding occurs between orthogonal reaction pairs (i.e., adenine to thymine and cytosine to guanine) to yield sequence-selective molecular ladders encoded in base-4, where there exists 4^n unique sequences of DNA strands bearing n bases. This phenomena lead Nadrian Seeman to assemble the first non-linear self-assembled structure in 1982, heralding the field of nucleic acid nanotechnology (Figure 1.1)[14]. In this seminal work, Seeman employed several asymmetric oligonucleotides, dissimilar to their biological, sequence symmetric counterparts, to afford molecular junctions. Moreover, these junctions can be directly linked by designing overhangs, termed ‘sticky ends’, within each junction that facilitate the ‘sticking together’ of multiple junction species by maximizing the Watson-Crick base pairing[15,16]. Although fundamental and undeveloped, the process actualized by Seeman was able to assemble a cubic molecule composed exclusively of DNA fragments (Figure 1.2)[17].

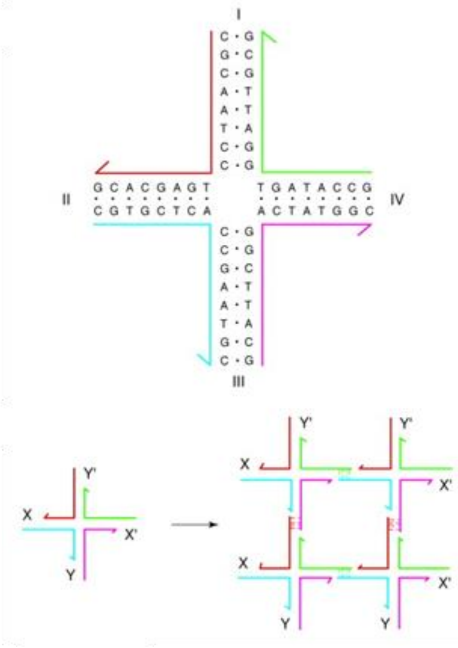


Figure 1.1 A branched DNA molecule with four arms that is able to form multi-junction structures through the use of overhangs or ‘sticky ends’ [14,15].

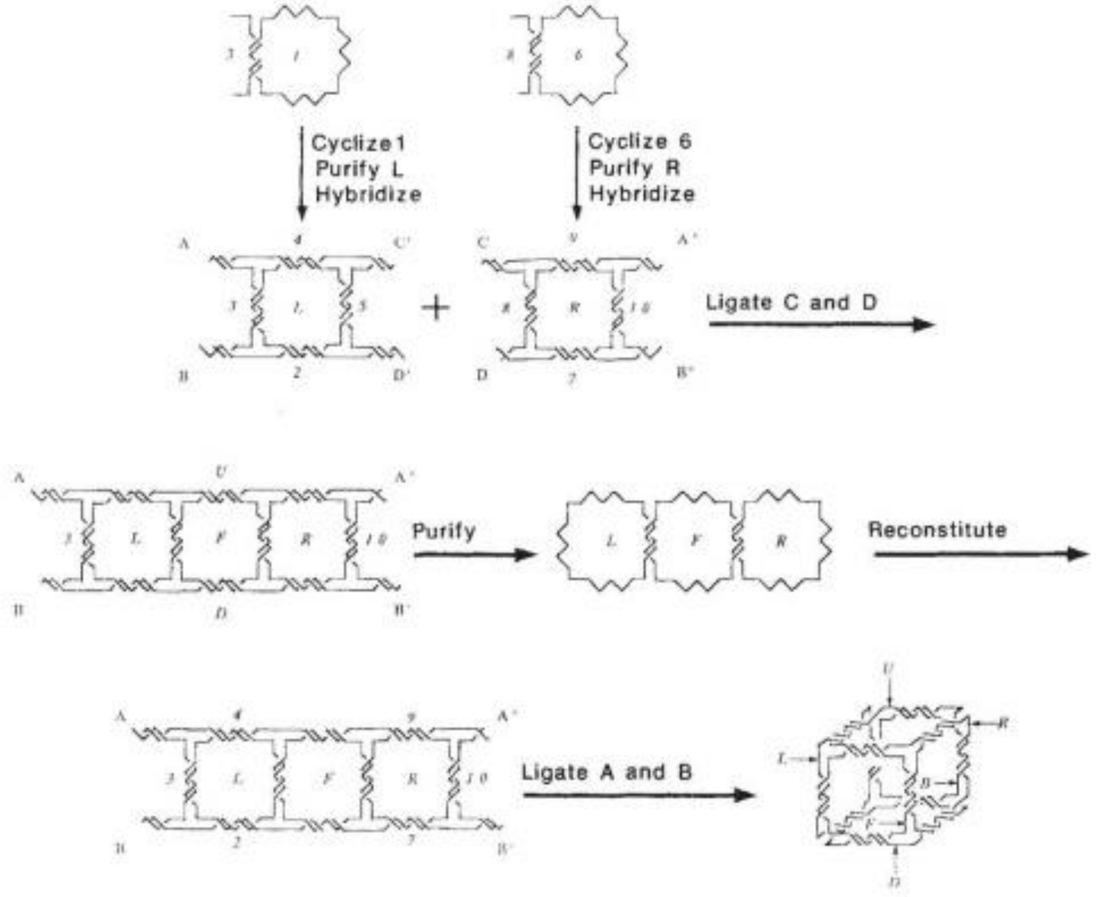

Cyclize 1
Purify L
Hybridize
Cyclize 6
Purify R
Hybridize
Ligate C and D
Purify
Reconstitute
Ligate A and B

Figure 1.2 The generation of a cube entirely from DNA strands by Junghuei Chen and Nadrian Seeman[17].

While the selective hybridization of nucleic acids has been used to fabricate complex nanostructures from organic/inorganic hybrid components through the affixation of complementary DNA sequences on separate metallic or semiconductor nanoparticles to effect aggregation[18], nanostructures composed exclusively of DNA affords more defined architectures with an increased capacity for functionalization and tailoring for applications[19]. However, until fairly recently, these nanoconstructions were limited in complexity owing to the need for exact stoichiometric control and often purification of error-prone species[11]. The introduction of a technique known as ‘DNA origami’ by Paul Rothemund in 2006 was a major catalyst for the construction of DNA architectures (Figure 1.3)[20]. In DNA origami, pre-determined shapes are generated by the raster-filling of a long single-strand of DNA by relatively short, oligomeric 'staple-strands.' The key development in Rothemund's work was establishment of a variable equilibrium method in which strands were dissociated by heating above the denaturing temperature of DNA (≈90°C) before slowly cooling the strands in a process known as annealing, where the steady reduction in temperature allows remarkable selectivity of sequences to form multi-dimensional structures. This process can take up to 20 hours for complex architectures. The flexibility of this technique has been demonstrated to produce a variety of structures including images[19], 'nano-breadboards'[21–23], and molecular containers[24,25].

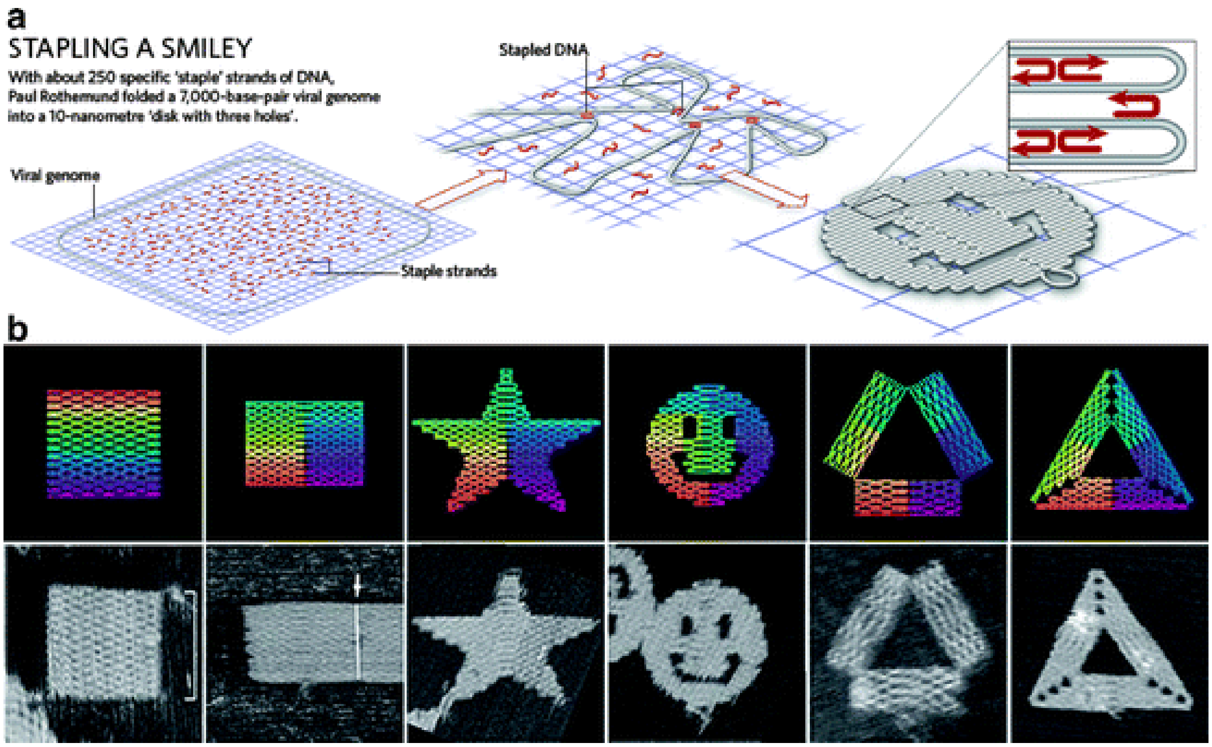

a
STAPLING A SMILEY
With about 250 specific 'staple' strands of DNA, Paul Rothemund folded a 7,000-base-pair viral genome into a 10-nanometre 'disk with three holes'.
Viral genome
Staple strands
Stapled DNA
b

Figure 1.3 The principle design behind DNA origami[26] and resultant structures[20].

Additionally, DNA origami can be used to build 'DNA tiles' and 'DNA bricks' which act similar to Lego bricks, enabling the assembly of complex 3D structures with unprecedented sophistication (Figure 1.4)[27–29]. Resulting structures are large enough to interact with cells, enabling the programming of polymer systems for direct-cell interaction[28].

Despite the significant progress in nucleic acid self-assembly technology, limitation still exist with respect to mechanical, thermal, and chemical stabilities owing to the relatively weak hydrogen bonding holding the structures in place[30]. Hydrogen bonding has a bond strength of only 5-30 kJ/mol while covalent bonds have strengths around 100-1000 kJ/mol. Moreover, deoxyribonucleic acid backbones and their base pairs are limited in functionalization and modification without negatively affecting their capacity for self-assembly. Thus, novel hybridization chemistries, i.e. dynamic covalent interactions, as well as synthetic polymer backbones are required for realizing the potential of the information-directed self-assembly of encoded polymeric species.

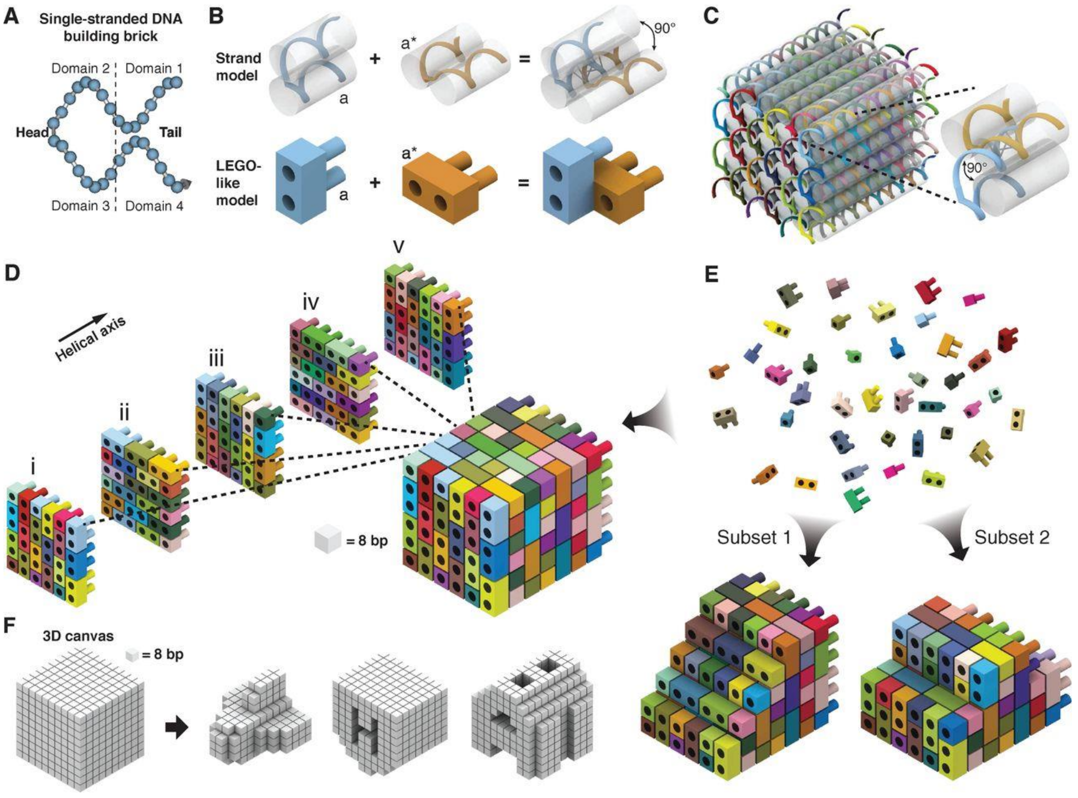

A
Single-stranded DNA building brick
Domain 2
Domain 1
Head
Tail
Domain 3
Domain 4
B
Strand model
a
a*
90°
LEGO-like model
a
a*
C
90°
D
Helical axis
i
ii
iii
iv
v
= 8 bp
E
Subset 1
Subset 2
F
3D canvas
= 8 bp

Figure 1.4 DNA bricks used to form three-dimensional letter structures[28].

1.3. Dynamic Covalent Interactions

Dynamic covalent interactions are a class of reversible covalent bond forming reactions where equilibrium can be adjusted by moderate changes in the environment such as pH, temperature, or light intensity.[31] Examples of such bonds are cleavable alkoxyamine, ester, disulfide, imine, boronic ester, oxime, hydrazone and the Diels-Alder cycloaddition (Figure 1.5). Given the reversibility of these bonds, they can be employed in self-assembly processes and have the potential for realization of error-correction as they find a thermodynamically favored product (Figure 1.6). Outlined below are three dynamic covalent reactions with the greatest potential for utilization in self-assembly.

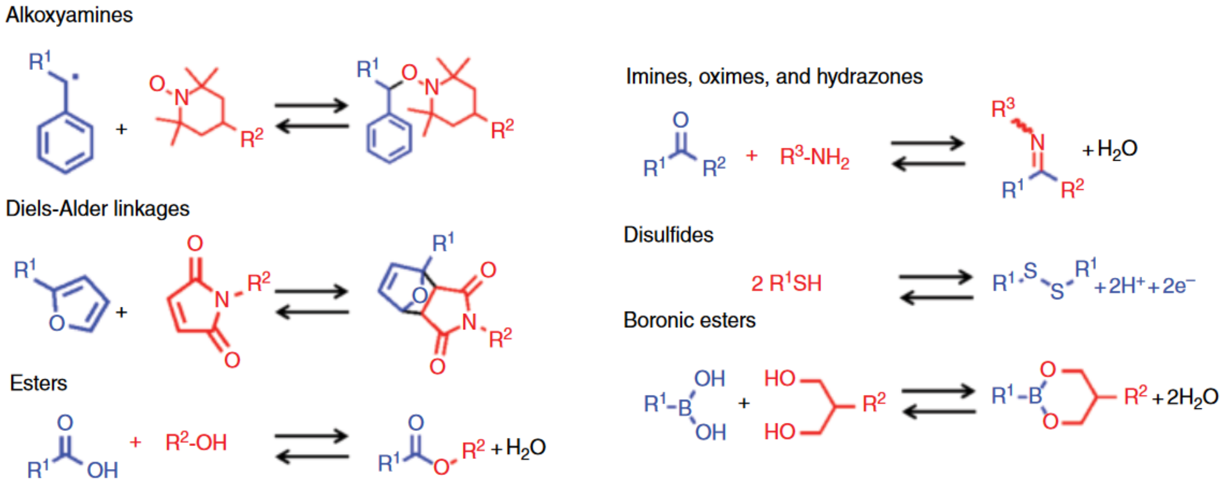

Alkoxyamines
Diels-Alder linkages
Esters
Imines, oximes, and hydrazones
Disulfides
Boronic esters
R1
R2
R3-NH2
+H2O
2 R1SH
+2H++2e−
R2-OH
+2H2O

Figure 1.5 Examples of dynamic covalent interactions[32].

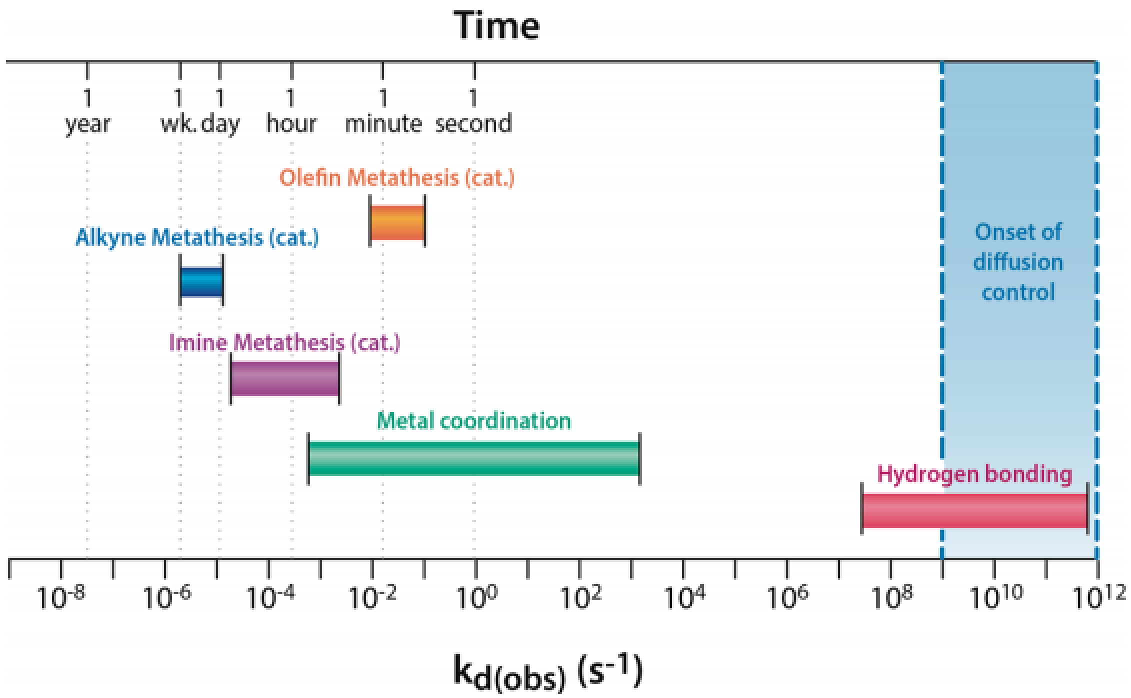


Figure 1.6 Literature values comparing the dissociation rate coefficients for a variety of reversible reactions (k_d(obs))[33].

1.3.1. Imine Chemistry

The formation of an imine, “Schiff base,” from a(n) ketone/aldehyde and an amine is a well-established and dominant dynamic covalent interaction extensively employed in literature[34]. The reversible condensation between amines and aldehydes, whereby amino and carbonyl groups react to form a C=N bond either intra- or intermolecularly with the removal of a water molecule, was discovered by the German chemist Hugo Schiff in 1864 and was named after him[35]. Although several factors affect the equilibrium of imine, including solvent, concentration, temperature, as well as steric and electronic factors, control over equilibrium is generally applied through regulation of pH, whereby ketone/aldehyde and amine reagents are preferred under acidic condition with the imine-product forming under basic environments. However, the use of azeotropic distillation by a Dean-Stark apparatus or the addition of drying agents to the reaction mixture are often seen to drive the equilibrium forward. Interestingly, imines are attractive dynamic covalent reaction products as they're capable of participating in three reactions; imine formation, transimination and imine metathesis (Figure 1.7)[34,36]. Here, imine formation is the reversible, condensation reaction between the amine- and aldehyde-species. Upon introduction of a second amine, the original imine may undergo transimination, contributing to the exchange of R groups. Similarly, the introduction of a second imine can produce the same exchange of R groups, producing two new imines in imine metathesis. These secondary reaction, transamination and imine metathesis, are desirable in self-assembly as they provide multiple pathways and mechanisms for the realization of thermodynamically favoured products, and as such will be the first dynamic covalent interaction studied in the work described herein.

To note, Lehn and colleagues have extensively studied the rate of the transimination reactions in the presence of lanthanide ions as catalysts in which they discovered an inverse relationship between exchange rates and ionic radius, with scandium triflate, $Sc(OTf)_3$, affording the highest rate of exchange with turnover frequencies up to 3600 h-1 and rate accelerations up to 6 × 10^5 [37].

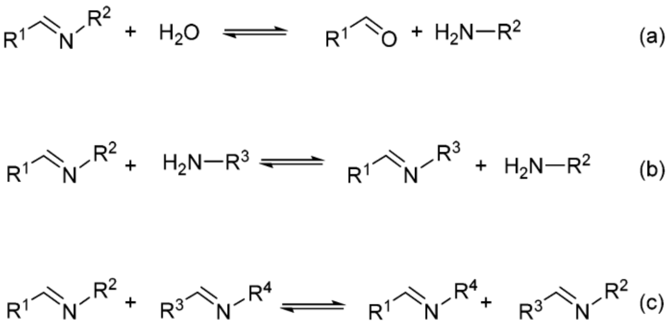


Figure 1.7 (a) Imine hydrolysis, (b) transimination, and (c) imine metathesis reactions[36].

1.3.2. Boronate Ester Chemistry

Another pH-sensitive, condensation reaction that affords a reversible covalent bond is the condensation reaction between a boronic acid and diols[8]. Here again, under acidic conditions, the boronic acid and diol starting materials are favoured, with condensation occurring at elevated pHs to afford boronate esters. Frequently used in materials chemistry as targeted drug delivery systems and crystal engineering, boronate esters have proven to be robust and invaluable products[38,39]. Importantly, boronate ester chemistries have been used in conjunction with imine chemistries to afford molecular structures via self-assembly processes, albeit structures utilizing repeat units[40–42]. However, significant challenges are present with boronate ester chemistries. Foremost, diols are often oxygen sensitive requiring self-assembly under aerobic conditions and complicate characterization of resultant structures[43]. Additionally, boronic acids are capable of participating in undesirable reactions, e.g., self-condensation to form boroxines, potentially impeding the formation of complex, multi-dimensional structures due to a reduction in constituent fidelity (Figure 1.8)[44].

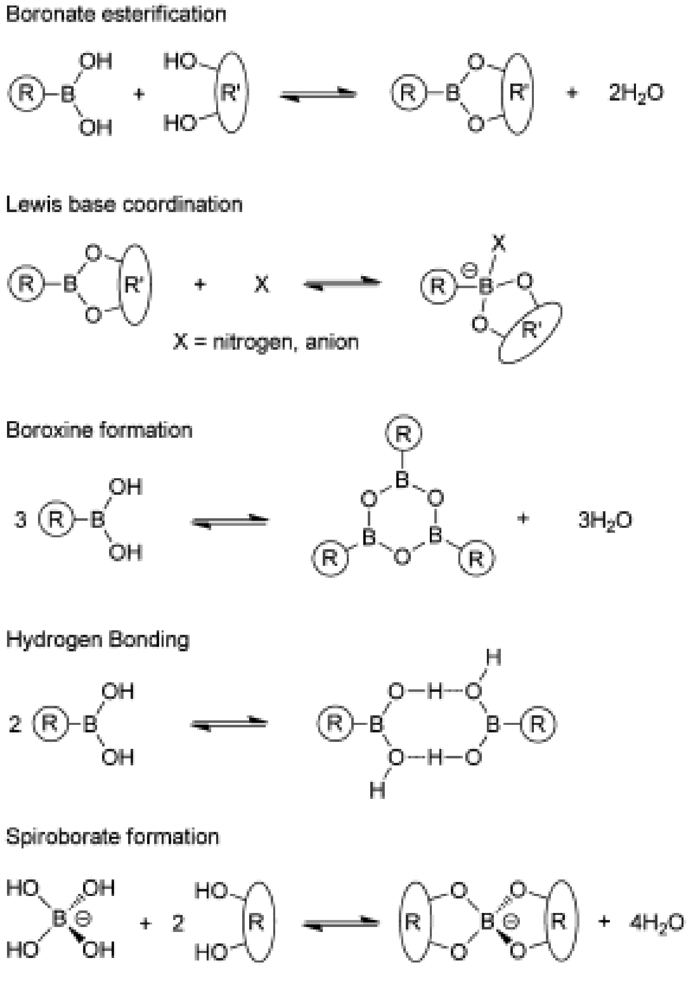

Boronate esterification
+ 2H$_2$O
Lewis base coordination
X = nitrogen, anion
Boroxine formation
+ 3H$_2$O
Hydrogen Bonding
Spiroborate formation
+ 4H$_2$O

Figure 1.8 Key interactions of boron, with respect to self-assembly[44].

1.3.3. Diels-Alder Chemistry

The reversibility of the cycloaddition of Diels-Alder reaction has been acknowledged for over 50 years and is often taught in basic organic chemistry classes[45]. It is inherently exothermic with the reverse reaction occurring at elevated temperatures and necessitates the tailoring of the diene and dienophile starting materials as the number of Diels-Alder reactions who experience dynamically reversible adduct formation under mild conditions is very bound[8]. Equilibration is generally achieved through the prolonged heating of reaction mixtures between 70°C and 100°C with studies showing that the furan substitution dictates the reactivity and reversibility of the system[9,46].

For the past 10 years, several groups have demonstrated the use of reversible Diels-Alder adduct formation in the design and synthesis of novel organic materials, primarily employing electron-rich furan derivatives and electron-poor maleimide derivatives for healable polymeric materials[47,48]. The [4+2] cycloaddition of furan and maleimide can be achieved at moderate temperatures but can exhibit relatively slow rates of formation as compared with imine and boronate ester chemistries.

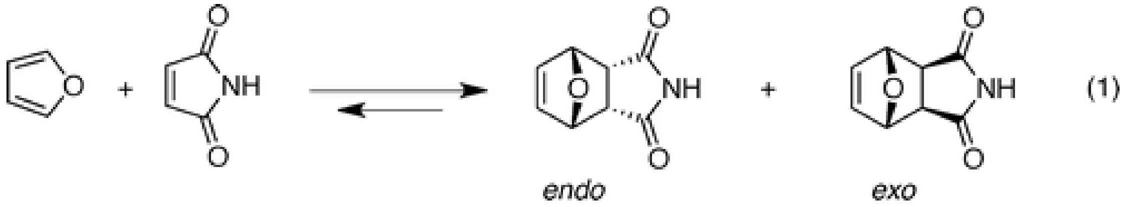


Figure 1.9 Parent [4+2] Cycloaddition of unsubstituted furan with maleimide to provide the kinetic endo and thermodynamic exo diels-alder adducts[49].

Importantly, the Diels–Alder reaction produces a mixture of two diastereomers, an *endo*- and an *exo*-isomer where cyclo-reversion temperature of the *endo*-isomer is at a significantly lower temperature than the *exo*-adduct (Figure 1.9) [46]. The ratio between *endo* and *exo*-adducts differs in relation to the substituents of the Diels–Alder partners and experimental parameters (Figure 1.10 and 1.11)[49].

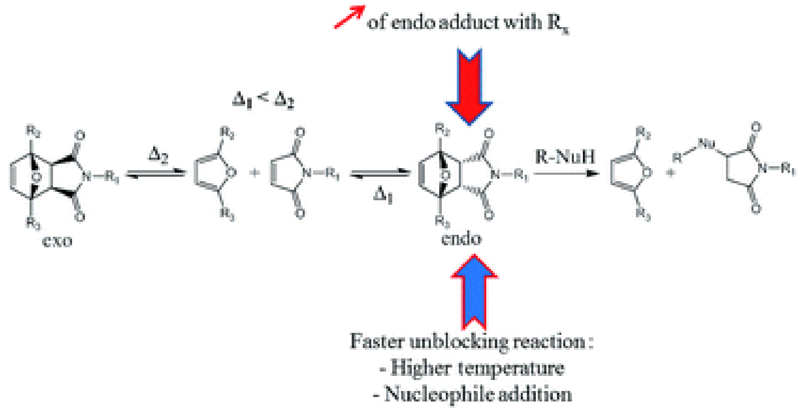

of endo adduct with R_x
$\Delta_1 < \Delta_2$
Δ_2
Δ_1
R-NuH
exo
endo
Faster unblocking reaction :
- Higher temperature
- Nucleophile addition

Figure 1.10 Diels-Alder and retro Diels-Alder reaction between furan derivatives and maleimide for the creation of new materials[49].

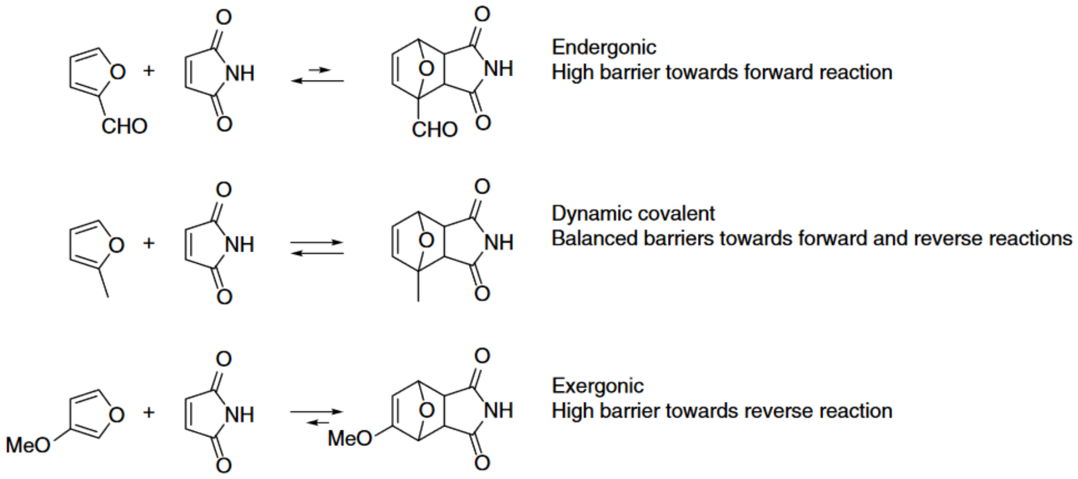

CHO
NH
Endergonic
High barrier towards forward reaction
Dynamic covalent
Balanced barriers towards forward and reverse reactions
MeO
Exergonic
High barrier towards reverse reaction

Figure 1.11 Furan substitution patterns that influence reversibility[50].

1.4. Abiotic, Sequence-defined Polymers

Sequence-defined polymers, uniform macromolecules with precise, though irregular, sequences[51,52] found extensively in nature (e.g., nucleic acids, peptides, etc.), have elicited significant interest owing to their facility for high density data storage[53–55], capacity for their encoded information to direct the self-construction of complex nanostructures[11], and potential for semi-conservative replication and templating macromolecule synthesis[56–58] (Figure 1.12).

The synthesis of DNA sugar-phosphate backbone analogues to afford new classes of abiotic, information-bearing polymers has received tremendous research attention in recent decades[59,60]. Interestingly, several backbone analogues of DNA, termed xeno-nucleic acids (XNAs), exhibit the ability to both self-pair and cross-pair with DNA or RNA to form stable duplex structures via canonical Watson-Crick base pairing with similar hybridization fidelity to double-stranded DNA[12]. Moreover, cross-pairing with naturally occurring nucleic acids has enabled the templated enzymatic synthesis of DNA from various synthetic analogues[58]. As there exists abundant literature on the properties and applications of XNAs, the reader is directed to previous reviews on the subject[59–62].

The encoding of information into the monomeric residue sequence of synthetic polymers, via either the polymer backbone or pendant groups, is by now well-established[52]; indeed, numerous methods for synthesizing sequence-defined polymers exist[51]. Solid phase synthesis, where sequential reaction and purification steps are performed using cross-linked support resins, has proven to be a powerful technique to yield sequence-specific oligomers since Merrifield synthesized the first tetrapeptide on-resin[63]. Additionally, living polymerization[64–66], reversible-deactivation radical polymerization[67,68], and solution-phase iterative methods have been successfully developed to afford sequence-defined macromolecules[69].

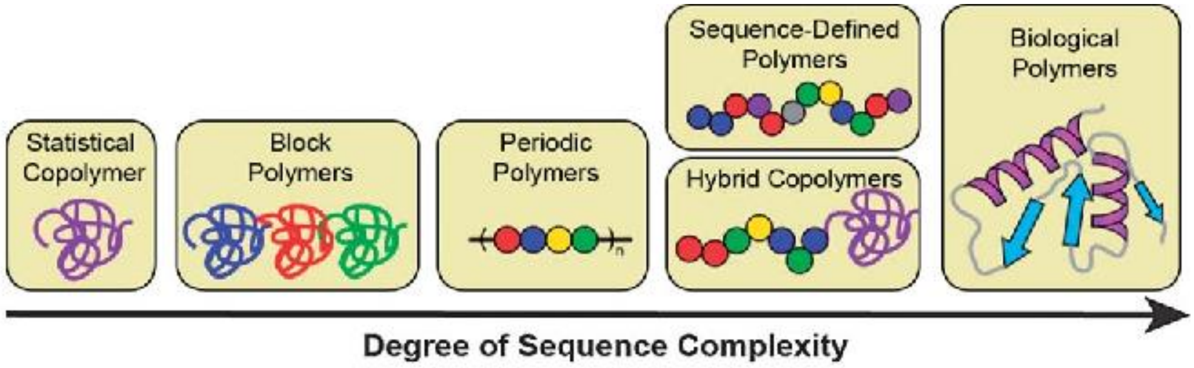


Figure 1.12 Furan substitution patterns that influence reversibility[70].

1.4.1. Solid-phase Synthesis

The step-wise, iterative synthesis of oligomers on insoluble cross-linked support resins allows for the washing of unreacted reagents and minimizes errors in polymer growth[71,72]. Moreover, utilizing orthogonal chemistries and strategies, sequence-specific formation of oligomers is possible through the designed addition of distinct monomeric residues. Although developed by Merrifield for the preparation of peptides, the facile synthetic technique has been employed for the synthesis of a wide-range of resin-bound organic compounds including DNA, RNA, and peptidomimetic polymers[73]. Poly(N-substituted glycine)s (i.e., peptoids), structural isomers of peptides in which side chains are nitrogen-bound instead of on the α-carbon, are examples of abiotic peptidomimetic polymers that are routinely synthesized on-resin[74,75]. Specifically, Zuckerman's sub-monomer solid phase synthetic method is often utilized in the preparation of sequence-defined oligomers (Figure 1.13). In this method, peptoids are synthesized in an iterative, step-wise manner via bromoacetic acid coupling to free amines, followed by halide displacement using a primary amine monomer[76]. The synthetic accessibility of primary amines, the monomer responsible for peptoid pendant groups, enables the facile modification and functionalization of peptoid species. Moreover, the peptoid backbone is inherently neutral, achiral, and lacks hydrogen bond donor sites[74].

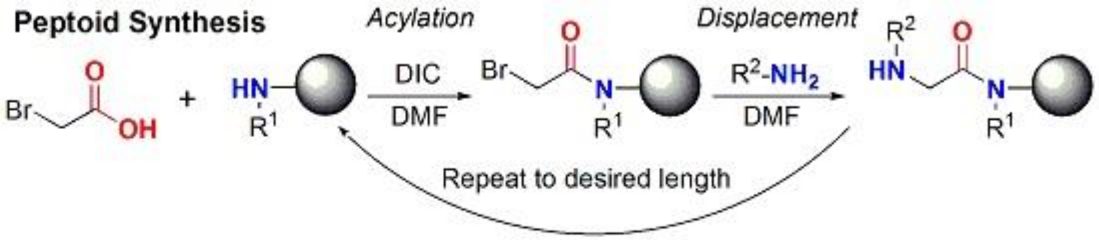


Figure 1.13 Peptoid synthesis via Zuckerman's submonomer aporach to solid-phase synthesis[74,75].

Given their facile synthesis and the accessibility of functional group monomers, peptoids are primarily used for the self-assembly process described herein, however other methods of synthesizing sequence-defined polymeric species will be discussed briefly as they have potential as backbones for similar self-assembly processes.

1.4.2. Solution-phase Synthesis

As with solid-phase synthesis, the total control of macromolecular sequence in solution-phase synthesis requires highly efficient coupling chemistries, known as "click chemistries." However, dissimilar from solid-phase synthesis, monomers are often added in a 1:1 ratio with growing polymer chains or require extensive purification steps between coupling steps, substantially impeding time for total synthesis[77,78]. Iterative methods are often used to achieve a sequence-defined polymer through multiple, orthogonal additions or deprotection and coupling steps[79]. However, recently, to improve the rates of polymer formation several rapid methods for solution-phase generation of polymers have emerged (Figure 1.14). Iterative growth is one such strategy where sequential deprotection of growing polymer chains allows for the coupling of these chains to each other, doubling the length iteratively from monomers to dimers, dimers to tetramers, tetramers to octamers, and so forth[80]. Unfortunately, his results in repeated or palindromic oligomer sequences but allows for an exponential growth of chains. Similarly, a bi-functional monomer can initially be used with the iterative monomer addition to double chain length compared to mono-functional initial monomers, and results in a symmetric polymer sequence[81]. Similarly, a recently a method has been established in which a trimeric-linker grows three polymer chains simultaneously, but allows the facile purification between monomer residue additions through filtration, similar to solid-phase but using a membrane filter, and can be cleaved to obtain each polymer chain independently.

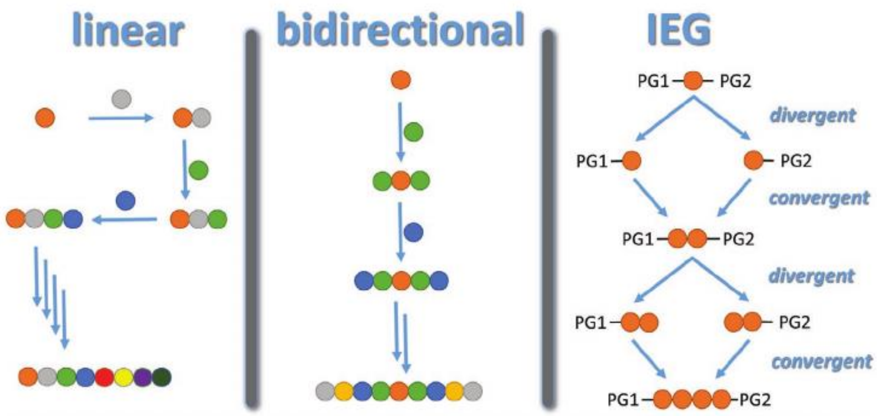

linear
bidirectional
IEG
PG1
PG2
divergent
convergent
divergent
convergent

Figure 1.14 Schematic representation of solution-phase synthesis of polymers by linear, bidirectional, and iterative exponential growth mechanisms[82].

1.4.3. Living Polymerization

Living polymerization techniques for the synthesis of sequence-defined polymeric species are of growing interest owing to the possibility of rapid polymer formation[64,66]. To date ring-opening polymerizations (ROPs) and ring-opening metathesis polymerizations (ROMPs) have been the most successful[65,83]. Here, monomers are programmed with a short sequence and then polymerized through ROP or ROMP to produce discrete, uniform polymeric chains with defined repeating sequences. This method also allows for the facile incorporation of chirality into the backbone. Controlled radical polymerizations have also been exploited for sequence control, but ultimately all living polymerization methods still exhibit chain-length dispersity and chain-to-chain deviations[67,84].

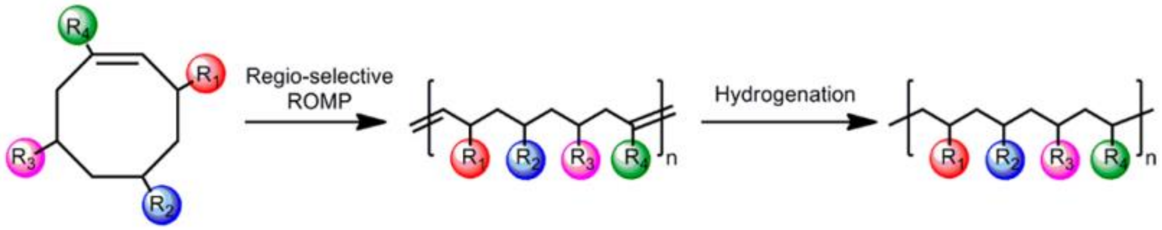

R1
R2
R3
R4
Regio-selective ROMP
Hydrogenation
n

Figure 1.15 Schematic representation of sequence-specific polymers by regioselective ROMP[83].

1.5. Nucleic Acid Mimicry

DNA, an archetypal sequence-defined polymer which selectively dimerizes with its sequence complement, has been extensively explored in many applications[10,85,86], a consequence of the relative dimer stability[87], the development of sophisticated synthetic modification techniques[88,89], and mature methods for replication and amplification[90,91]. Nevertheless, the development of abiotic sequence-defined macromolecules capable of mimicking the properties of DNA has gained prominence, yielding backbone and pendant group diversity and providing routes to tailor chemical and physical properties. Owing to rapid progress in the field of synthetic polymer chemsitry, it is now possible to realize wholly synthetic systems that exhibit features analogous to DNA[92]. This section seeks to provide an overview of the current state of abiotic, sequence-defined polymers dissimilar to the DNA backbone while mimicking nucleic acid attributes, and examines their applications in *(i)* data storage and retrieval, *(ii)* self-assembly of duplexes, and *(iii)* replication and synthetic templating of new macromolecules.

1.5.1 Encoding and Decoding Abiotic Polymers

The utilization of macromolecules as information storage media was explored in 1988 in the work known as ‘Microvenus’, where information to depict a graphic image was encoded in a sequence of DNA[93]. More recently, DNA has been employed as an information storage medium for several large works including 70 billion copies of a 53,426 word book[86], Shakespeare’s Sonnets[94], and the 16 GB entirety of the Wikipedia website[95]. Indeed, with a theoretical maximum of 455 exabytes per gram of single strand[86], roughly 1 kg of DNA would be sufficient to store the world’s expected total digital data in 2040 (3×10^{24} bytes) – equivalent to the storage capacity of approximately 10^9 kg of silicon wafers[96].

Given DNA's extraordinary information density and potential as a data-storage technology, chemists have sought to replicate this capability with synthetic polymers, potentially exceeding the information density of nucleobases and enabling data repositories in environments ill-suited for biological matter[55]. Colquhoun and Lutz identified two essential requirements for the consideration of a synthetic polymer as a molecular data storage medium: (i) a writing mechanism, and (ii) a sequence-reading mechanism[97]. As noted previously, multiple robust methods have been developed for writing or storing data in abiotic polymers by tailoring monomer sequence during synthesis[52,78,98]. In contrast, the complete decoding of a sequence-defined macromolecule (i.e., sequencing) remains the more challenging requirement. Whereas DNA can be sequenced via a number of approaches[54,91,99], few exist for synthetic polymers[54]. Tandem mass spectrometry (MS), first established for peptide and peptoid sequencing, has emerged as the primary technique to determine the sequence of abiotic polymers[100,101]. Whereas alternative sequencing methods, including thermal degradation of the polymeric chain and subsequent characterization of the resultant fragments by MS[102], have been explored though seldom used.

In 2015, Lutz, a pioneer in the field of information storage by synthetic polymeric materials, was the first to consciously design abiotic oligomers for the purpose of writing and sequencing encoded data[103]. Inspired by Porel and Alibi's work utilizing tandem MS to sequence an encoded 16-mer polymer[104], Lutz and coworkers investigated non-natural polymers which could readily be dissociated in tandem MS, but in a robust manner such that strong fragmentation patterns were discerned – simplifying reading of the sequence. They determined that poly(alkoxyamine amide)s synthesized using solid phase protecting-group-free chemoselective coupling steps underwent facile sequencing by MS/MS as low-energy fragmentation proceeded between a TEMPO comonomer and a coding unit[105] (Figure 1.16a). Sequences were encoded by alternating the formation of an N-substituted amide by the reaction of a primary amine with one of two interchangeable symmetric anhydride species (2-bromo-isobutyric acid or 2-bromopropionic acid permitting either a 1 or 0 amide synthon-bit to be recognized by MS) and subsequently, forming an alkoxyamine by coupling a carbon-centered radical with a nitroxide. Interestingly, sequencing was achieved by reconstruction from both ends of the oligomers, identified by either a carboxymethyl moiety (α end-group, c fragments) or a bromine (ω termination, y fragments), as only the largest radical in a given complementary fragment (i.e., y_4^+ given the c_1^+/y_4^+ pentamer fragment pair) could be detected. In this initial work, oligomeric sequences of up to five bits were encoded and sequenced.

Another information-storage polymers to come from Lutz and coworkers, poly(triazole amide)s exhibited a more complicated fragmenting pattern than poly(alkoxyamine amide)s as oligomers underwent both amide cleavage and dissociation of last ethylene oxide unit; an issue potentially resolved by substitution of the oligo(ethylene glycol) segments in each chain with alkyl chains (Figure 1.16b)[106]. Interestingly, the preferential charge state (z) for sequence fragments increased with polymerization degree by the relationship z = integer part of [(n +1)/2]. Thus for a pentamer fraction, the triply-charged state was the major detected signal with doubly- and quadruply- charged states each contributing less than 2.0% of the signal. In contrast, MS/MS spectra of poly(alkoxyamine amide)s obtained in the negative ion mode exclusively exhibited one charge signature. Although this elaborate charge state pattern complicated sequencing by MS/MS, the successful sequencing of an 8-bit, or 1 byte, encoded-oligomer was demonstrated. Notably, the authors suggest the additional sequencing complexity afforded by this system could be particularly advantageous for applications requiring extensive security included encrypted materials and anticounterfeit technologies.

Soon after these initial works, Lutz and coworkers expanded the library of synthetic polymers exhibiting information storage capabilities to include poly(alkoxyamine phosphodiester)s (Figure 1.16c)[107]. Here, alternating phosphoramidite coupling with nitroxide radical coupling affords oligomeric sequences containing both the readily cleavable alkoxyamine bonds previously studied and the synthetic accessibility of various phosphoramidite monomers. Unlike the poly(alkoxyamine amide)s system, all theoretical fragments were observed for the poly(alkoxyamine phosphodiester)s, ensuring unambiguous sequence reconstruction and leading to the sequencing of a strand containing up to eight bits. Ouahabi et al. later employed the weak NO-C bond afforded by the alkoxyamine group to design polymeric sequences in which readily cleavable alkoxyamine groups were placed in between segments of poly(phosphodiester)s encoded with a byte of information, enabling initial fragmentation and sequence identification of byte order by MS/MS at low collision energies and subsequent MS^3 sequencing of each poly(phosphodiester) byte with increased fragmentation (Figure 1.17)[108]. Bytes were labelled with commercially available nucleotides as mass tags, permitting knowledge of the location in the final 64 bit (8 bytes) reconstructed sequence.

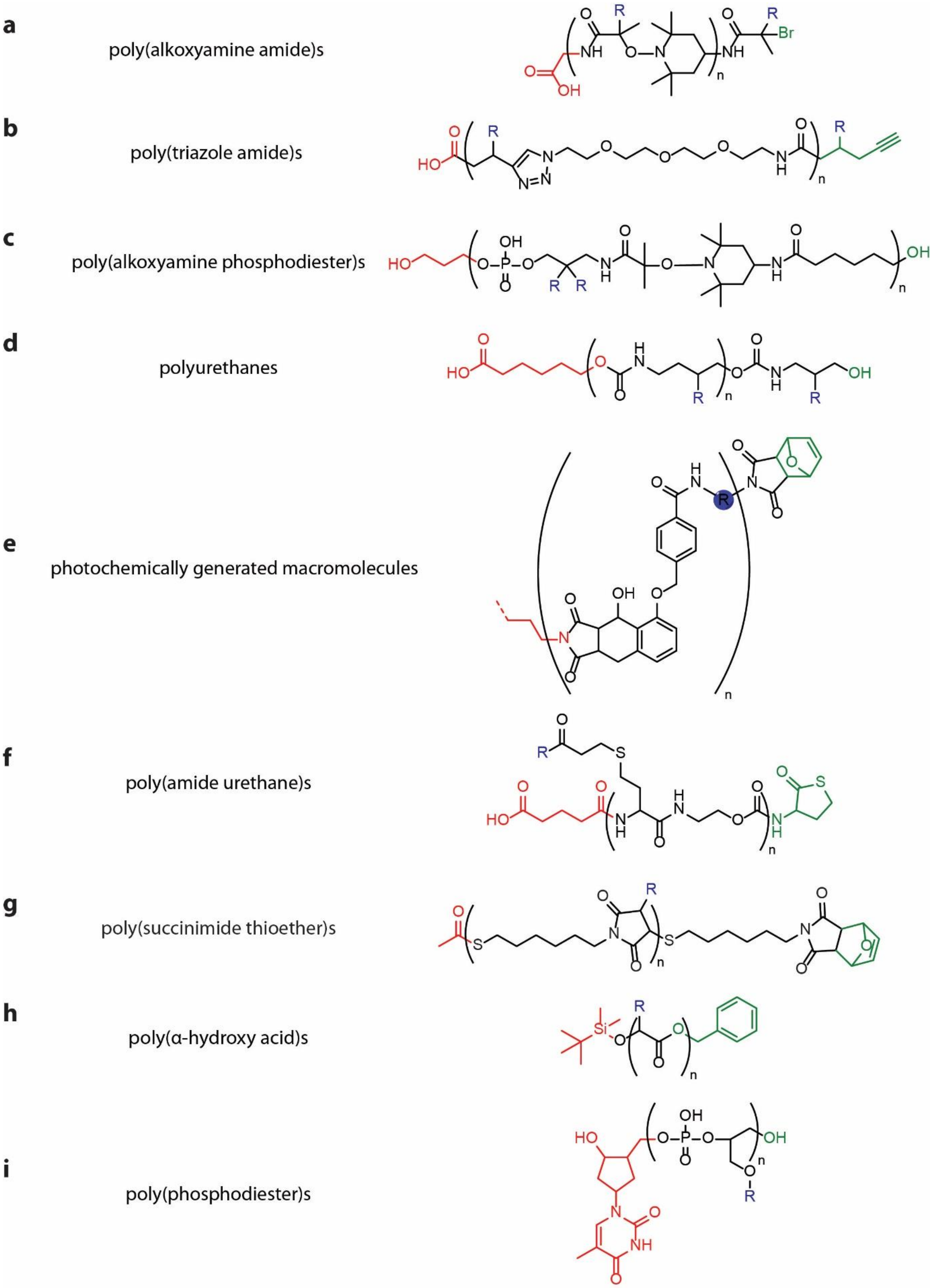

a
poly(alkoxyamine amide)s
b
poly(triazole amide)s
c
poly(alkoxyamine phosphodiester)s
d
polyurethanes
e
photochemically generated macromolecules
f
poly(amide urethane)s
g
poly(succinimide thioether)s
h
poly(α-hydroxy acid)s
i
poly(phosphodiester)s

Figure 1.16 Reported synthetic sequence-defined macromolecules employed for data storage and retrieval.

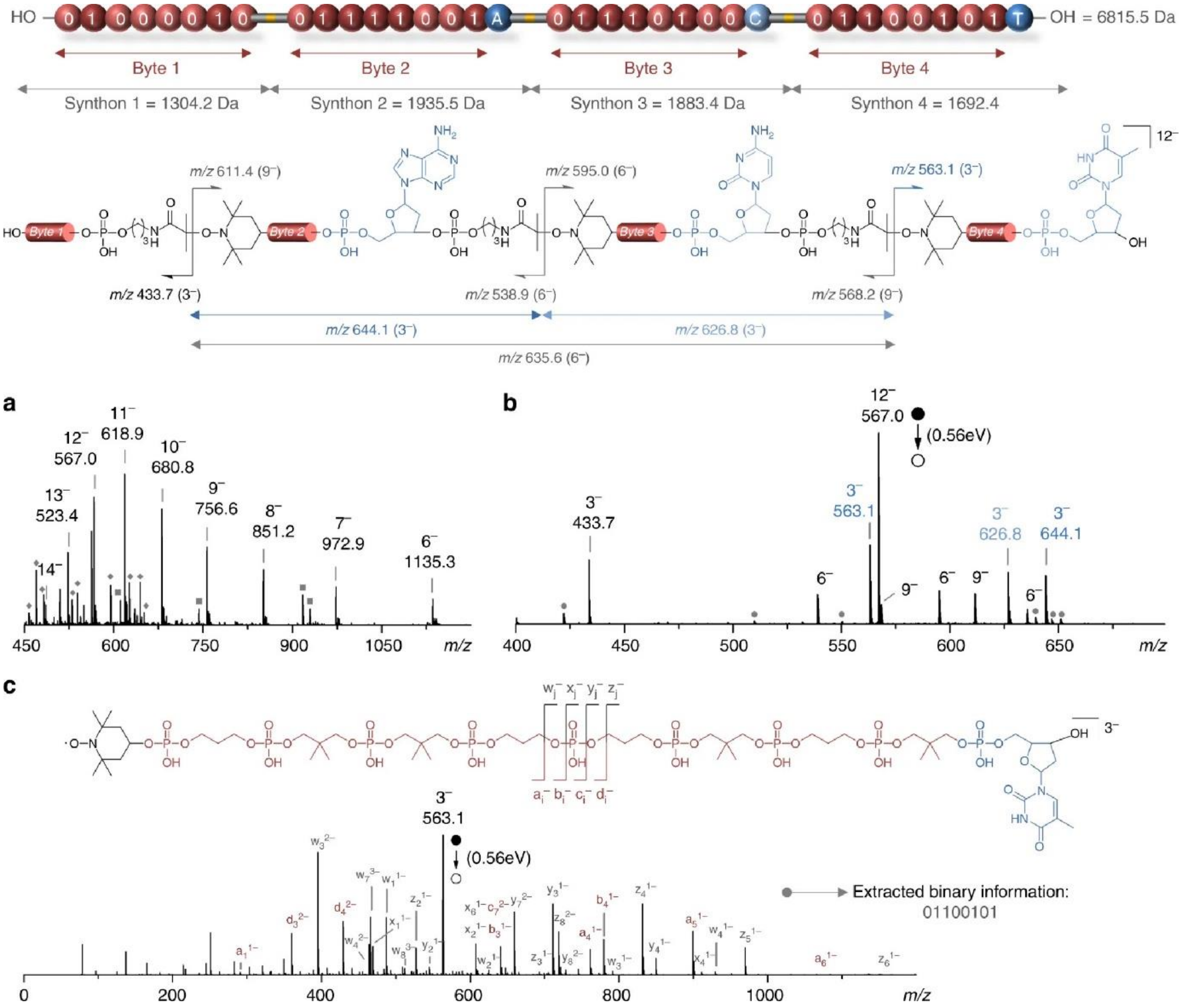


Figure 1.17 Sequencing of a 4-byte digital polymer that contains the ASCII-encoded word Byte. (A) High-resolution electrospray mass spectrum (MS^1) obtained in the negative ion mode for a 4-byte digital polymer. (b) MS^2 spectrum (0.56 eV, center of mass frame) obtained by collision-induced dissociation of the $[M\text{-}12H]^{12-}$ precursor ion. (c) Molecular sequencing (pseudo-MS^3) of a byte fragment obtained by collision-induced dissociation (0.56 eV, center of mass frame) of the precursor trianion $[M\text{-}3H]^{3-}$. For clarity, only the sequencing of byte 4 is shown as an example in this figure.[108]

Given the wide use of polyurethanes in commercial applications, Lutz and coworkers investigated their capacity to form sequence-defined strands and their ensuing decoding by tandem MS. Encoded oligomeric sequences were achieved by a chemoselective multistep-growth process; where first a hydroxyl group couples with *N,N'*-disuccinimidyl carbonate and then forms a carbamate linkage with an amino alcohol[109]. Discreet, uniform polyurethanes are synthesized by repeating these two steps and encoded with amino alcohols bearing or omitting methyl side groups (Figure 1.16d). Interestingly, while positive mode MS/MS results in complicated spectras as a consequence of ionized C-O and C-N carbamate fragmentations, the resulting polymers were easily sequenced in negative mode MS/MS where the C-O carbamate bond cleavage is detected exclusively. In this work, Gunay et al. explored incorporation of the encoded polyurethane sequences as molecular tags in both polystyrene membrane films and methacrylate-based 3D prints. In both examples, the sequence-coded polyurethane could be imprinted and later decoded by removing a small portion of material, dissolving in methanol, and analysing by MS/MS sequencing. Lutz and coworkers went on to further explore the ability for the chemically resistant polyurethanes to be employed as molecular barcodes for intraocular implants[110].

Shortly after Lutz and coworkers first published their initial work on poly(alkoxyamine amide)s for information storage, Zydziak et al. employed matrix-assisted laser desorption/ionisation-time-of-flight (MALDI-ToF) to decode a novel class of sequence-defined, photochemically generated macromolecules (Figure 1.16e)[81,111]. Oligomers were prepared *via* photoligation and thermal deprotection; photoreactive monomers (two unique monomers were investigated to encode information) containing a furan-protected maleimide were reacted with unprotected maleimides under ultraviolet irradiation, heating of the product released the furan and exposed the maleimide for further reaction with fresh monomer extending the polymeric-chain[81]. Although limited to the sequencing of low molecular weight oligomers by MALDI-ToF-ToF (ca. 3000 m/z), Zydziak et al. successfully decoded symmetric hexamers[111]. An advantage of the MALDI-ToF-ToF system is the lack of higher-order charge states, providing easy-to-read spectrum considering fragmentation of these oligomers predominantly occurred between monomer units.

In an effort to expand the information density possible per repeat unit, Boukis and Meier combined the Passerini three-component reaction and Biginelli three-component reaction to develop a method for encoding sequence-defined polymers with up to 24 bits per repeat unit[112]. With a library of 116 components and six possible components to substitute, over 2×10^7 unique repeat units are theoretically possible, although only 33 different components were demonstrated resulting in 13 bits per repeat unit. Recognizing the importance of distinct fragmentation patterns for sequencing, Boukis and Meier conducted collision experiments at various fragmentation energies to identify decarboxylation, Mc Lafferty rearrangements, and isocyanide formation as the dominant fragmentation mechanisms. With this knowledge, tetramers were decoded containing 97 bits, however, the authors acknowledged that the large number of fragments per repeat unit introduced a limitation in the decrypting of longer sequences.

With an increase in the length of information-encoded sequences, came the heightened difficulty in manually reconstructing the sequence from MS/MS-spectra. Inspired by this challenge, Martens et al. developed Chemreader, an automated sequencing tool for MS/MS-spectra[113]. Given their controlled fragmentation, oligo(amide-urethane) hexamers (Figure 1.16f) were employed to develop the Chemreader algorithm. Inputted with the length of the sequence and the masses of individual encoded monomers, Chemreader predicts all possible fragments, searches for their masses in the MS/MS spectra, and reconstructs the sequence – decoding a 20-character octamer in milliseconds. To demonstrate Chemreader’s potential for reading human-readable data, the question “TO WRITE OR NOT TO WRITE ON OLIGOS?” was encoded on short oligomers with each oligomer representing one word and a 19 acrylate alphabet representing the letters and question mark. Mass tags labelled the position of the encoded oligomers in the final question (“1TO 2WRITE 3OR 4NOT 5TO 6WRITE 7ON 8OLIGOS?”) which was correctly reconstructed by Chemreader. Unfortunately, oligomers must be analysed individually, prompting the authors to suggest employing liquid chromatography to separate a given mixture of encoded oligomers prior to tandem MS, a technique currently applied in peptide chemistry[114].

Recently, Zhu and coworkers used dithiosuccinimide motifs, with low C-S bond dissociation energies, to encode information on monodispersed polymers with large degree of polymerization at the gram scale (Figure 1.16g)[115]. However, decoding was merely demonstrated with oligomers up to twelve bits[116]. Additionally, Lee et al. employed poly(phenyllactic-*co*-lactic acid), with an information density (bit/Da) 50% higher than DNA, to encode and decode a 64-bit word, 'SEQUENCE' from eight ASCII characters (Figure 1h)[117]. Polymers consisting of more than 64 repeat units were unable to be directly sequenced by tandem MS, and so the ester groups found in the main chain were hydrolysed to generate constituent monomers decoded in MALDI-TOF, affording the sequencing of a 128-bit strand.

With the expanding library of potential polymers for data storage Lutz and coworkers moved towards spatially organizing unique encoded-strands for use as a molecular archive[118]. Polyphosphodiesters (d-PPDE, Figure 1i)) were selected given their synthetic accessibility and encoded with the citation: "La chimie crée son odjet. Cette faculté créatrice, semblable á celle de l'art lui-même, la distingue essentíellement des sciences naturelles et historiques" from the French chemist, Marcellin Berthelot. Each polymer contained 10 bytes of information, thus the 160 characters, encoded in ASCII, required 16 different encoded-polymer sequences of 81 monomers. Deposition with layer-by-layer (LbL) assembly employing both non-coded poly(allylamine hydrochloride) and poly(sodium 4-styrenesulfonate) layers between d-PPDE layers, allowed for linear growth and controlled layer separation. The 16 d-PPDE layers required about one hour to deposit by spin-assisted deposition. While Szweda et al. suggested sequencing by time-of-flight secondary ion mass spectrometry (ToF-SIMS) or deposition electrospray ionization; decoding was not shown for the resultant films.

From invisible inks[119] to fluorescent sensors[120], chemicals have long been used for the purpose of espionage – an application quickly adapted to encrypted, abiotic macromolecules (Figure 1.18). To this end, Boukis et al. proposed employing an Ugi four-component reaction to synthesize 500,000 unique molecular keys for use as decryption codes[121]. Libraries of the four components (noted as A, B, C, and D) used in the Ugi reaction are assigned arbitrary numbers resulting in an alphanumerical code for decryption (A001, B002…). The chosen 18-bit oligomeric key was deposited on a material previously agreed upon between the message transmitter and recipient (i.e., adsorbed on the upper right-hand corner of an envelope). Upon delivery of the envelope and an encrypted file, the message recipient extracted and purified the key, readout the alphanumerical code by tandem-MS, and ultimately applied the code to decrypt the secret communication. Similarly, Charles and coworkers used poly(alkoxyamine phosphodiester)s to encode an encrypted message sequence with a decryption key built into the alkoxyamine moieties retrieved by MS/MS-ion mobility spectrometry[122]. The decryption key identified the encryption rules used to encode the polymers and, after reconstruction of the sequence from MS/MS data collected at a high collision energy, enabled the recipient to decode the secret message.

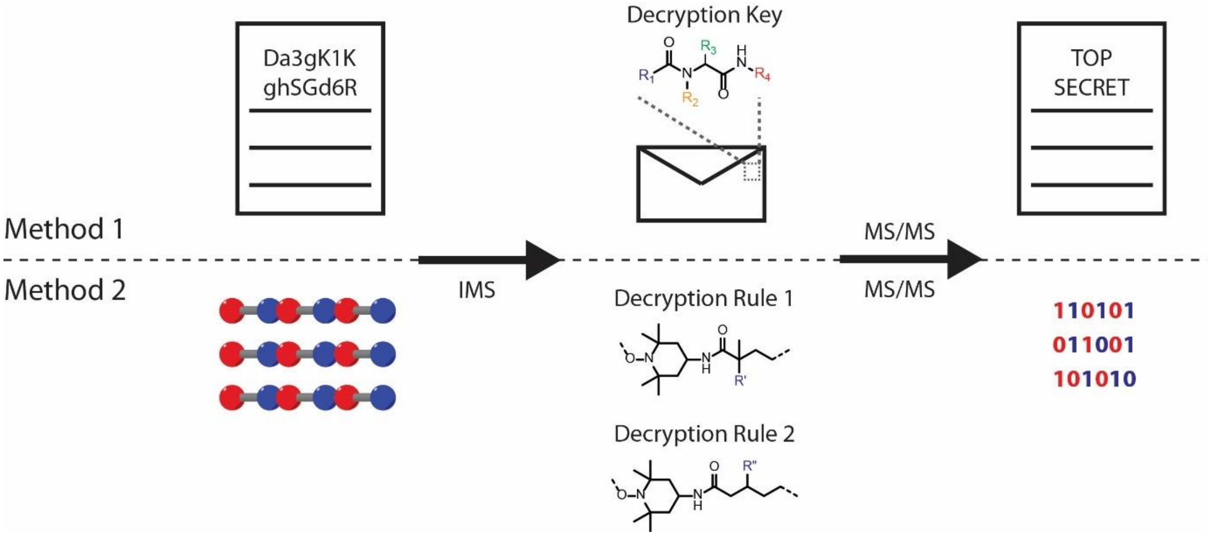


Figure 1.18 Abiotic macromolecule espionage. In method 1, a four-component molecular key is deposited on an envelope delivering an encoded message[121]. The molecular key is isolated and characterized, revealing the decryption rules for the encoded message. Method 2 relies on encoded the message in sequence-defined macromolecules. Decryption rules can be elucidated from individual strands by ion mobility spectrometry enabling the decoding of the binary data gathered from MS/MS sequencing[122].

Adding another level of security to encrypted polymeric-messages, Lutz and coworkers recently developed photo-editable poly(phosphodiester)s where light triggered sequence mutations enabled the photo-erasing or photo-revealing of encoded-information (Figure 1.19)[123]. The use of monomers bearing either *o*-nitrobenzyl or *o*-nitroveratryl pendant groups afforded information encoding by mass differentiation, however, upon UV irradiation, both of these groups are cleaved affording indistinguishable hydroxyl groups. Conversely, polymers employing *o*-nitrobenzyl and *p*-nitrobenzyl cannot be decrypted by MS/MS, until irradiated with light, in which case the *o*-nitrobenzyl is exclusively cleaved, elucidating a mass difference in monomers.

The two requirements necessary for information storage, (i) a writing mechanism and (ii) a sequence-reading mechanism, also remain the greatest challenges for the widespread implementation of this technology. Plagued by slow writing mechanisms, sequence-defined macromolecules have been largely limited to short oligomeric sequences. With the advancement of synthesis techniques and new chemistries, such as those by Zhu and coworkers[115], the writing speed will continuously improve – potentially exceeding the rates of DNA strand synthesis. Unfortunately, little progress has been made towards decoding technologies. While DNA strands can be sequenced by a variety of methods (e.g., nanopore sequencing, sequencing-by-synthesis, etc.[99]) synthetic encoded polymers rely on destructive methods such as MS/MS sequencing. This thesis will demonstrate a non-destructive methodology in which abiotic sequences bearing reactive binding pairs selectively assemble with complementary sequences, thereby affording MS identification by hybridizing unknown information-encoded oligomeric species with a library of known mass-tagged sequences[124]. This technique, and the concept of information-encoded hybridizations is explored in the following section.

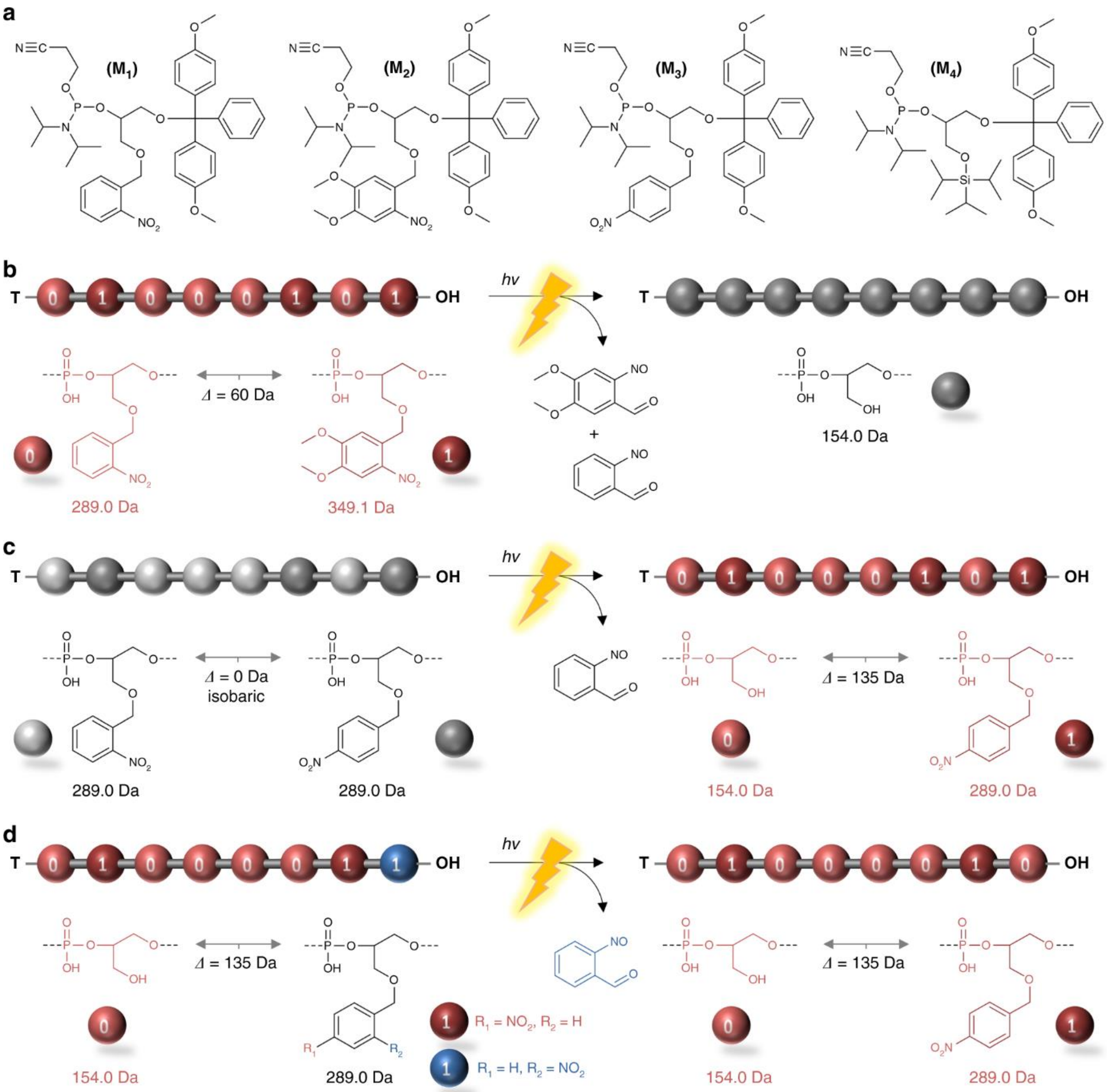

a
(M₁)
(M₂)
(M₃)
(M₄)
N≡C
NO₂
O₂N
Si
b
T
0 1 0 0 0 1 0 1
OH
hv
Δ = 60 Da
289.0 Da
349.1 Da
NO
154.0 Da
c
Δ = 0 Da
isobaric
289.0 Da
289.0 Da
Δ = 135 Da
154.0 Da
289.0 Da
d
0 1 0 0 0 0 1 1
0 1 0 0 0 0 1 0
Δ = 135 Da
154.0 Da
289.0 Da
R₁ = NO₂, R₂ = H
R₁ = H, R₂ = NO₂
Δ = 135 Da
154.0 Da
289.0 Da

Figure 1.19 Photo-editable poly(phosphodiester)s where light triggered sequence mutations enabled the photo-erasing or photo-revealing of encoded-information[123].

1.5.2. Self-Assembly of Ladder Species

The antiparallel duplex formed from two complementary strands of DNA, a fundamental molecular structure, has enabled the sophisticated self-assembly of nanoparticles with conceivable applications in drug delivery, molecular computing and enzyme immobilization[11,85]. Researchers have long put considerable effort into synthetically achieving this level of structural control to design elegant and complex supramolecular structures, beginning with the formation of molecular ladders or multiple-stranded foldamers. There exists several in-depth reviews for molecular duplexes stabilized by hydrogen-bonding (Figure 1.20a)[125–129], typically from heterocyclic amide- or urea-linked monomers, and metal coordination (Figure 1.20b)[125,130] .

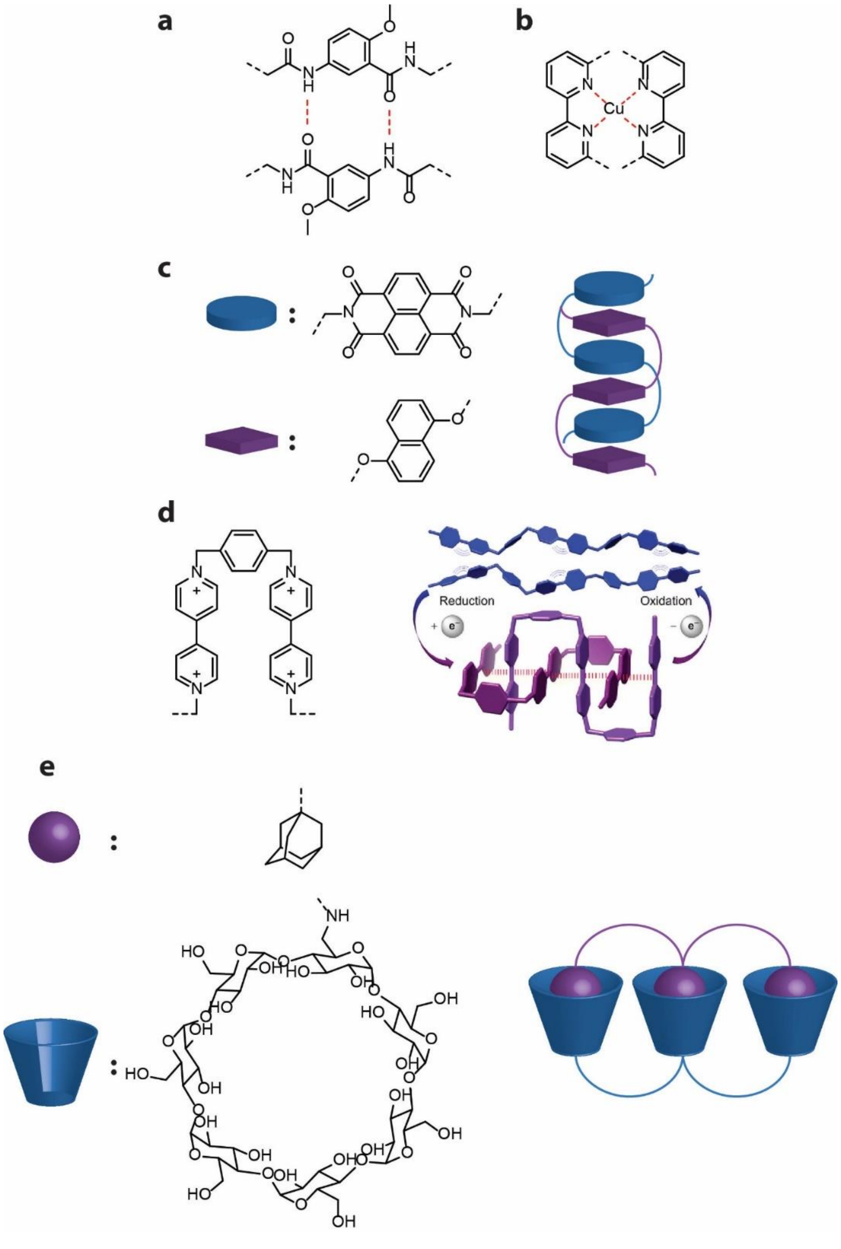


Figure 1.20 Noncovalent linkage mechanisms used to form abiotic molecular ladder; (a) hydrogen bonding, (b) metal coordination, (c) π-stacking, (d) conjugated radical cations, and (e) multivalent molecular recognition.

In an attempt to stabilize metal-free duplexes, scientists investigated the synthesis of polymeric strands capable of forming intermolecular aromatic-aromatic interactions (Figure 1.20c)[127,131]. While hydrogen-bonding interactions remained the primary driving force for early attempts at aromatic-aromatic stabilized complexation, in 2002, Gabriel and Iverson demonstrated an abiotic duplex relying extensively on aromatic interactions that could assemble in aqueous environments[132]. Iverson's group had previously demonstrated the π-stacking of a foldamer with incorporated electron rich 1,5-dialkoxy-napthalene (Dan) and electron-deficient 1,4,5,8-napthalene-tetracarboxylic diimide (Ndi) substituents[133]. By synthesizing two complementary homo-oligomers of either repeating Dan or Ndi residues, Gabriel and Iverson created stable duplexes of up to four repeating units with an association constant of 350,000 M^{-1} – three orders of magnitude larger than the binding of a single unit[132]. This work was soon followed by Goto et al. who synthesized oligoresorcinols which produced random coils in common organic solvents, but aggregated in water to form a double helical molecular ladder assisted by parallel stacking of the aromatic protons[134]. Duplexes consisting of 11-mer oligomers were demonstrated. Interestingly, incorporation of chiral moieties at both ends of the oligoresorcinols generated a one-handed screw. These early discoveries lead to a wide variety of abiotic oligomers capable of duplex formation including ethynylhelicene oligomers[135,136], quinoline oligoamides[137–139], and naphthyridine oligoamides[140].

Another promising mechanism, molecular ladders generated from conjugated radical anions or cations has gained interest in recent years (Figure 1.20d)[141]. For example, Wang et al. incorporated radical cationic forms of 4,4'-bipyridinium ($BIPY^{2+}$) into oligomers by linking the $BIPY^{2+}$ units with *p*-xylylene bridges to form oligoviologens[142]. These oligoviologens duplexes, demonstrated with up to 12 repeat units, are induced by radical-radical interactions. Furthermore, the assemblies are redox active, whereby reduction of concentrated oligoviologens induces duplex formation, while oxidation triggers an extension of the chains, and thus dissociated strands.

Kurlemann and Ravoo briefly examined sequence-selective molecular ladder formation by multivalent molecular recognition, employing cyclodextrin (CD)-based strands and strands modified with guest molecules (Figure 1.20e)[143]. Complementary binding pairs *n*-butyl : α-CD and adamantine : β-CD were selected as the monomeric species exhibited preferential binding to their complementary motif. However, upon incorporation of these motifs into trimers, the n-butyl pendant group provided insufficient binding discrimination between α-CD and β-CD resulting in the formation of higher-order multimeric species; in contrast, adamantine remained selective. Surprisingly, no dimeric sequence resulted in the generation of in-registry molecular ladder formation, while only complementary homo-oligomers of adamantane and β-CD formed in-registry duplexes in a trimeric system.

While most molecular ladders rely upon hydrogen-bonding and metal coordination to form duplexes between single stranded oligomers, there has been a more recent focus to incorporate stronger covalent bonds in hybridized structures by utilizing dynamic covalent chemistry (DCC) functional groups. DCC refers to a set of reactions where manipulating specific reaction conditions can precisely control the equilibrium; the transient nature of these bonds provides a mechanism for error correction necessary for self-assembly. The Moore group was the first to fabricate molecular ladders by incorporating DCC. Specifically, they employed amine and aldehyde functional groups into complementary *m*-phenylene ethynylene oligomers that underwent scandium (III)-catalyzed imine formation and exchange reactions to form n-rung molecular ladders (Figure 1.21a)[144]; however, they were unable to form longer duplexes in suitable yields owing to the kinetic trapping experienced at four or more rungs[145]. Wei et al. were able to form molecular ladders of lengths up to 16 rungs through the hybridization of sequence-defined peptoid-based oligomers designed to undergo an amine/aldehyde condensation reaction[5]. The duplexes were able to overcome the kinetic limitation of the *m*-phenylene ethynylene ladders owing to the flexibility of the peptoid backbone. Through MALDI-TOF kinetic analysis and Förster resonance energy transfer (FRET) analysis techniques, Wei et al. determined the ladder formation mechanism to follow an initial 'zipping up' primary mechanism to rapidly fabricate out-of-registry ladders, proceeded by a secondary, rate-limiting shuffling mechanism *via* transamination and imine metathesis of imine bonds - generating in-registry molecular ladders. Furthermore, by employing non-commensurate complementary aldehyde- and amine-functionalized oligomer strands, large molecular ladders constituting several short strands could be formed from the lowest common multiple of functionalities of the oligomers (i.e., trialdehyde/tetraamine strands form a 12-rung ladder), a process known as Vernier templating (Figure 1.22)[146]. In addition, the Anslyn group examined the enhanced antimicrobial activity of peptide-based molecular ladders bearing hydrazone-rungs[147] and the Scott group demonstrated the rapid formation of molecular ladders from amine/aldehyde bearing β-peptoids[148].

The relative synthetic accessibility and stability of Schiff base reactants as DCC functionalities has promoted its primary usage in covalent molecular ladders. Indeed, in 2017 the Yashima group was the first to demonstrate a new class of covalent molecular ladders bearing boronate-esters[149]. Interestingly, the di-functionalized *m*-terphenyl oligomers, either bearing diethyl boronates or diols, adopted a helical conformation upon hybridization. Recently, Dunn et al. demonstrated the formation of ladders up to six-rungs utilizing boronic-acid- and catechol-bearing peptoid strands in an aqueous environment (Figure 1.21b)[150]. Moreover, by exploiting the peptoid geometry in which functional groups on adjacent residue are presented on opposite sides of the backbone, Dunn et al. were able to generate molecular grids where two catechol flanking strands covalently-bonded to a peptoid center-strand bearing boronic-acid moieties on both edges. Moreover, peptides and conjugated oligo(phenylene ethynylene) derivatives have also been shown to form molecular ladders with boronate-ester-rungs[151,152].

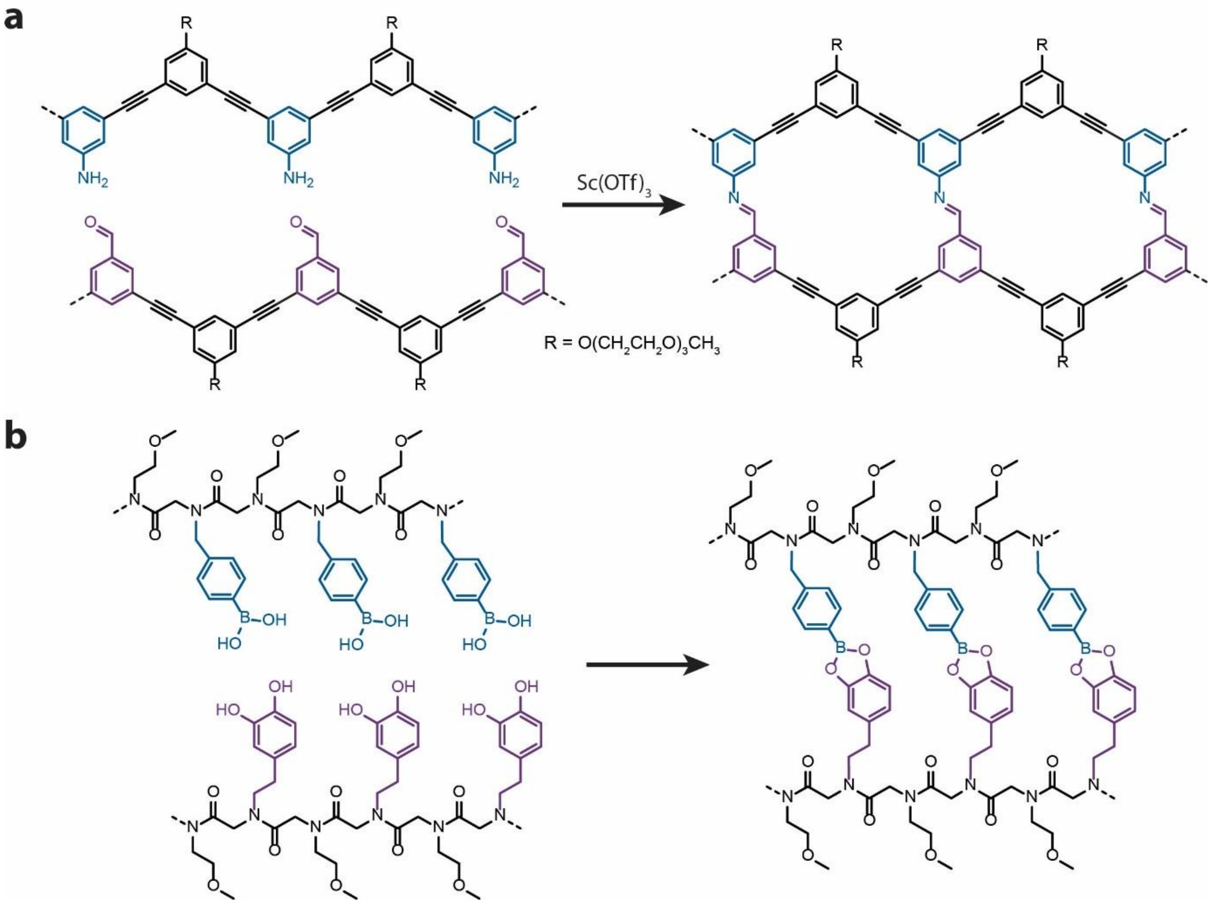


Figure 1.21 Reported approaches to forming covalently-linked molecular ladders bearing (a) imine or (b) boronate ester linkages.

The production of intricate nanostructures from a self-assembly process will require the use of information-bearing polymer strands to undergo sequence-selective hybridization. While there have been several successful demonstrations of abiotic sequence-selectivity by synthetic polymers utilizing hydrogen-bonding[153] and metal coordination[154,155], up until recently, molecular ladders bearing novel molecular recognition systems have consisted of complementary homo- or di-block oligomer pairings[156]. This thesis will report the one-pot sequence-selective covalent hybridization of a library of information-bearing oligomeric species by using a Lewis acidic catalyst to affect the amine/aldehyde condensation reaction equilibrium to dissociate and subsequently catalyze bond rearrangement upon extraction, analogous to the thermal dissociation/annealing process employed in DNA origami. The developed Lewis acid dissociation/extraction/annealing process mitigates the kinetic trapping of covalent molecular ladders first observed by Moore's group[145] and has the capacity to catalyze the selective assembly of molecular ladders even in environments containing excess of competitive binding species[124]. Subsequently, encoded molecular ladders and grids with both boronate-ester and imine dynamic covalent reactant pairs will be employed to illustrate base-4 covalent assembly[157].

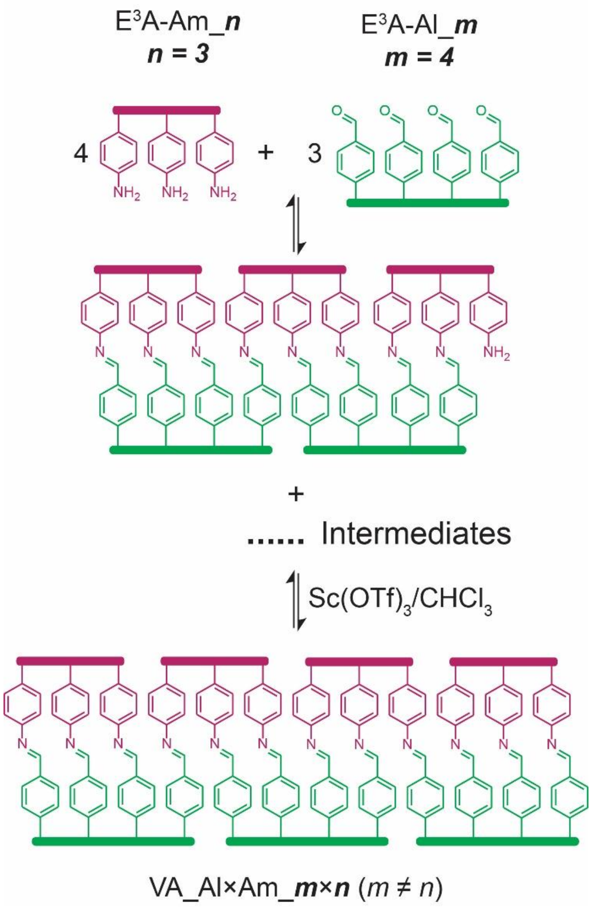


Figure 1.22 Schematic representation of an imine-bearing molecular ladder formed by Vernier templating[146].

1.5.3. Replication of Abiotic Polymers

In 1995, in an effort to throw new light on the origins of life, Leslie Orgel presented what he defined as the Holy Grail of modern organic chemistry: "the development of nonenzymatic reactions in which replication and information transfer occur together without the help of protein enzymes"[158]. While no 'pure knight' has successfully accomplished the replication of synthetic, sequence-defined polymer systems to the degree of DNA replication, significant effort has been made to advance this technology.

Early attempts at synthetic replication consisted of "minimal replicating systems," or the autocatalytic synthesis and amplification of small molecules[159,160]. A primary example of autocatalyzed self-replication is Terfort and Kiedrowski's work on the condensation of 3-aminobenzamadines and 2-formylphenoxyacetic acids (Figure 1.23)[161]. Here, a previously condensed anil preorganizes two unreacted monomers by way of amidinium-carboxylate salt bridges - catalysing the condensation reaction between the two monomers. Ideally, as was the case in this example, the salt bridge is less stable upon formation of the template molecule, facilitating the dissociation of the newly formed molecule and the template, both of which complex with fresh monomer inducing exponential autocatalysis.

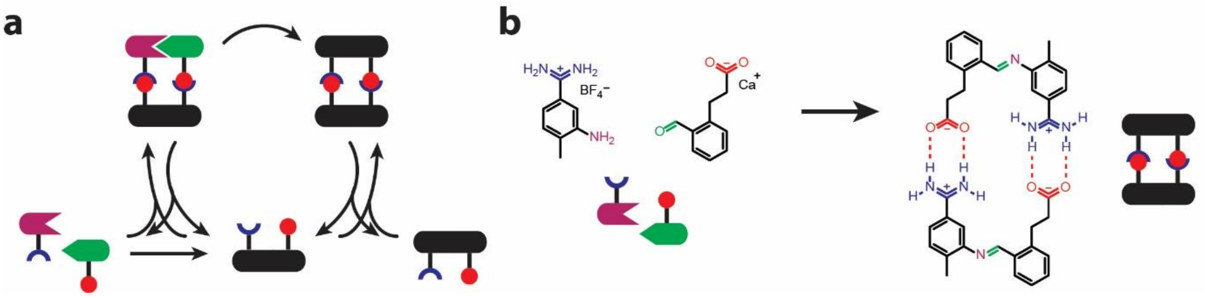


Figure 1.23 Autocatalyzed self-replication (a) mechanism schematic and (b) application employing a salt bridge medium[159,160].

Autocatalytic synthesis can also be applied to larger systems, as was demonstrated by Carnall et al.[162]. In this work, macrocycles were produced from peptide sequences alternating hydrophobic (leucine) and hydrophilic (lysine) residues terminated with dithiol groups capable of forming disulphide linkages. In solution, the macrocycles formed trimer and tetramer species that were stable over a four day period, but subsequently formed heptamer species that were stable for 7.5 months. Interestingly, shaking rather than stirring the reaction mixture produced hexamer macrocycles. To investigate the autocatalytic properties of the system, reaction mixtures were seeded after four days with hexamer or heptamer species resulting in increased rates of hexamer/heptamer formation. Cryogenic transmission electron microscopy (cryo-TEM) images of the reaction solutions revealed long thin fibers, up to 1 or 2 μm in length, indicating a self-organization event that facilitated the autocatalysis of macrocycle formation (Figure 1.24).

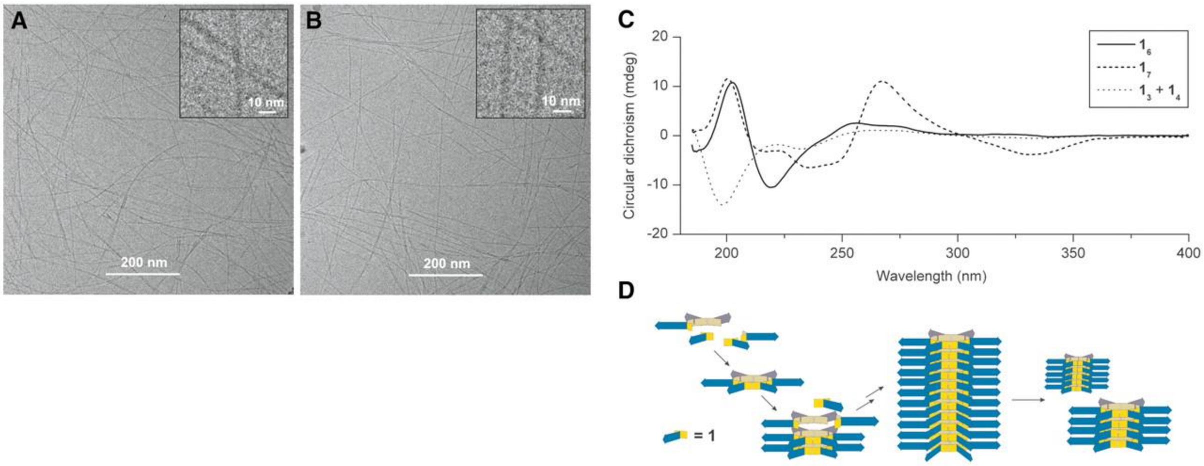


Figure 1.24 (a and b) Cryo-TEM images, (c) circular dichroism spectra, and (d) schematic of the macrocycles formed from Canrall's work[162].

A significant contribution to the field of synthetic polymer replication came from the Liu group's work of 2013 in which sequence-defined polymers of about 26 kDa were synthesized employing a DNA template without the use of an enzyme (Figure 1.25)[163]. In this work, a peptide nucleic acid (PNA) pentamer adapter was attached to a synthetic polymer building block by disulphide linkers which could be selectively cleaved using dithiothreitol. The polymer building block was functionalized with terminal groups capable of coupling together to allow for the formation of the larger polymer sequence. Several coupling reactions [e.g. amine acylation, reductive amination, oxime and hydrazone formation, and Cu(I)-catalyzed alkyne-azide cycloaddition (CuAAC)] and architectures ('AB' or 'AA/BB') were examined to determine optimal coupling technique, with the CuAAC coupling reaction and an AA/BB architecture providing the ideal coupling strategy. As the PNA adapters were sequence-selective towards the DNA template codon, information could be transcribed into the synthetic polymer by varying the polymer building blocks attached to the PNA adapters. Ultimately the coupling of 16 building blocks, or 90 side-chain functionalized β-amino acid residues, was demonstrated.

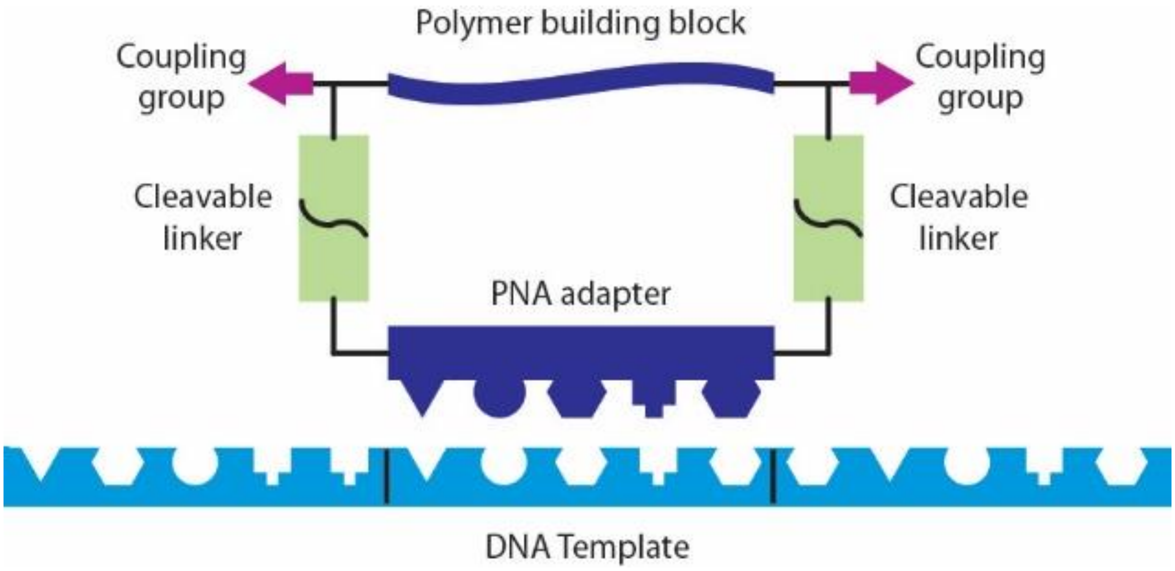


Figure 1.25 Synthetic, sequence-defined polymer replication using a DNA template[163].

Here, it is important to note the technological advancements provided by individuals working in the area of ladderphanes whose efforts on the templated synthesis of polymers with low polydispersities (PDI), although not discrete nor sequence-defined, mark early attempts at replication of larger polymeric species from abiotic templates. In 2007, South and Weck examined the template-enhanced ring-opening metathesis (TROM) polymerization of norbornene-based monomers (Figure 1.26a)[164]. Template polymer strands of 20 replicating residues were synthesized in solution bearing diaminopyridine functionalities (PDI = 1.30) and coupled to a polystyrene resin allowing for facile separation of daughter strand and repeated template use. Thymine bearing norbornene monomers were complexed, through hydrogen-bonding, with the diaminopyridine functionalities of the template strands and subjected to ROMP, resulting in daughter strands of 20 repeat units and a polydispersity of 1.19. Similarly, Sleiman and coworkers used a thymine diblock copolymer template, synthesized by ROMP with a PDI of 1.07, to complex with adenine-bearing monomers able to undergo Sonogashira coupling to form daughter strands[165]. Without templating, the adenine-bearing daughter strands culminated in a PDI of 2.77 with an average degree of polymerization of 9.8; templated strands showed a significantly lower PDI of 1.2, comparable to the template strand PDI, and 25.3 repeat units on average. Continued advancements in both ROMP catalysts and monomers have enabled the formation of polymers with even lower polydispersities and improved templating capabilities[166,167].

To achieve large, uniform polymeric species from templated synthesis, without the challenges associated with the synthesis of large discrete templates, the Anderson group investigated the use of Vernier templating to enable the synthesis of large π-conjugated macrocycles from small template subsets (Figure 1.27)[168–170]. To this end, a 12-porphyrin nano-ring, the largest π-conjugated macrocycle published at the time, was synthesized from the palladium-catalyzed oxidative coupling of three poryphyrin tetramers templated *via* metal-coordination to two hexapyridyl templates (T6) with a 39% yield[168]. Additionally, the 12-porphyrin ring was synthesized using classical methods with a dodecapyridyl template (T12) and three tetramers (35% yield), and although the yields were similar, T12 required a low-yielding ten-step synthesis, while T6 was achieved after only two steps from commercially available materials. Moreover, by employing both a greater number and larger pyridyl templates, the Anderson group was able to expand the nano-ring to include 24[169] and then 30 and 40 porphyrin units[170]. The group later templated linear porphyrin oligomers of up to twelve residues with a maximum yield of 38%[171].

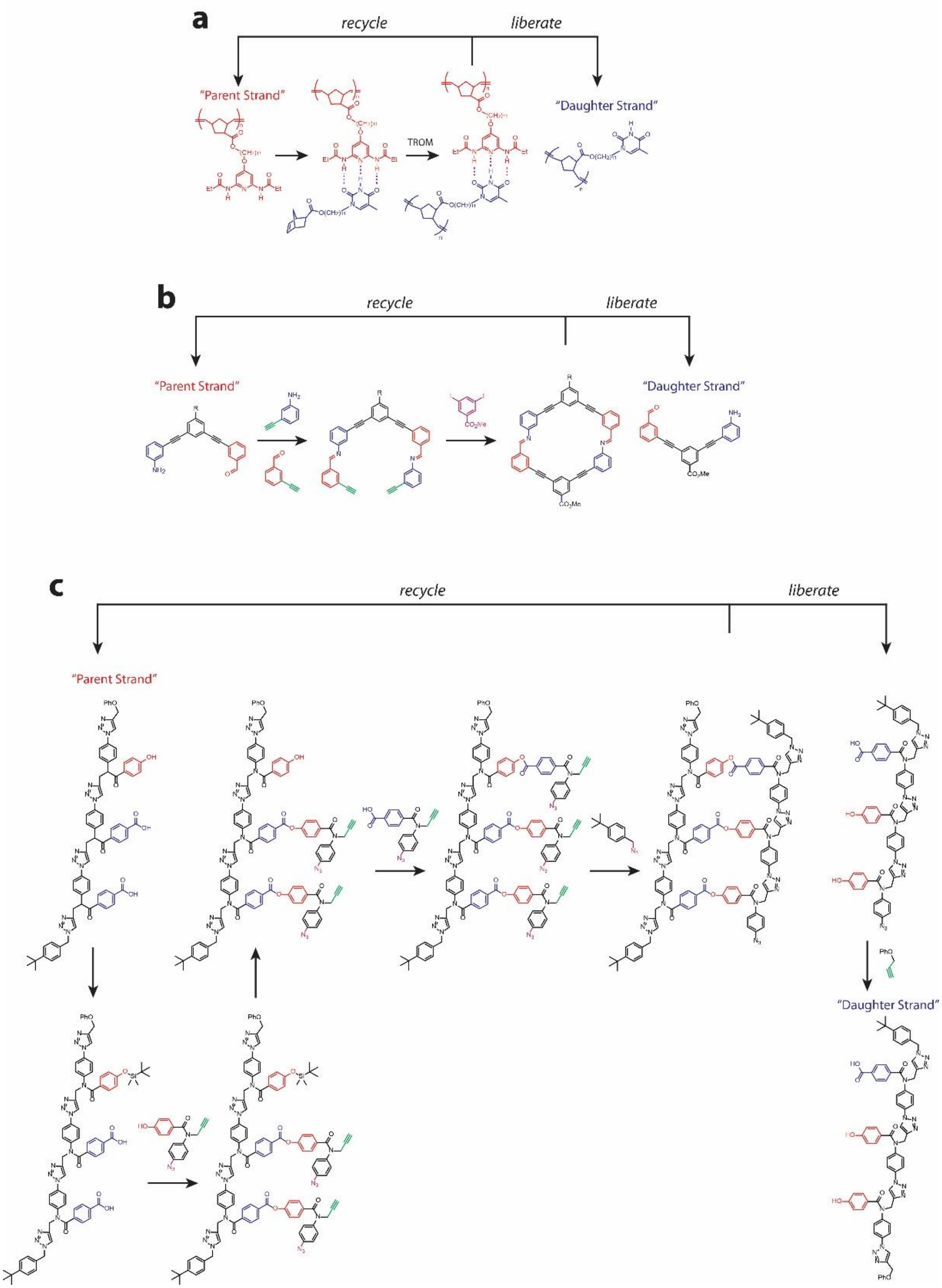

a
recycle
liberate
"Parent Strand"
TROM
"Daughter Strand"
b
recycle
liberate
"Parent Strand"
"Daughter Strand"
c
recycle
liberate
"Parent Strand"
"Daughter Strand"

Figure 1.26 Abiotic polymer replications using parent/daughter templating techniques in which the parent strands can be reused. Synthetic pathways employing (a) salt bridge linkages and template-enhanced ring-opening metathesis polymerization[164], (b) imine chemistry and Sonogashira coupling[172], and (c) ester formation with CuACC[173]..

In the last year, two templating techniques applying covalent interactions to complex parent and daughter strands have emerged resulting in the templating of sequence-defined oligomers. First, Strom et al. used the amine/aldehyde condensation reaction to bind monomers to a parent dimer before coupling with the Sonogashira reaction[172]. Template dimers (e.g. diamine, dialdehyde, or amine-aldehyde dimers) were initially synthesized and complexed with an excess of the complementary monomers to promote imine formation, unreacted monomers were removed by vacuum sublimation overnight, and finally, the monomers bound to the parent strands were linked *via* Sonogashira coupling to template either homo- and di-functionalized dimers. Shortly after, Hunter and coworkers templated sequence-defined trimers employing a phenol monomer (P) and a benzoic acid monomer (A), both functionalized with an alkyne and azide for CuAAC[174]. To demonstrate the system, the trimer template AAP was synthesized by an iterative CuAAC/deprotection strategy, followed by the selective protection of the phenol pendant group to prevent intramolecular folding of the parent strand. Free phenol monomers were then coupled with the benzoic acid pendant groups to form esters, permitting the deprotection of the phenol pendant group, and subsequent ester formation with free benzoic acid monomers. The monomers were zipped together using CuAAC and the resulting ladder was dissociated by cleaving the ester base-pairs. Although the antiparallel sequence-complementary daughter strand, APP, was the major product (72%), there was a significant amount of the scrambled sequence PAP (17%). Unfortunately, the use of asymmetric monomers with a directional zipping mechanism limits the system to trimer replication as larger systems introduce the trapping of species by the coupling of monomeric species in opposite directions. By combining the sequence-selectivity of the monomers in this work with the non-directional zipping of replication in the area of template enhanced ROMP, the replication and amplification of longer sequence-defined polymeric materials can be achieved.

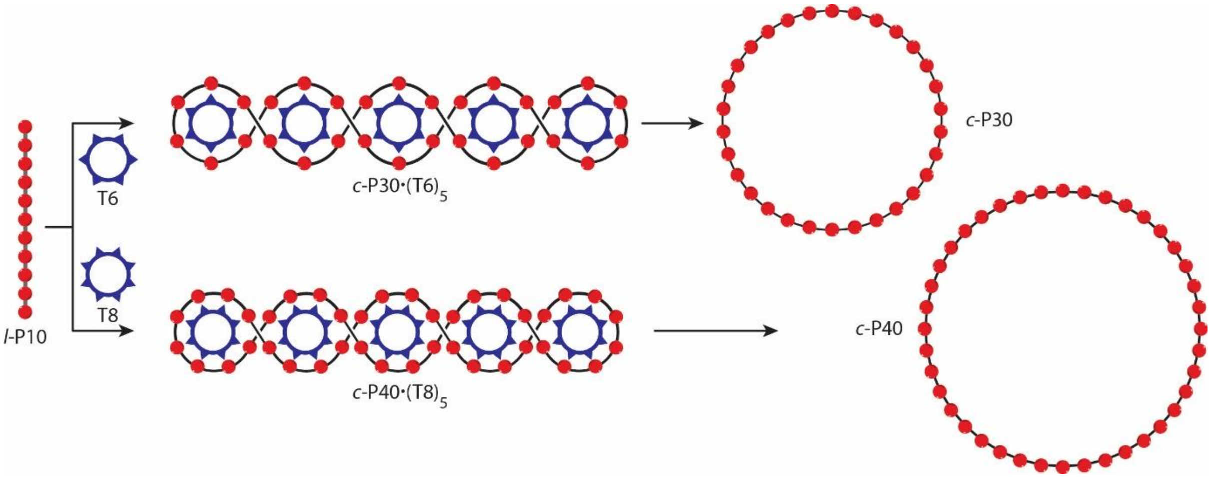


Figure 1.27 Synthesis of nanorings by way of Vernier templating[171].

1.6. Dynamic Covalent Assembly

The ultimate goal of this project is to provide a mechanism and pathway forward for the generation of intricate nanomolecular constructs by way of dynamic covalent interactions. While the self-assembly of molecular ladders from dynamic covalent interactions has been mentioned in previous sections, in this section we will look more generally into dynamic covalent self-assembly of architectures, often in three dimensions, albeit without the benefits and complexity of information-directed construction. Thus, in the following sections, dynamic covalent interactions will be discussed as they pertain to micelle,

1.6.1. Dynamic Covalent Micelles

With nanometer sized ranges and an ability to encapsulate hydrophobic therapeutic agents, micelles have long shown to be useful molecules in drug delivery systems[175]. Polymeric micelles are self-assembled from the aggregation of amphiphilic block copolymers where one block contains a hydrophilic group that forms the outer shell of the micelle in an aqueous environment and another block forms the hydrophobic core of the micelle. The simple nature of micelle formation allows for the employment of a variety of chemistries and modifications to tailor micelles to their intended applications[176]. Moreover, by employing reversible chemistries, micelle formation and dissociation can be controlled. For example, Minkenberg et al. converted a nonamphiphilic surfactant comprised of a polar head group with an aromatic aldehyde into an amphiphile by exposing the surfactant to an aliphatic amine that upon reversible imine bond formation between the aldehyde and amine left an apolar tail[177]. These species would then self-assemble into pH-controlled micelles as shown in Figure 1.28. This allowed them to control the micelle formation by manipulating the pH, and readily demonstrated the potential utility of this technology to be used in drug delivery systems by controlling the uptake and release of Nile Red, a hydrophobic organic dye.

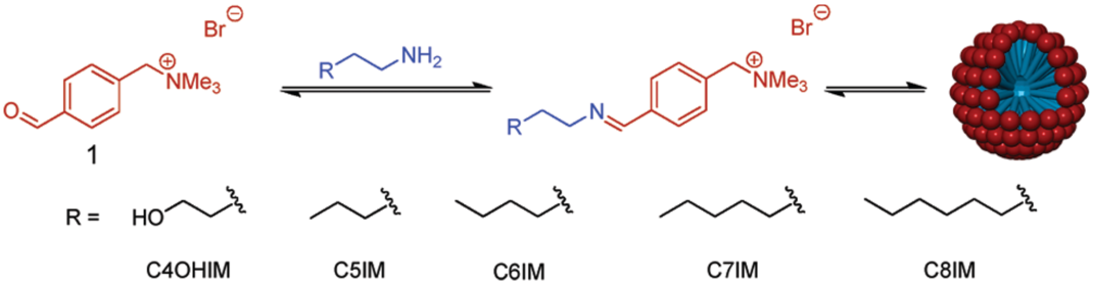


Figure 1.28 Controlled formation of micellar aggregates from the aromatic aldehyde polar head group reversibly reacting with various amines to form an amphiphilic surfactant that can self-assemble into micelles.

Further building upon this work, Minkenberg et al. combined two water soluble surfactants containing either aldehyde or amine functional groups to generate vesicles composed of reversible imine bonds[178]. Interestingly, these vesicles would cluster to form vesicle networks leading to pH-reversible gels shown in Figure 1.29. While not demonstrated by Minkenberg et al., such assemblies could potentially afford self-healing materials, mechanisms for drug delivery systems.

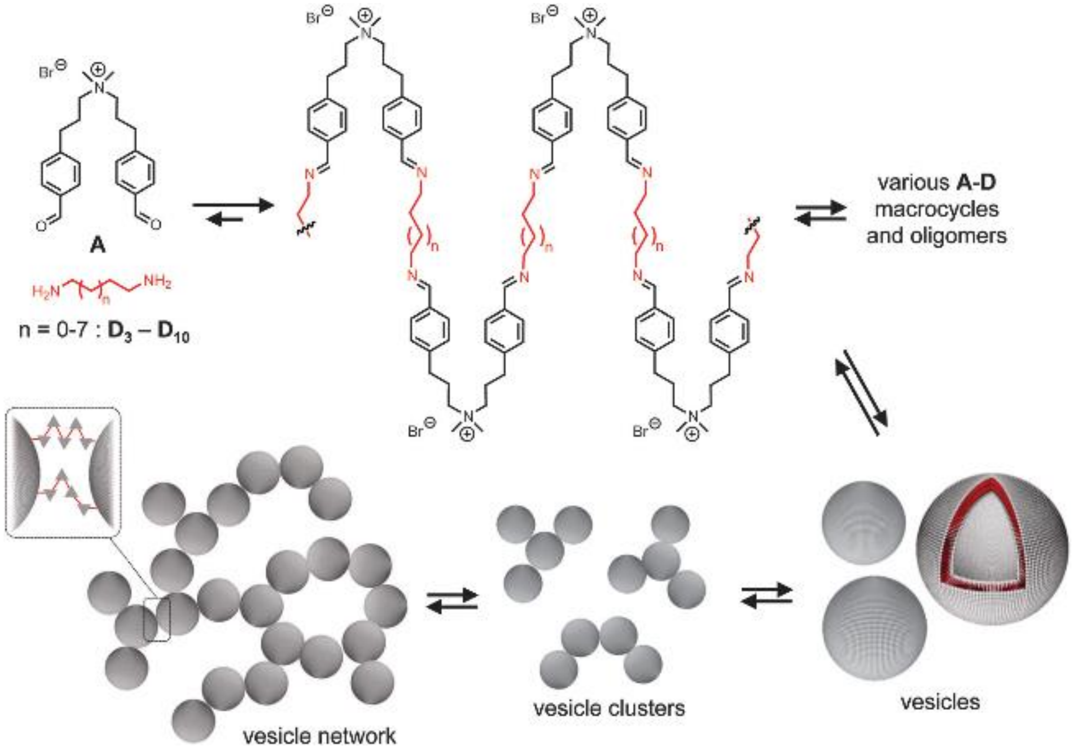


Figure 1.29 Schematic showing the formation of imine bonds from difunctionalized aldehydes and aliphatic amines to form macrocycles and oligomers leading to the vesicles, vesicle clusters, and finally vesicle networks capable of forming pH-reversible hydrogels.

Ever the visionary, Minkenberg et al. continued this work with imine micelles to fabricate Gemini surfactants, characterized by having multiple amphiphilic functional groups on each short strand, by utilizing aliphatic amines of various lengths and a bisaldehyde-functionalized quaternary ammonium head group[179]. Then, the researchers applied the Gemini surfactants to make wormlike micelles as shown in Figure 1.30, yielding the first evidence of reversible aggregation of Gemini surfactants to form wormlike micelles. Gemini surfactants have shown great potential in both biological applications such as gene transfection and industrial applications as viscoelastic additives for cosmetics.

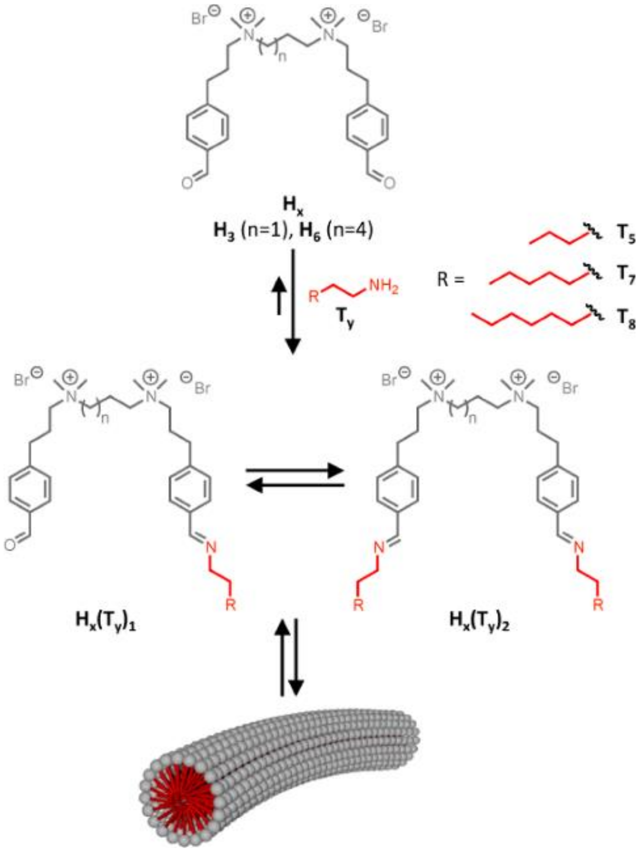


Figure 1.30 The formation of wormlike micelles from Gemini surfactants comprised of a bisaldehyde head group and amine tail groups[179].

Other groups have demonstrated the formation of micelles with chemistries other than imine formation, such as the use of disulfide chemistry by Li et al. to fabricate responsive cross-linked micelles[180]. These micelles could be loaded with the therapeutic, paclitaxel(PTX), for controlled release on tumor sites as shown in Figure 1.31. By introducing cysteine into linear-dendritic system of polyethylene glycol(PEG) and cholic acid, the thiols could be oxidized to form disulfide bonds in the core of the micelle. This cross-linked micelle had increased stability than non-cross-linked micelles; however, the cross-linked micelles had a slower release of PTX. Release of the PTX was controlled by altering the amount of cellular glutathione. In animal studies, they found that the cross-linked micelles more efficiently collected and released at the tumor site of mice carrying the “SKOV-3 ovarian cancer xenograft.”

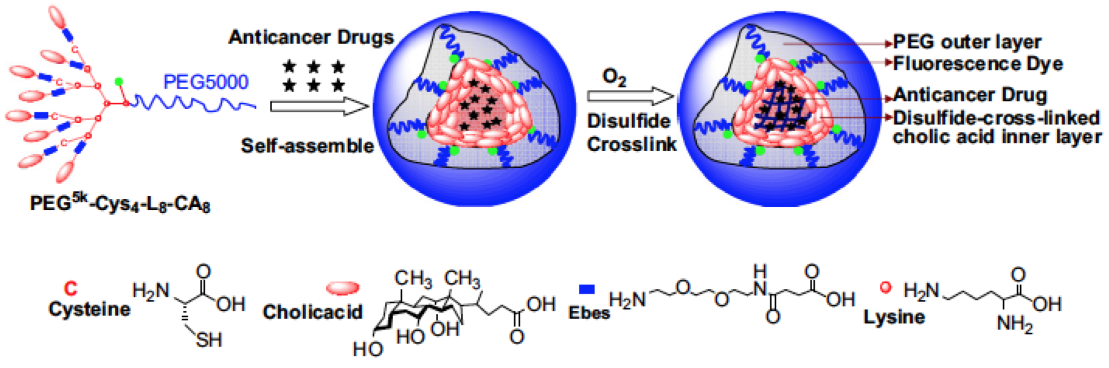

Anticancer Drugs
PEG5000
Self-assemble
PEG5k-Cys4-L8-CA8
O2
Disulfide
Crosslink
PEG outer layer
Fluorescence Dye
Anticancer Drug
Disulfide-cross-linked
cholic acid inner layer
C
Cysteine
Cholicacid
Ebes
Lysine

Figure 1.31 The formation of PTX loaded micelles through disulfide bond formation[180].

1.6.2. Dynamic Covalent Hydrogels

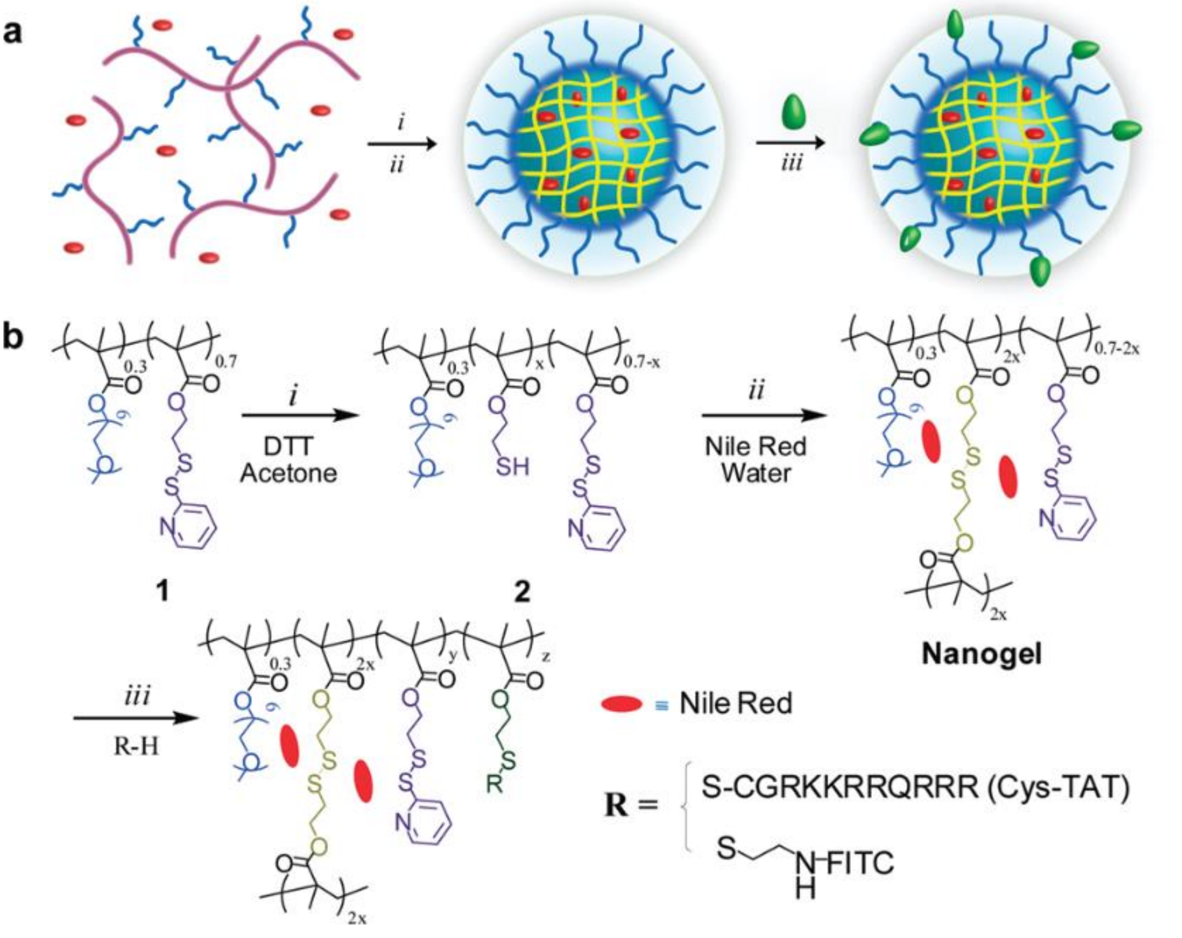

a
i
ii
iii
b
0.3
0.7
i
DTT
Acetone
x
0.7-x
SH
ii
Nile Red
Water
2x
0.7-2x
1
2
Nanogel
iii
R-H
y
z
= Nile Red
R =
S-CGRKKRRQRRR (Cys-TAT)
S
N-FITC
H

Figure 1.32 Dyanmic covalent nanogel a) schematic showing the formation of a nanogel from functionalized nanoparticles b) structures of the nanogel including the disulfide chemistry[181]

Stimuli-responsive, drug delivery vehicles are a common application for dynamic covalent interactions. As such, it is not surprising to see addition to hydrogels. Ryu et al. developed a nonemulsion method for fabrication of nanoparticles that were incorporated into nanogels; allowing for the encapsulation of hydrophobic groups and stimuli-responsive disulfide bonds to form inter/intrastrand cross-linking in the gel formation (Figure 1.32)[181]. By reducing the disulfide bonds, disassembling the nanogel, the body could release of the encapsulated hydrophobic therapeutic moiety.

Similarly, Fairbanks et al. was able to incorporate a lithium acylphosphinate photoinitiator(LAP) into a hydrogel formed from the oxidation of thiol-functionalized 4-armed poly(ethylene glycol)(PEG) to fabricate a photo-responsive hydrogel where LAP generated radicals could break the disulfide bonds that cross-linked the gel (Figure 1.33)[182]. By manipulating the photointiator concentration to control the number of radicals formed in the sample, they were able to demonstrate the self-healing abilities of the hydrogel. Interestingly, the gel would break up into PEG groups upon generation of a similar concentration of sulfur radicals to the carbon- and phosphorus radicals formed from the photoinitiator. However, ff there were considerably more sulfur radicals in the sample, disulfide exchange would occur allowing for self-healing.

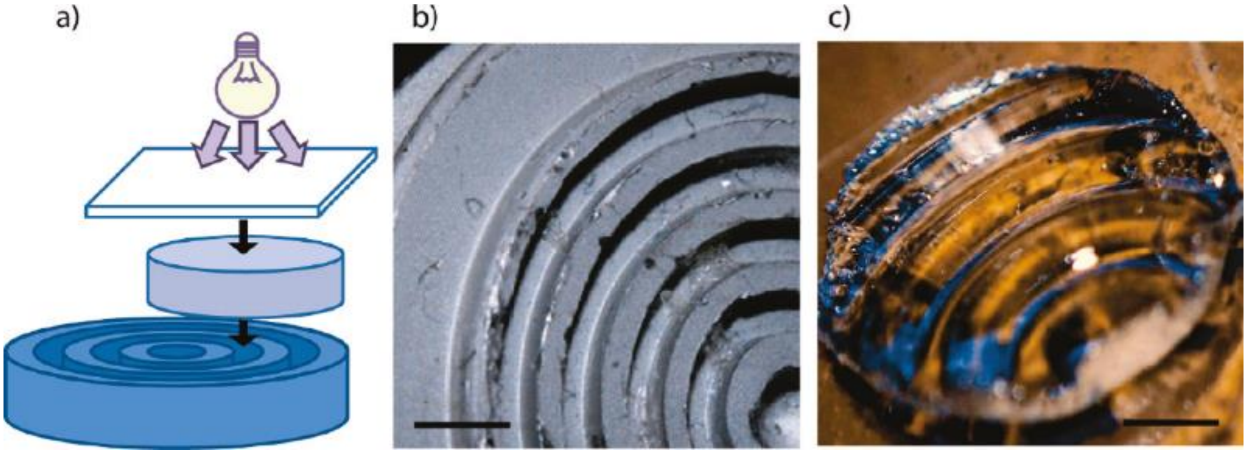


Figure 1.33 Dynamic covalent hydrogel a) Schematic showing the formation of a pattern into the photoadaptable hydrogel b) The patterned surface scale bar = 1.5mm c) The resulting patterned hydrogel scale bar = 1.5mm[182].

1.6.3. Covalent Organic Frameworks

Covalent organic frameworks (COFs) are two- and three-dimensional porous, crystalline network structures bonded by covalent bonds[183]. These networks are synthesized from a variety of well-defined, functionalized monomers to self-assemble into predictable structures[184]. Pore dimensions and size inside of the network are readily controlled and programmed by the length of the monomer linkages, enabling the tailoring of COFs to their applications. Typical applications include the storage and separation of gasses such as hydrogen, methane, and nitrogen[185]. As a developing chemistry and material, COFs are not commercially applied as of yet, but their facile modification and stability has interested a whole generation of scientists and holds a promising and bright future.

COFs were first developed by the Yhagi group at the University of Michigan (Go Blue!) in 2005 through the self-condensation of 1,4-benzendiboronic acid or co-condesation with hexahydroxytriphenylene[186]. Dynamic covalent interactions are used in the generation of COFs as their reversibility enables error-correction of the constituents to form highly ordered, crystalline structures. While boronic-acid condensation reactions were initially employed, COFs are currently synthesized with a variety of dynamic covalent reactions include imine and triazine chemistries[185]. Although a variety of chemistries exist which have the potential to mediate the formation of COFs, in practice, it is difficult to find reaction conditions which promote high degrees of crystallinity that are not detrimental to monomer stability. Harsh conditions are commonly employed for the synthesis of COFs, such as the formation of triazine-linked COFs at 400°C in molten $ZnCl_2$[187]. Similarly, imine linked COFs are often formed under solvothermal conditions in which relatively high concentrations of acetic acid are used at elevated temperatures[188]. Recently, an example of crystalline imine-linked COFs was demonstrated using monomeric aniline to promote rearrangement of the imine-linked network, however to form crystalline structures required over a month of self-assembly[185]. Other methods include the utilization of two orthogonal chemistries subsequently allowing for the self-assembly of constituents with less functionalization per constituent[189].

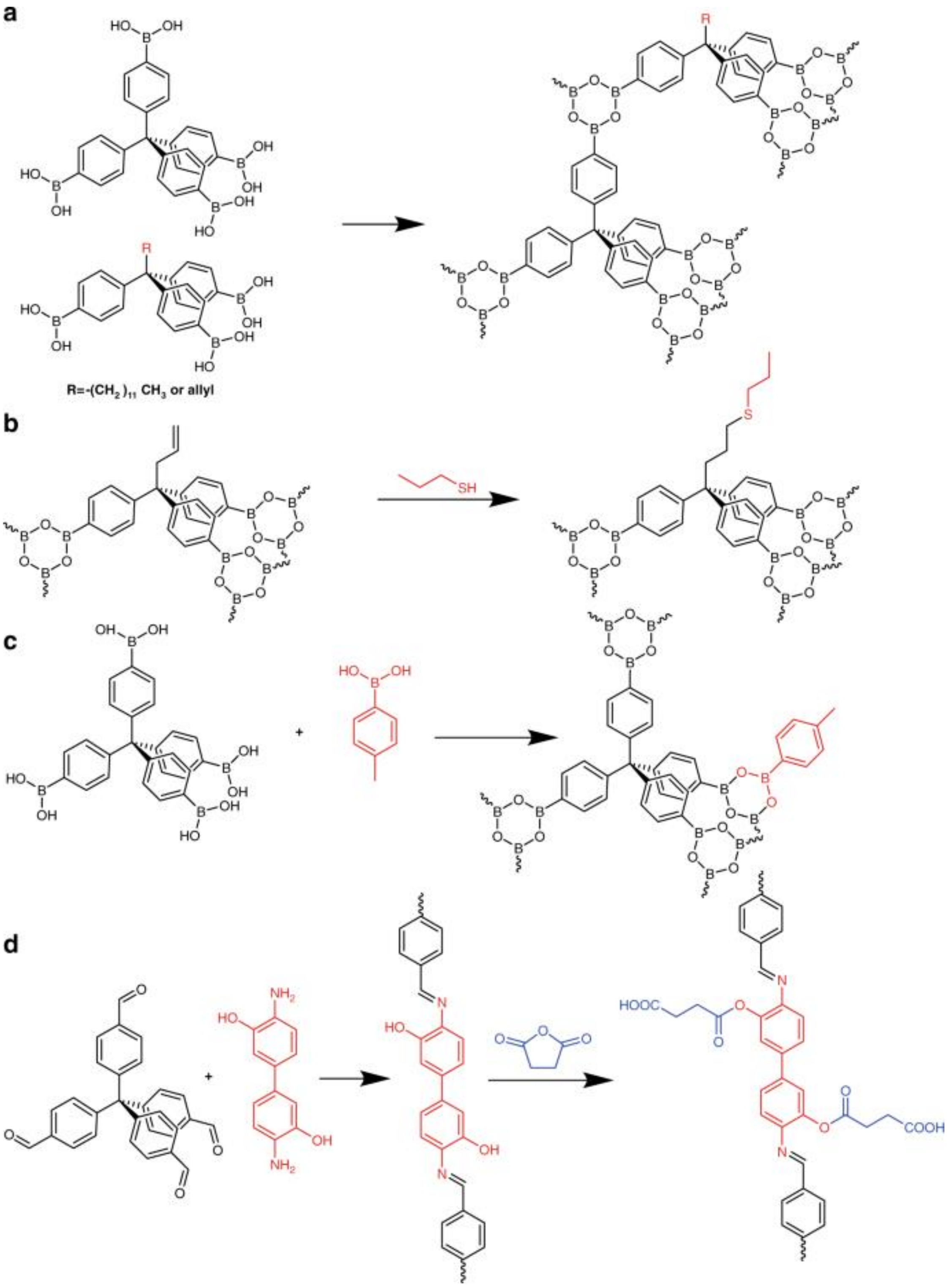

a
R=-(CH2)11 CH3 or allyl
b
SH
S
c
d
HOOC
COOH

Figure 1.34 Synthetic schemes for the functionalization of covalent organic frameworks; (a) COF-102 is modified by (b) thiol-ene or (c) boronic acid, chemistries. (d) Modification of 3D-OH-COF[185].

As synthetic techniques have advanced, COF material scientists have begun investigating the post-synthetic modification of COFs to further tune pore dimensions and selectivity of networks to external molecules (Figure 1.34). To enable post-synthetic modifications, linkers must be stable to post-synthetic conditions, reactions must be selective and form few byproducts, reactions must have high conversions as resultant structures cannot be purified, and the resulting structures must be characterizable, typically solid-state NMR, to confirm modification. These severe limitations have reduced the number of possible chemsitries available. The azide-alkynes click reaction is one of the most promising as it is orthogonal to most chemistries and yields a high rate of conversion[88]. Thiol-ene chemistries are also employed but are difficult due to the reactivity of both functional group precursors[185]. Hydroxyl-modification has also been employed successfully through a variety of chemistries including conversion to ethers. The locking or freezing of the dynamic covalent interactions post-synthesis may enable further modification and increase stabilities of COFs by preventing continuous adaptation and reaction of the network. For example, the imine dynamic covalent interaction can be reduced efficiently with the use of sodium borohydride or milder reducing agents to form inert secondary amines[185].

1.6.4. Dynamic Covalent Cages

Although, examples exist since 1991, covalent cages, defined as ‘polycyclic compounds with the shape of cages’ by IUPAC, are one of the most recent three dimensional structures to gain popularity through the use of dynamic covalent interactions (Figure 1.35)[190]. Mimicking the defined cavities found in nature, organic cages are typically produced with the intention of acting as receptor or sensors for guest molecules such as hydrocarbons or anions[191,192]. Other applications include stabilizing of guest-molecules or enabling the regioselectivity of uncommon reactions[6]. The most regularly occurring geometries found in molecular cages are tetrahedrons, cubes, octahedrons, dodecahedrons, and icosahedrons[190]. Conventionally designed through coordination of metal ions, recent advancements in dynamic covalent reactions has enabled the generation of new cage species (Figure 1.36)[6]. Cram et al. were the first to apply dynamic covalent interactions to the synthesis of covalent organic cages, having been synthesized from irreversible covalent bonds initially[193]. Given the relative simplicity of hollow organic cages as compared with large COFs, a larger variety of dynamic covalent chemistries have been employed for the formation of dynamic covalent cages; imine, boronate ester, alkene metathesis, and disulfide interactions are several examples[194]. Additionally, imine and boronate ester chemistries have been used simultaneously to form intricate molecular cages(Figure 1.37)[42].

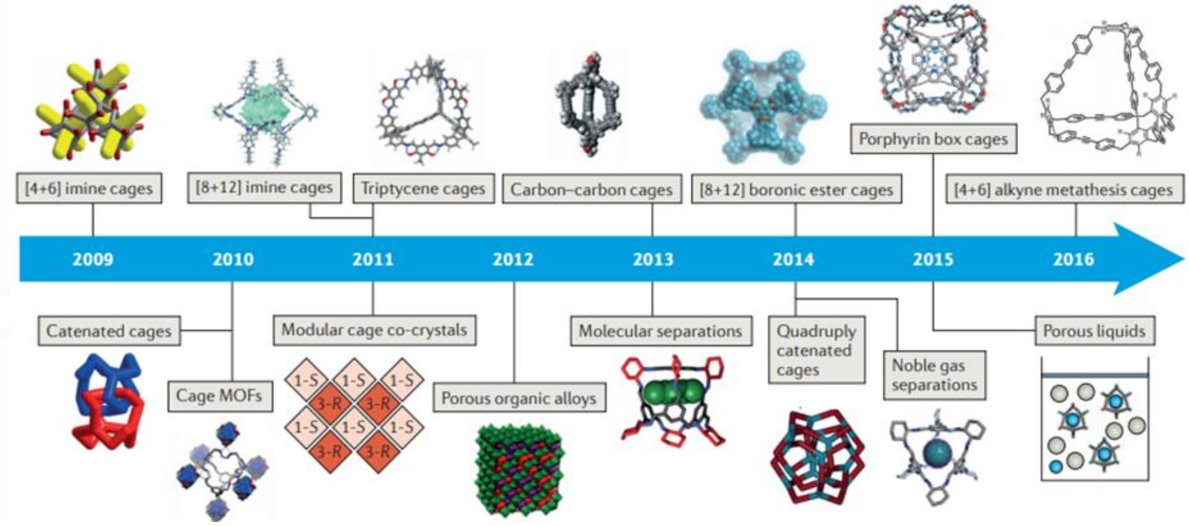

[4+6] imine cages
[8+12] imine cages
Triptycene cages
Carbon–carbon cages
[8+12] boronic ester cages
Porphyrin box cages
[4+6] alkyne metathesis cages
2009
2010
2011
2012
2013
2014
2015
2016
Catenated cages
Cage MOFs
Modular cage co-crystals
1-S
3-R
Porous organic alloys
Molecular separations
Quadruply catenated cages
Noble gas separations
Porous liquids

Figure 1.35 A timeline of porous organic cages[195].

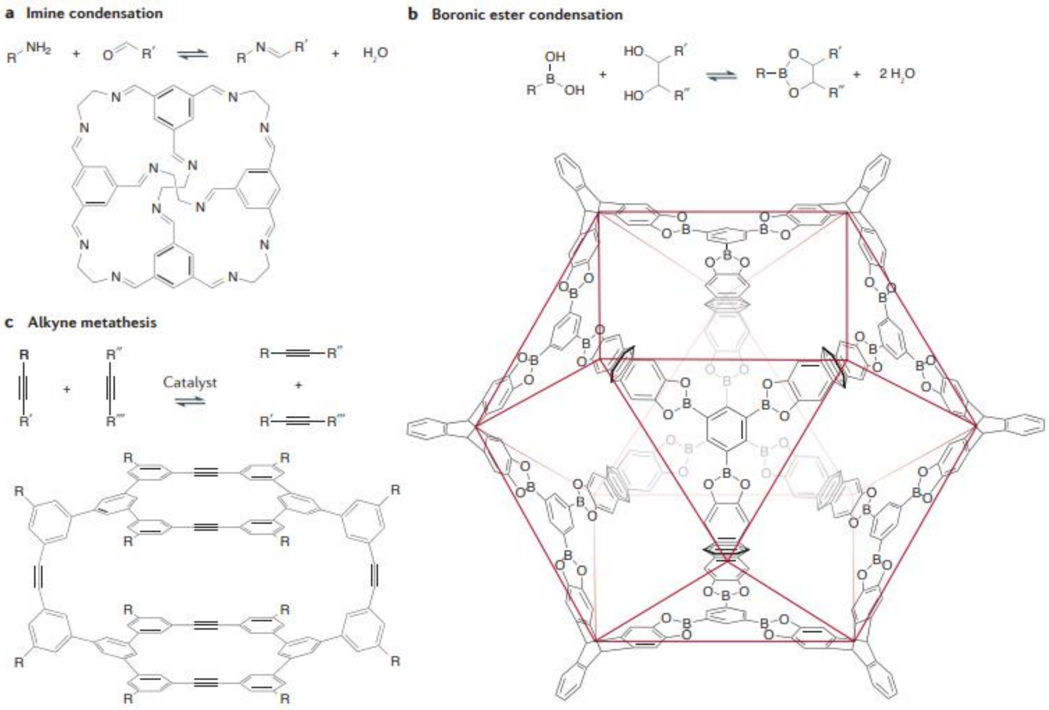

a Imine condensation
b Boronic ester condensation
c Alkyne metathesis
Catalyst

Figure 1.36 Synthetic pathways for the formation of porous organic cages by (a)imine, (b) boronate ester, and (c) alkyne metathesis chemistries[195].

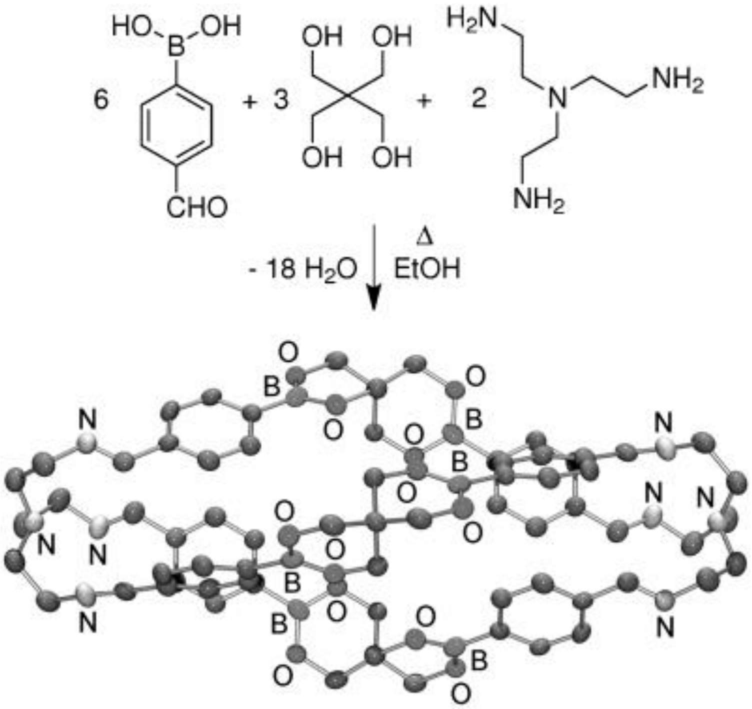

6
+ 3
+
2
- 18 H2O
Δ
EtOH

Figure 1.37 Formation of a molecular cage bearing imine- and boronate ester-linkages[42].

1.7. Overview of Subsequent Chapters

The remaining chapters of this dissertation will experimentally explore the ideas introduced in this chapter and are arranged as follows:

Chapter 2 will discuss the sequence-selective hybridization of molecular ladders bearing imine-rungs. This chapter details the method in which $Sc(OTf)_3$ is employed as a multi-role catalyst to deprotect aldehyde species in situ, dissociate information-bearing oligomeric strands, and promote interstrand shuffling and error-correction of assembled species. The selectivity of hybridization is characterized and the simultaneous assembly of several distinct molecular constructs is demonstrated. Moreover, the established technique is applied for non-destructive data storage and sequencing.

Chapter 3 highlights the combination of two dynamic covalent chemistries in the synthesis of a base-4 information system that has been designed to form molecular ladders and molecular grids. Herein we describe the conditions necessary to form specific molecular architectures comprised of imine and boronate ester linkages by first exploring reaction conditions for the individual systems, then assembling molecular grids in a sequential manner whereby the two dynamic covalent interactions process separately, and finally the assembly of multi-functional and irregular sequences are demonstrated.

Chapter 4 details the formation of molecular ladders utilizing a thermally reversible dynamic covalent reaction between a furan and a maleimide to affect oligomer hybridization. This chapter details the reaction conditions necessary to form monomers and oligomers bearing the reversible functional groups, and the subsequent conditions necessary to mediate hybridization. Degree of molecular ladder alignment and temperature effects on hybridization are investigated in this chapter. Sequence-selective hybridization using a thermal cycle analogous to nucleic acid self-assembly is shown.

Chapter 5 will detail the preparation of diagnostics tests for antibodies produced in response to SARS-CoV-2. The synthesis of SARS-CoV-2 epitopes, as peptide sequences, to selectively bind with antibodies produced by humans in response is outlined along with the modifications necessary for conjugation to red blood cells for use in a rapid gel card agglutination assays.

Chapter 6 will give an overall summary of the importance of this work, and it will provide a perspective on the future of the field.

1.8. References

1. Philp, D. & Stoddart, J. F. Self-Assembly in Natural and Unnatural Systems. *Angew. Chem. Int. Ed.* 1155–1196 (1996).
2. Mattia, E. & Otto, S. Supramolecular Systems Chemistry. *Nat. Nanotechnol.* **10,** 111–119 (2015).
3. Stupp, S. I. *et al.* Supramolecular Materials: Self-Organized Nanostructures. *Science* **276,** 384–389 (1997).
4. van der Asdonk, P., Keshavarz, M., Christianen, P. C. M. & Kouwer, P. Directed peptide amphiphile assembly using aqueous liquid crystal templates in magnetic fields. *Soft Matter* **12,** 6518–6525 (2016).
5. Wei, T., Furgal, J. C., Jung, J. H. & Scott, T. F. Long, self-assembled molecular ladders by cooperative dynamic covalent reactions. *Polym. Chem.* **8,** 520–527 (2017).
6. Mastalerz, M. Shape-Persistent Organic Cage Compounds by Dynamic Covalent Bond Formation. *Angew. Chem. Int. Ed.* 5042–5053 (2010). doi:10.1002/anie.201000443
7. Jiang, S. *et al.* Oriented Two-Dimensional Porous Organic Cage Crystals. *Angew. Chem. Int. Ed.* **56,** 9391–9395 (2017).
8. Jin, Y., Yu, C., Denman, R. J. & Zhang, W. Recent advances in dynamic covalent chemistry. *Chem. Soc. Rev.* **42,** 6634–6654 (2013).
9. Reuther, J. F., Dahlhauser, S. D. & Anslyn, E. V. Tunable Orthogonal Reversible Covalent (TORC) Bonds: Dynamic Chemical Control over Molecular Assembly. *Angew. Chem. Int. Ed.* **58,** 74–85 (2019).
10. Pinheiro, A. V., Han, D., Shih, W. M. & Yan, H. Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotechnol.* **6,** 763–772 (2011).
11. Saccà, B. & Niemeyer, C. M. DNA origami: The art of folding DNA. *Angew. Chem. Int. Ed.* **51,** 58–66 (2012).
12. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254,** 1497–1500 (1991).
13. Crick, F. & Watson, J. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose

Nucleic Acid. *Nature* **171,** 737 (1953).
14. Kallenbach, N. R., Ma, R. I. & Seeman, N. C. An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **305,** 829–831 (1983).
15. Seeman, N. C. DNA engineering and its application to nanotechnology. *Trends Biotechnol.* **17,** 437–443 (1999).
16. Seeman, N. C. DNA Nicks and Nodes and Nanotechnology. *Nano Lett.* **1,** 22–26 (2001).
17. Chen, J. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350,** 631–633 (1991).
18. Lin, Q. Y. *et al.* Building superlattices from individual nanoparticles via template-confined DNA-mediated assembly. *Science* **359,** 669–672 (2018).
19. Wang, P., Meyer, T. A., Pan, V., Dutta, P. K. & Ke, Y. The Beauty and Utility of DNA Origami. *Chem* **2,** 359–382 (2017).
20. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).
21. Hung, A. M. *et al.* Large-area spatially ordered arrays of gold nanoparticles directed by lithographically confined DNA origami. *Nat. Nanotechnol.* **5,** 121–126 (2010).
22. Acuna, G. P. *et al.* Distance dependence of single-fluorophore quenching by gold nanoparticles studied on DNA origami. *ACS Nano* **6,** 3189–3195 (2012).
23. Maune, H. T. *et al.* Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nat. Nanotechnol.* **5,** 61–66 (2010).
24. Ke, Y. *et al.* Scaffolded DNA origami of a DNA tetrahedron molecular container. *Nano Lett.* **9,** 2445–2447 (2009).
25. Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459,** 73–76 (2009).
26. Sanderson, K. What to make with DNA origami. *Nature* **464,** 158–159 (2010).
27. Zhao, Z., Liu, Y. & Yan, H. Organizing DNA origami tiles into larger structures using preformed scaffold frames. *Nano Lett.* **11,** 2997–3002 (2011).
28. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338,** 1177–1183 (2012).
29. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485,** 623–626 (2012).
30. Clausen-Schaumann, H., Rief, M., Tolksdorf, C. & Gaub, H. E. Mechanical stability of single DNA molecules. *Biophys. J.* **78,** 1997–2007 (2000).
31. Herrmann, A. Dynamic combinatorial/covalent chemistry: a tool to read, generate and modulate the bioactivity of compounds and compound mixtures. *Chem. Soc. Rev.* **43,** 1899–1933 (2014).
32. Rowan, S. J., Cantrill, S. J., Cousins, G. R. L., Sanders, J. K. M. & Stoddart, J. F. *Dynamic covalent chemistry. Angew. Chem. Int. Ed.* **41,** (2002).
33. Cencer, M. M., Greenlee, A. J. & Moore, J. S. Quantifying Error Correction through a Rule-Based Model of Strand Escape from an [n]-Rung Ladder. *J. Am. Chem. Soc.* **142,** 162–168 (2020).
34. Belowich, M. E. & Stoddart, J. F. Dynamic imine chemistry. *Chem. Soc. Rev.* **41,** 2003–2024 (2012).
35. Schiff, H. Mittheilungen aus dem Universitätslaboratorium in Pisa: Eine neue Reihe organischer Basen. *Justus Liebigs Ann. Chem.* **131,** 118–119 (1864).
36. Ciaccia, M., Pilati, S., Cacciapaglia, R., Mandolini, L. & Stefano, S. Di. Effective

catalysis of imine metathesis by means of fast transiminations between aromatic–aromatic or aromatic–aliphatic amines. *Org. Biomol. Chem. Pap. View* 3282–3287 (2014). doi:10.1039/c4ob00107a

37. Giuseppone, N., Schmitt, J. L., Schwartz, E. & Lehn, J. M. Scandium(III) catalysis of transimination reactions. Independent and constitutionally coupled reversible processes. *J. Am. Chem. Soc.* **127,** 5528–5539 (2005).
38. Bull, S. D. *et al.* Exploiting the reversible covalent bonding of boronic acids: Recognition, sensing, and assembly. *Acc. Chem. Res.* **46,** 312–326 (2013).
39. Stubelius, A., Lee, S. & Almutairi, A. The Chemistry of Boronic Acids in Nanomaterials for Drug Delivery. *Acc. Chem. Res.* **52,** 3108–3119 (2019).
40. Hutin, M., Bernardinelli, G. & Nitschke, J. R. An iminoboronate construction set for subcomponent self-assembly. *Chem. Eur. J.* **14,** 4585–4593 (2008).
41. Christinat, N., Scopelliti, R. & Severin, K. Multicomponent assembly of boron-based dendritic nanostructures. *J. Org. Chem.* **72,** 2192–2200 (2007).
42. Christinat, N., Scopelliti, R. & Severin, K. Multicomponent assembly of boronic acid based macrocycles and cages. *Angew. Chem. Int. Ed.* **47,** 1848–1852 (2008).
43. Dunn, M. F., Wei, T., Scott, T. F. & Zuckermann, R. N. Aqueous dynamic covalent assembly of molecular ladders and grids bearing boronate ester rungs. *Polym. Chem.* **10,** 2337–2343 (2019).
44. Nishiyabu, R., Kubo, Y., James, T. D. & Fossey, J. S. Boronic acid building blocks: Tools for self assembly. *Chem. Commun.* **47,** 1124–1150 (2011).
45. Ward, D. E. & Abaee, M. S. Intramolecular Diels-Alder Reaction by Self-assembly of the Components on a Lewis Acid Template. *Org. Lett.* **2,** 3937–3940 (2000).
46. Discekici, E. H. *et al.* Endo and Exo Diels-Alder Adducts: Temperature-Tunable Building Blocks for Selective Chemical Functionalization. *J. Am. Chem. Soc.* **140,** 5009–5013 (2018).
47. Adzima, B. J., Aguirre, H. A., Kloxin, C. J., Scott, T. F. & Bowman, C. N. Rheological and chemical analysis of reverse gelation in a covalently cross-linked diels-alder polymer network. *Macromolecules* **41,** 9112–9117 (2008).
48. Liu, Y. L. & Chuo, T. W. Self-healing polymers based on thermally reversible Diels-Alder chemistry. *Polym. Chem.* **4,** 2194–2205 (2013).
49. Froidevaux, V. *et al.* Study of the diels-alder and retro-diels-alder reaction between furan derivatives and maleimide for the creation of new materials. *RSC Adv.* **5,** 37742–37754 (2015).
50. Rowan, S. J., Cantrill, S. J., Cousins, G. R. L., Sanders, J. K. M. & Stoddart, J. F. *Dynamic covalent chemistry*. *Angewandte Chemie - International Edition* **41,** (2002).
51. Lutz, J. F. Aperiodic copolymers. *ACS Macro Lett.* **3,** 1020–1023 (2014).
52. Lutz, J.-F., Ouchi, M., Liu, D. R. & Sawamoto, M. Sequence-controlled polymers. *Science* **341,** 1238149 (2013).
53. Rutten, M. G. T. A., Vaandrager, F. W., Elemans, J. A. A. W. & Nolte, R. J. M. Encoding information into polymers. *Nat. Rev. Chem.* **2,** 365–381 (2018).
54. Mutlu, H. & Lutz, J. F. Reading polymers: Sequencing of natural and synthetic macromolecules. *Angew. Chem. Int. Ed.* **53,** 13010–13019 (2014).
55. Lutz, J. F. Coding Macromolecules: Inputting Information in Polymers Using Monomer-Based Alphabets. *Macromolecules* **48,** 4759–4767 (2015).
56. Giuseppone, N. Toward self-constructing materials: A systems chemistry approach. *Acc.*

Chem. Res. **45,** 2178–2188 (2012).

57. Loakes, D. & Holliger, P. Polymerase engineering: Towards the encoded synthesis of unnatural biopolymers. *Chem. Commun.* 4619–4631 (2009). doi:10.1039/b903307f
58. Pinheiro, V. B. & Holliger, P. Towards XNA nanotechnology: New materials from synthetic genetic polymers. *Trends Biotechnol.* **32,** 321–328 (2014).
59. Pinheiro, V. B., Loakes, D. & Holliger, P. Synthetic polymers and their potential as genetic materials. *BioEssays* **35,** 113–122 (2013).
60. Anosova, I. *et al.* The structural diversity of artificial genetic polymers. *Nucleic Acids Res.* **44,** 1007–1021 (2016).
61. Pinheiro, V. B. & Holliger, P. The XNA world: Progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **16,** 245–252 (2012).
62. Austin, J. H. Synthetic Genetic Polymers Capable of Heredity and Evolution. *Meditating Selflessly* 341–345 (2019). doi:10.7551/mitpress/8876.003.0036
63. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149–2154 (1963).
64. Grubbs, R. B. & Grubbs, R. H. 50th Anniversary Perspective: Living Polymerization - Emphasizing the Molecule in Macromolecules. *Macromolecules* **50,** 6979–6997 (2017).
65. Gutekunst, W. R. & Hawker, C. J. A General Approach to Sequence-Controlled Polymers Using Macrocyclic Ring Opening Metathesis Polymerization. *J. Am. Chem. Soc.* **137,** 8038–8041 (2015).
66. Lutz, J. F., Lehn, J. M., Meijer, E. W. & Matyjaszewski, K. From precision polymers to complex materials and systems. *Nat. Rev. Mater.* **1,** (2016).
67. De Neve, J., Haven, J. J., Maes, L. & Junkers, T. Sequence-definition from controlled polymerization: The next generation of materials. *Polym. Chem.* **9,** 4692–4705 (2018).
68. Ouchi, M. & Sawamoto, M. Sequence-controlled polymers via reversible-deactivation radical polymerization. *Polym. J.* **50,** 83–94 (2018).
69. Nanjan, P. & Porel, M. Sequence-defined non-natural polymers: Synthesis and applications. *Polym. Chem.* **10,** 5406–5424 (2019).
70. Rosales, A. M., Segalman, R. A. & Zuckermann, R. N. Polypeptoids: A model system to study the effect of monomer sequence on polymer properties and self-assembly. *Soft Matter* **9,** 8400–8414 (2013).
71. Fields, G. B. & Noble, R. L. Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.* **35,** 161–214 (1990).
72. Palomo, J. M. Solid-phase peptide synthesis: An overview focused on the preparation of biologically relevant peptides. *RSC Adv.* **4,** 32658–32672 (2014).
73. Cherkupally, P. *et al.* Immobilized coupling reagents: Synthesis of amides/peptides. *ACS Comb. Sci.* **16,** 579–601 (2014).
74. Chongsiriwatana, N. P. *et al.* Peptoids that mimic the structure, function, and mechanism of helical antimicrobial peptides. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 2794–2799 (2008).
75. Zuckermann, R. N., Kerr, J. M., Moosf, W. H. & Kent, S. B. H. Efficient Method for the Preparation of Peptoids [Oligo(N-substituted glycines)] by Submonomer Solid-Phase Synthesis. *J. Am. Chem. Soc.* **114,** 10646–10647 (1992).
76. Zuckermann, R. N., Kerr, J. M., Kent, S. B. H. & Moos, W. H. Efficient Method for the Preparation of Peptoids [Oligo(N-substituted glycines)] by submonomer solid-phase synthesis. *J. Am. Chem. Soc.* **114,** 10646–10647 (1992).
77. Webster, A. M. & Cobb, S. L. Recent Advances in the Synthesis of Peptoid Macrocycles.

Chem. Eur. J. **24,** 7560–7573 (2018).
78. Solleder, S. C., Schneider, R. V., Wetzel, K. S., Boukis, A. C. & Meier, M. A. R. Recent Progress in the Design of Monodisperse, Sequence-Defined Macromolecules. *Macromol. Rapid Commun.* **38,** 1–45 (2017).
79. Pfeifer, S., Zarafshani, Z., Badi, N. & Lutz, J. F. Liquid-phase synthesis of block copolymers containing sequence-ordered segments. *J. Am. Chem. Soc.* **131,** 9195–9197 (2009).
80. Barnes, J. C. *et al.* Iterative exponential growth of stereo- and sequence-controlled polymers. *Nat. Chem.* **7,** 810–815 (2015).
81. Zydziak, N., Feist, F., Huber, B., Mueller, J. O. & Barner-Kowollik, C. Photo-induced sequence defined macromolecules via hetero bifunctional synthons. *Chem. Commun.* **51,** 1799–1802 (2015).
82. Meier, M. A. R. & Barner-Kowollik, C. A New Class of Materials: Sequence-Defined Macromolecules and Their Emerging Applications. *Adv. Mater.* **31,** 1–5 (2019).
83. Zhang, J., Matta, M. E. & Hillmyer, M. A. Synthesis of sequence-specific vinyl copolymers by regioselective ROMP of multiply substituted cyclooctenes. *ACS Macro Lett.* **1,** 1383–1387 (2012).
84. Pfeifer, S. & Lutz, J. F. A facile procedure for controlling monomer sequence distribution in radical chain polymerizations. *J. Am. Chem. Soc.* **129,** 9542–9543 (2007).
85. Tørring, T., Voigt, N. V., Nangreave, J., Yan, H. & Gothelf, K. V. DNA origami: A quantum leap for self-assembly of complex structures. *Chem. Soc. Rev.* **40,** 5636–5646 (2011).
86. Church, G. M., Gao, Y. & Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science* (2012). doi:10.1038/21092
87. Kool, E. T., Morales, J. C. & Guckian, K. M. Mimicking the structure and function of DNA: Insights into DNA stability and replication. *Angew. Chem. Int. Ed.* **39,** 990–1009 (2000).
88. Haque, M. M. & Peng, X. DNA-associated click chemistry. *Sci. China Chem.* **57,** 215–231 (2014).
89. Verma, S. & Eckstein, F. MODIFIED OLIGONUCLEOTIDES: Synthesis and Strategy for Users. *Annu. Rev. Biochem.* **67,** 99–134 (1998).
90. Zhao, Y., Chen, F., Li, Q., Wang, L. & Fan, C. Isothermal Amplification of Nucleic Acids. *Chem. Rev.* **115,** 12491–12545 (2015).
91. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145 (2008).
92. Lutz, J. Can Life Emerge from Synthetic Polymers ? 1–10 (2020). doi:10.1002/ijch.201900110
93. Davis, J. Microvenus. *Art J.* **55,** 70–74 (1996).
94. Goldman, N. *et al.* Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494,** 77–80 (2013).
95. Shankland, S. Startup packs all 16GB of Wikipedia onto DNA strands to demonstrate new storage tech. *CNET* (2019). Available at: https://www.cnet.com/news/startup-packs-all-16gb-wikipedia-onto-dna-strands-demonstrate-new-storage-tech/.
96. Panda, D. *et al.* DNA as a digital information storage device: hope or hype? *3 Biotech* **8,** 1–9 (2018).
97. Colquhoun, H. & Lutz, J. F. Information-containing macromolecules. *Nat. Chem.* **6,** 455–

456 (2014).

98. Vandenbergh, J., Reekmans, G., Adriaensens, P. & Junkers, T. Synthesis of sequence controlled acrylate oligomers via consecutive RAFT monomer additions. *Chem. Commun.* **49,** 10358–10360 (2013).
99. Shendure, J. *et al.* DNA sequencing at 40: Past, present and future. *Nature* **550,** (2017).
100. Heerma, W. *et al.* Comparing mass spectrometric characteristics of peptides and peptoids. *J. Mass Spectrom.* **32,** 697–704 (1997).
101. Ruijtenbeek, R. *et al.* Characterization of a phosphorylated peptide and peptoid and peptoid-peptide hybrids by mass spectrometry. *J. Mass Spectrom.* **37,** 47–55 (2002).
102. Samperi, F., Montaudo, M. S., Puglisi, C., Di Giorgi, S. & Montaudo, G. Structural characterization of copolyamides synthesized via the facile blending of polyamides. *Macromolecules* **37,** 6449–6459 (2004).
103. Roy, R. K. *et al.* Design and synthesis of digitally encoded polymers that can be decoded and erased. *Nat. Commun.* **6,** 1–8 (2015).
104. Porel, M. & Alabi, C. A. Sequence-defined polymers via orthogonal allyl acrylamide building blocks. *J. Am. Chem. Soc.* **136,** 13162–13165 (2014).
105. Charles, L., Laure, C., Lutz, J. F. & Roy, R. K. MS/MS Sequencing of Digitally Encoded Poly(alkoxyamine amide)s. *Macromolecules* **48,** 4319–4328 (2015).
106. Amalian, J. A., Trinh, T. T., Lutz, J. F. & Charles, L. MS/MS Digital Readout: Analysis of Binary Information Encoded in the Monomer Sequences of Poly(triazole amide)s. *Anal. Chem.* **88,** 3715–3722 (2016).
107. Cavallo, G., Al Ouahabi, A., Oswald, L., Charles, L. & Lutz, J. F. Orthogonal Synthesis of 'Easy-to-Read' Information-Containing Polymers Using Phosphoramidite and Radical Coupling Steps. *J. Am. Chem. Soc.* **138,** 9417–9420 (2016).
108. Al Ouahabi, A., Amalian, J. A., Charles, L. & Lutz, J. F. Mass spectrometry sequencing of long digital polymers facilitated by programmed inter-byte fragmentation. *Nat. Commun.* **8,** 1–8 (2017).
109. Gunay, U. S. *et al.* Chemoselective Synthesis of Uniform Sequence-Coded Polyurethanes and Their Use as Molecular Tags. *Chem* **1,** 114–126 (2016).
110. Karamessini, D., Petit, B. E., Bouquey, M., Charles, L. & Lutz, J. F. Identification-Tagging of Methacrylate-Based Intraocular Implants Using Sequence Defined Polyurethane Barcodes. *Adv. Funct. Mater.* **27,** (2017).
111. Zydziak, N. *et al.* Coding and decoding libraries of sequence-defined functional copolymers synthesized via photoligation. *Nat. Commun.* **7,** 1–10 (2016).
112. Boukis, A. C. & Meier, M. A. R. Data storage in sequence-defined macromolecules via multicomponent reactions. *Eur. Polym. J.* **104,** 32–38 (2018).
113. Martens, S. *et al.* Multifunctional sequence-defined macromolecules for chemical data storage. *Nat. Commun.* **9,** 1–8 (2018).
114. Biniossek, M. L. & Schilling, O. Enhanced identification of peptides lacking basic residues by LC-ESI-MS/MS analysis of singly charged peptides. *Proteomics* **12,** 1303–1309 (2012).
115. Huang, Z. *et al.* Combining Orthogonal Chain-End Deprotections and Thiol–Maleimide Michael Coupling: Engineering Discrete Oligomers by an Iterative Growth Strategy. *Angew. Chem. Int. Ed.* **56,** 13612–13617 (2017).
116. Ding, K. *et al.* Easily encodable/decodable digital polymers linked by dithiosuccinimide motif. *Eur. Polym. J.* **119,** 421–425 (2019).

117. Lee, J. M. *et al.* High-density information storage in an absolutely defined aperiodic sequence of monodisperse copolyester. *Nat. Commun.* **11,** 1–9 (2020).
118. Szweda, R., Tschopp, M., Felix, O., Decher, G. & Lutz, J. F. Sequences of Sequences: Spatial Organization of Coded Matter through Layer-by-Layer Assembly of Digital Polymers. *Angew. Chem. Int. Ed.* **57,** 15817–15821 (2018).
119. Jamil, T. Steganography: the art of hiding information in plain sight. *IEEE Potentials* **18,** 10–12 (1999).
120. Sarkar, T., Selvakumar, K., Motiei, L. & Margulies, D. Message in a molecule. *Nat. Commun.* **7,** 1–9 (2016).
121. Boukis, A. C., Reiter, K., Frölich, M., Hofheinz, D. & Meier, M. A. R. Multicomponent reactions provide key molecules for secret communication. *Nat. Commun.* **9,** 1–10 (2018).
122. Amalian, J. A., Cavallo, G., Al Ouahabi, A., Lutz, J. F. & Charles, L. Revealing Data Encrypted in Sequence-Controlled Poly(Alkoxyamine Phosphodiester)s by Combining Ion Mobility with Tandem Mass Spectrometry. *Anal. Chem.* **91,** 7266–7272 (2019).
123. Ouahabi, A. Al, Oswald, L. & Szweda, R. Photo-editable macromolecular information. *Nat. Commun.* 1–9 (2019). doi:10.1038/s41467-019-11566-2
124. Leguizamon, S. C. & Scott, T. F. Sequence-selective dynamic covalent assembly of information-bearing oligomers. *Nat. Commun.* **11,** 784 (2020).
125. Hill, D. J., Mio, M. J., Prince, R. B., Hughes, T. S. & Moore, J. S. A field guide to foldamers. *Chem. Rev.* **101,** 3893–4011 (2001).
126. Haldar, D. & Schmuck, C. Metal-free double helices from abiotic backbones. *Chem. Soc. Rev.* **38,** 363–371 (2009).
127. Berl, V., Huc, I., Khoury, R. G., Krische, M. J. & Lehn, J. M. Interconversion of single and double helices formed from synthetic molecular strands. *Nature* **407,** 720–723 (2000).
128. Stross, A. E., Iadevaia, G. & Hunter, C. A. Cooperative duplex formation by synthetic H-bonding oligomers. *Chem. Sci.* **7,** 94–101 (2016).
129. Gong, B. Molecular duplexes with encoded sequences and stabilities. *Acc. Chem. Res.* **45,** 2077–2087 (2012).
130. Koert, U., Harding, M. M. & Lehn, J. M. DNH deoxyribonucleohelicates: Self assembly of oligonucleosidic double-helical metal complexes. *Nature* **346,** 339–342 (1990).
131. Bisson, A. P. *et al.* Synthesis and recognition properties of aromatic amide oligomers: Molecular zippers. *J. Am. Chem. Soc.* **122,** 8856–8868 (2000).
132. Gabriel, G. J. & Iverson, B. L. Aromatic oligomers that form hetero duplexes in aqueous solution. *J. Am. Chem. Soc.* **124,** 15174–15175 (2002).
133. Zych, A. J. & Iverson, B. L. Synthesis and conformational characterization of tethered, self-complexing 1,5-dialkoxynaphthalene/1,4,5,8-naphthalenetetracarboxylic diimide systems. *J. Am. Chem. Soc.* **122,** 8898–8909 (2000).
134. Goto, H., Katagiri, H., Furusho, Y. & Yashima, E. Oligoresorcinols fold into double helices in water. *J. Am. Chem. Soc.* **128,** 7176–7178 (2006).
135. Sugiura, H., Nigorikawa, Y., Saiki, Y., Nakamura, K. & Yamaguchi, M. Marked effect of aromatic solvent on unfolding rate of helical ethynylhelicene oligomer. *J. Am. Chem. Soc.* **126,** 14858–14864 (2004).
136. Amemiya, R., Saito, N. & Yamaguchi, M. Hetero-double-helix formation by an ethynylhelicene oligomer possessing perfluorooctyl side chains. *J. Org. Chem.* **73,** 7137–7144 (2008).
137. Shang, J. *et al.* Self-association of aromatic oligoamide foldamers into double helices in

water. *Org. Lett.* **16,** 4992–4995 (2014).
138. Gan, Q. *et al.* Quadruple and double helices of 8-fluoroquinoline oligoamides. *Angew. Chem. Int. Ed.* **47,** 1715–1718 (2008).
139. Gan, Q. *et al.* Heteromeric double helix formation by cross-hybridization of chloro-and fluoro-substituted quinoline oligoamides. *Chem. Commun.* **46,** 297–299 (2010).
140. Ferrand, Y., Kendhale, A. M., Garric, J., Kauffmann, B. & Huc, I. Parallel and antiparallel triple helices of naphthyridine oligoamides. *Angew. Chem. Int. Ed.* **49,** 1778–1781 (2010).
141. Zhang, D., Tian, J., Chen, L., Zhang, L. & Li, Z. Dimerization of conjugated radical cations : an emerging non-covalent interaction for self-assembly. *Chem. Asian J.* **10,** 56–68 (2015).
142. Wang, Y. *et al.* Folding of oligoviologens induced by radical-radical interactions. *J. Am. Chem. Soc.* **137,** 876–885 (2015).
143. Kurlemann, M. & Ravoo, B. J. Towards the sequence-specific multivalent molecular recognition of cyclodextrin oligomers. *Beilstein J. Org. Chem.* **10,** 2428–2440 (2014).
144. Hartley, C. S., Elliott, E. L. & Moore, J. S. Covalent assembly of molecular ladders. *J. Am. Chem. Soc.* **129,** 4512–4513 (2007).
145. Elliott, E. L., Hartley, C. S. & Moore, J. S. Covalent ladder formation becomes kinetically trapped beyond four rungs. *Chem. Commun.* **47,** 5028–5030 (2011).
146. Wei, T., Jung, J. H. & Scott, T. F. Dynamic covalent assembly of peptoid-based ladder oligomers by vernier templating. *J. Am. Chem. Soc.* **137,** 16196–16202 (2015).
147. Reuther, J. F. *et al.* Dynamic covalent chemistry enables formation of antimicrobial peptide quaternary assemblies in a completely abiotic manner. *Nat. Chem.* **10,** 45–50 (2018).
148. Furgal, J. C., Dijck, J. M. Van, Leguizamon, S. C. & Scott, T. F. Accessing sequence specific hybrid peptoid oligomers with varied pendant group spacing. *Eur. Polym. J.* 306–311 (2019).
149. Iida, H. *et al.* Double-Stranded Helical Oligomers Covalently Bridged by Rotary Cyclic Boronate Esters. *Chem. Asian J.* **12,** 927–935 (2017).
150. Dunn, M. F., Wei, T., Zuckermann, R. N. & Scott, T. F. Aqueous dynamic covalent assembly of molecular ladders and grids bearing boronate ester rungs. *Polym. Chem.* **10,** 2337–2343 (2019).
151. Hebel, M. *et al.* Sequence Programming with Dynamic Boronic Acid/Catechol Binary Codes. *J. Am. Chem. Soc.* **141,** 14026–14031 (2019).
152. Drogkaris, V. & Northrop, B. H. Discrete boronate ester ladders from the dynamic covalent self-assembly of oligo(phenylene ethynylene) derivatives and phenylenebis(boronic acid). *Org. Chem. Front.* 1082–1094 (2020). doi:10.1039/d0qo00083c
153. Gong, B. Molecular duplexes with encoded sequences and stabilities. *Acc. Chem. Res.* **45,** 2077–2087 (2012).
154. Coppock, M. B., Miller, J. R. & Williams, M. E. Assembly of a trifunctional artificial peptide into an anti-parallel duplex with three Cu(II) cross-links. *Inorg. Chem.* **50,** 949–955 (2011).
155. Marquis, A. *et al.* Messages in molecules: Ligand/cation coding and self-recognition in a constitutionally dynamic system of heterometallic double helicates. *Chem. Eur. J.* **12,** 5632–5641 (2006).
156. Wei, T., Furgal, J. C. & Scott, T. F. In situ deprotection and dynamic covalent assembly

using a dual role catalyst. *Chem. Commun.* **53,** 3874–3877 (2017).

157. Leguizamon, S. C., Dunn, M. & Scott, T. F. Sequence-directed Dynamic Covalent Assembly of Base-4-encoded Oligomers. *Chem. Commun.* **56,** 7817–7820 (2020).
158. Orgel, L. E. Unnatural Selection in Chemical Systems. *Acc. Chem. Res.* **5800,** 109–118 (1995).
159. Vidonne, A. & Philp, D. Making molecules make themselves - The chemistry of artificial replicators. *European J. Org. Chem.* 593–610 (2009). doi:10.1002/ejoc.200800827
160. Tjivikua, T., Ballester, P. & Rebek, J. A Self-Replicating System. *J. Am. Chem. Soc.* **112,** 1249–1250 (1990).
161. Terfort, A. & Kiedrowski, G. von. Self-replication by condensation of 3-amino-benzamidines and 2-formylphenoxyacetic acids. *Angew. Chem. Int. Ed.* **31,** 654–656 (1992).
162. Carnall, J. M. A. *et al.* Mechanosensitive self-replication driven by self-organization. *Science* **327,** 1502–1507 (2010).
163. Niu, J., Hili, R. & Liu, D. R. Enzyme-free translation of DNA into sequence- defined synthetic polymers structurally unrelated to nucleic acids. *Nat. Chem.* **5,** 282–292 (2013).
164. South, C. R. & Weck, M. Template-Enhanced Ring-Opening Metathesis Polymerization. **2,** 1386–1394 (2007).
165. Lo, P. K. & Sleiman, H. F. Nucleobase-templated polymerization : copying the chain length and polydispersity of living polymers into conjugated polymers. *J. Am. Chem. Soc.* **131,** 4182–4183 (2009).
166. Zhou, Z. & Palermo, E. F. Templated Ring-Opening Metathesis (TROM) of Cyclic Olefins Tethered to Unimolecular Oligo(thiophene)s. *Macromolecules* **51,** 6127–6137 (2018).
167. Ke, Y. *et al.* Well-defined condensation polymers with narrow polydispersity via unsymmetrical ladderphanes by sequential polymerization. *Macromolecules* **46,** 6712–6722 (2013).
168. Sullivan, M. C. O. *et al.* Vernier templating and synthesis of a 12-porphyrin nano-ring. *Nature* 10–13 (2011). doi:10.1038/nature09683
169. Kondratuk, D. V. *et al.* Two vernier-templated routes to a 24-porphyrin nanoring. *Angew. Chem. Int. Ed.* **51,** 6696–6699 (2012).
170. Kondratuk, D. V *et al.* Supramolecular nesting of cyclic polymers. *Nat. Chem.* **7,** 317–322 (2015).
171. Kamonsutthipaijit, N. & Anderson, H. L. Template-directed synthesis of linear porphyrin oligomers: classical, Vernier and mutual Vernier. *Chem. Sci.* **8,** 2729–2740 (2017).
172. Strom, K. R., Szostak, J. W. & Prywes, N. Transfer of Sequence Information and Replication of Diimine Duplexes. *J. Org. Chem.* **84,** 3754–3761 (2019).
173. Núñez-Villanueva, D. & Hunter, C. A. Molecular replication using covalent base-pairs with traceless linkers. *Org. Biomol. Chem.* **17,** 9660–9665 (2019).
174. Núñez-Villanueva, D., Ciaccia, M., Iadevaia, G., Sanna, E. & Hunter, C. A. Sequence information transfer using covalent template-directed synthesis. *Chem. Sci.* **10,** 5258–5266 (2019).
175. Zhang, Y., Huang, Y. & Li, S. Polymeric micelles: Nanocarriers for cancer-targeted drug delivery. *AAPS PharmSciTech* **15,** 862–871 (2014).
176. Claya, N. E. *et al.* Chemical and Mechanical Modulation of Polymeric Micelle Assembly. *Nanoscale* **9,** 5194–5204 (2017).

177. Minkenberg, C. B., Florusse, L., Eelkema, R., Koper, G. J. M. & Van Esch, J. H. Triggered self-assembly of simple dynamic covalent surfactants. *J. Am. Chem. Soc.* **131,** 11274–11275 (2009).
178. Minkenberg, C. B. *et al.* Dynamic covalent assembly of stimuli responsive vesicle gels. *Chem. Commun.* **48,** 9837–9839 (2012).
179. Minkenberg, C. B. *et al.* Responsive wormlike micelles from dynamic covalent surfactants. *Langmuir* **28,** 13570–13576 (2012).
180. Haldar, U., Bauri, K., Li, R., Faust, R. & De, P. Polyisobutylene-based pH-responsive self-healing polymeric gels. *ACS Appl. Mater. Interfaces* **7,** 8779–8788 (2015).
181. Ryu, J. H., Jiwpanich, S., Chacko, R., Bickerton, S. & Thayumanavan, S. Surface-functionalizable polymer nanogels with facile hydrophobic guest encapsulation capabilities. *J. Am. Chem. Soc.* **132,** 8246–8247 (2010).
182. Fairbanks, B. D., Singh, S. P., Bowman, C. N. & Anseth, K. S. Photodegradable, photoadaptable hydrogels via radical-mediated disulfide fragmentation reaction. *Macromolecules* **44,** 2444–2450 (2011).
183. Lyle, S. J., Waller, P. J. & Yaghi, O. M. Covalent Organic Frameworks: Organic Chemistry Extended into Two and Three Dimensions. *Trends Chem.* **1,** 172–184 (2019).
184. Feng, X., Ding, X. & Jiang, D. Covalent organic frameworks. *Chem. Soc. Rev.* **41,** 6010–6022 (2012).
185. Ma, X. & Scott, T. F. Approaches and challenges in the synthesis of three-dimensional covalent-organic frameworks. *Commun. Chem.* **1,** (2018).
186. Ockwig, N. W., Co, A. P., Keeffe, M. O., Matzger, A. J. & Yaghi, O. M. Porous , Crystalline , Covalent Organic Frameworks. **310,** 1166–1171 (2005).
187. Tahir, N., Krishnaraj, C., Leus, K. & Van Der Voort, P. Development of covalent triazine frameworks as heterogeneous catalytic supports. *Polymers (Basel).* **11,** (2019).
188. Wei, H. *et al.* The microwave-assisted solvothermal synthesis of a crystalline two-dimensional covalent organic framework with high CO2 capacity. *Chem. Commun.* **51,** 12178–12181 (2015).
189. Zeng, Y. *et al.* Covalent organic frameworks formed with two types of covalent bonds based on orthogonal reactions. *J. Am. Chem. Soc.* **137,** 1020–1023 (2015).
190. Zhang, G. & Mastalerz, M. Organic cage compounds-from shape-persistency to function. *Chem. Soc. Rev.* **43,** 1934–1947 (2014).
191. Ono, K. & Iwasawa, N. Dynamic Behavior of Covalent Organic Cages. *Chem. Eur. J.* **24,** 17856–17868 (2018).
192. Tozawa, T. *et al.* Porous organic cages. *Nat. Mater.* **8,** 973–978 (2009).
193. Quan, M. L. C. & Cram, D. J. Constrictive Binding of Large Guests by a Hemicarcerand Containing four Portals. *J. Am. Chem. Soc.* **113,** 2754–2755 (1991).
194. Jin, Y., Wang, Q., Taynton, P. & Zhang, W. Dynamic covalent chemistry approaches toward macrocycles, molecular cages, and polymers. *Acc. Chem. Res.* **47,** 1575–1586 (2014).
195. Hasell, T. & Cooper, A. I. Porous organic cages: Soluble, modular and molecular pores. *Nat. Rev. Mater.* **1,** (2016).

Chapter 2 Sequence-selective Dynamic Covalent Assembly of Information-bearing Oligomers

2.1 Original Publications Information

Leguizamon, S.C., Scott, T.F. Sequence-selective dynamic covalent assembly of information-bearing oligomers. Nat Commun 11, 784 (2020).

Leguizamon, S. C., Alqubati, A. F., Scott, T. F. Synthesis of Information-bearing Peptoids and their Sequence-directed Dynamic Covalent Self-assembly. J. Vis. Exp. (156), e60442, (2020)

Modifications have been made to the original documents in order to adapt the content to the proper format.

2.2 Abstract

Relatively robust dynamic covalent interactions have been employed extensively to mediate molecular self-assembly reactions; however, these assembly processes often do not converge to a thermodynamic equilibrium, instead yielding mixtures of kinetically-trapped species. Here we report a dynamic covalent self-assembly process that mitigates kinetic trapping such that multiple unique oligomers bearing covalently coreactive pendant groups are able to undergo simultaneous, sequence-selective hybridization with their complementary strands to afford biomimetic, in-registry molecular ladders with covalent rungs. Analogous to the thermal cycling commonly employed for nucleic acid melting and annealing, this is achieved by raising and lowering the concentration of a multi-role reagent to effect quantitative dissociation and subsequently catalyze covalent bond rearrangement, affording selective assembly of the oligomeric sequences. The hybridization specificity afforded by this process further enabled information encoded in

oligomers to be retrieved through selective hybridization with complementary, mass-labeled sequences.

2.3 Introduction

The capacity for sequence-specific polymer strands to selectively assemble into intricate, folded structures and multimeric complexes relies upon the information borne by their residue sequences[1,2]. Although there are several examples of the non-covalent assembly of information-bearing species[3–6], nucleic acids represent a particularly versatile class for producing nanostructures which, through careful consideration of their nucleobase sequence, can be designed to predictably self-assemble *via* the hybridization of complementary strands into arbitrary structures with nanometer-scale precision[7–10]. Unfortunately, the versatility of nucleic acids as nanoconstruction media is tempered by the thermal and mechanical instability of the resultant structures, attributable to the weakness of the hydrogen bonds that hold the strands together[11]. Efforts to improve the stability of double-stranded, nucleic acid-like duplexes include the utilization of abiotic, neutral backbones[12–14], the integration of isostere nucleobase mimetics[15,16], and the incorporation of photo-induced cross-links[17,18]; nevertheless, the assembly processes for these duplexes remain mediated by intermolecular interactions. In contrast, employing dynamic covalent interactions to mediate molecular self-assembly processes can afford inherently robust, covalently cross-linked structures[19–26], although impediments to connectivity rearrangement amongst the assembled components, including slow reaction rates and reaction site inaccessibility, often result in the kinetic trapping of non-equilibrium species[27–29]. Here, we describe a dynamic covalent assembly approach where mixtures of sequence-specific oligomers bearing covalently coreactive amine- and aldehyde-based pendant groups selectively associate and dimerize with their oligomeric complements to afford in-registry molecular ladders with imine-based covalent rungs, mimicking the selective, information-directed hybridization of complementary nucleic acid sequences.

2.4 Experimental

2.4.1 General Experimental Procedure

Unless otherwise stated, chemicals and reagents were obtained from commercial sources and used as received. ^{1}H NMR spectra were collected using Varian MR400 and Varian Inova 500 spectrometers. Diffusion ordered spectroscopy (DOSY) NMR spectra were collected using a Bruker 400 MHz NMR spectrometer. Chemical shifts were measured in δ (ppm) relative to residual solvent signals as internal standards ($CDCl_3$ – 7.24 for ^{1}H; CD_3CN – 1.94 for ^{1}H). Matrix-assisted laser desorption/ionization (MALDI) mass spectra were recorded using a Bruker Autoflex mass spectrometer, whereas electrospray ionization (ESI) mass spectra were recorded using an Agilent Q-TOF 1200 series spectrometer. MALDI analyses were performed in reflectron positive ion mode using 2-(4-hydroxyphenylazo)benzoic acid (HABA) as the matrix, where 2 μL of a solution of the sample (1 mM) was mixed with 6 μL of a mixture of 10 mg matrix in 200 μL acetonitrile, spotted on a MALDI sample plate (Bruker), and allowed to air dry. Reverse phase high performance liquid chromatography (RP-HPLC) was performed using a Shimadzu LC-6AD HPLC pump, equipped with a Shimadzu FRC 70A fraction collector, using analytical and preparative reversed phase Phenomenex Luna C18(2) columns with a linear gradient of water and acetonitrile as the eluent at 30°C, and monitored with a Shimadzu Prominence UV/vis detector at 214 nm. Gel permeation chromatography (GPC) was performed using a Shimadzu LC-20AD HPLC pump equipped with three Phenogel GPC/SEC columns (length 300 mm × diameter 7.8 mm, pore sizes of 500, 100, and 50 Å) in series, 94:4:2 (v/v/v) $CHCl_3$:MeOH:Et_3N as the eluent at room temperature, and monitored with a Shimadzu Prominence UV/vis detector at 313 nm. GPC calibration was performed using low dispersity polystyrene standards (low molecular weight Readycal Set, Fluka).

2.4.2 Monomer Synthesis

Primary amine monomers (Figure 2.1.) were purchased or synthesized according to the procedures below.

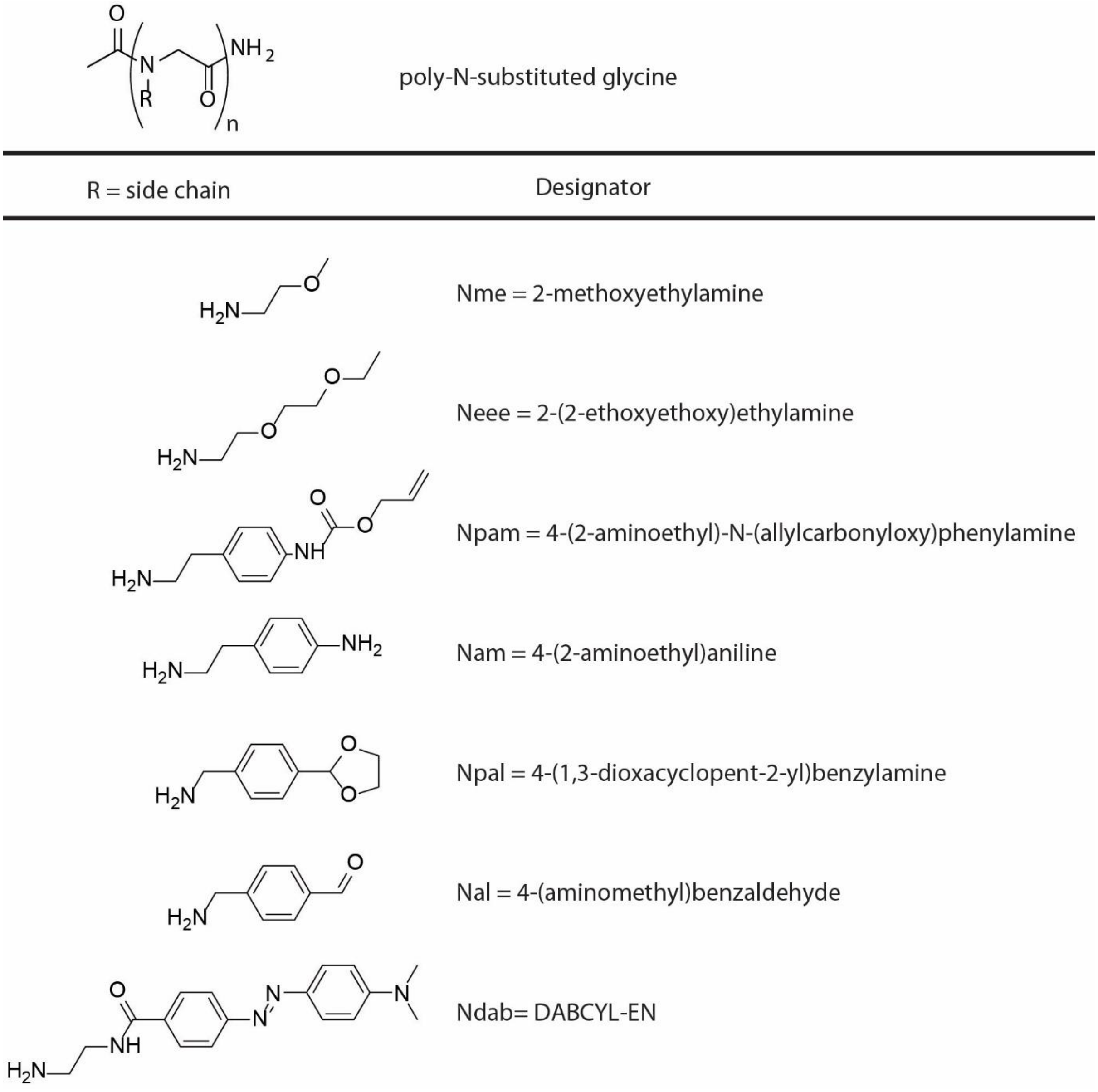

poly-N-substituted glycine
R = side chain
Designator
Nme = 2-methoxyethylamine
Neee = 2-(2-ethoxyethoxy)ethylamine
Npam = 4-(2-aminoethyl)-N-(allylcarbonyloxy)phenylamine
Nam = 4-(2-aminoethyl)aniline
Npal = 4-(1,3-dioxacyclopent-2-yl)benzylamine
Nal = 4-(aminomethyl)benzaldehyde
Ndab= DABCYL-EN

Figure 2.1 Primary amine monomers used in this study for peptoid synthesis.

Npal

Npal was synthesized by addition of 42.2 mL ethylene glycol, 0.02 g toluene *p*-sulfonic acid, and 25 g of 4-cyanobenzaldehyde to 200 mL of toluene. After refluxing overnight with a Dean-Stark trap to remove water generated during the reaction, the reaction mixture was cooled to room temperature and 40 mL of 5% $NaHCO_3$ aqueous solution was added. The organic layer was extracted, washed with deionized water three times, and evaporated to yield 4-(1,3-dioxacyclopent-2-yl)benzonitrile. 10 g of 4-(1,3-dioxacyclopent-2-yl)benzonitrile in 100 mL dry diethyl ether was added dropwise into 4.33 g of $LiAlH_4$ in 100 mL dry diethyl ether under nitrogen. The reaction mixture was stirred for 4 h at 0°C and 12 h at room temperature, then quenched by 95% ethanol, and further quenched by 50% ethanol in water. The ether supernatant was dried under reduced pressure to yield Npal.

Npam

4.9 g allyl chloroformate in 150 mL 1,4-dioxane was added into a solution of 4-(2-aminoethyl)aniline (5 g) in 150 mL 10% aq. acetic acid. The reaction mixture was stirred overnight at room temperature and then diluted with water prior to washing with diethyl ether. The aqueous layer was adjusted to pH 14 by 2 M NaOH (aq) and was extracted by diethyl ether. The organic fraction was evaporated to yield Npam.

Neee

50 mL of 6 M NaOH was added to 20 g of diethylene glycol monoethyl ether and 50 mL of tetrahydrofuran (THF) cooled to 0°C, followed by dropwise addition of 54 g tosyl chloride in 80 mL THF under N_2. After stirring for 1 h at 0°C, the reaction mixture was left for an additional hour at room temperature. The mixture was extracted with diethyl ether and the organic layer washed with 1 M NaOH and water before evaporating under vacuum to yield 2-(2-ethoxyethoxy)ethyl

tosylate. 40 g of 2-(2-ethoxyethoxy)ethyl tosylate and 250 mL of DMF were charged to a 500 mL flask under N_2. 31.5 g of NaN_3 was added to the reaction mixture which was then stirred at 60°C for 36 h before cooling to room temperature. The reaction mixture was extracted with diethyl ether and washed with water before evaporating under vacuum to produce 2-(2-ethoxyethoxy)ethyl azide. 2-(2-Ethoxyethoxy)ethyl azide (20 g) in 160 mL THF and 40 g triphenylphosphine were charged to a 250 mL flask under nitrogen. Upon stirring overnight, the reaction mixture was quenched with 220 mL water and allowed to stir for another day. The resulting solution was washed with toluene and dichloromethane, and evaporated under vacuum to yield Neee.

Ndab

To a 20 mL vial equipped with magnetic stirrer, 500 mg of DABCYL succinimide ester and 10 mL (excess) of distilled ethylene diamine were added. The reaction was allowed to stir for 12 h and then excess ethylene diamine was removed by rotary evaporation. The product was extracted by dichloromethane and then washed with distilled water to remove excess amine and succinimide. The organic layer was dried over Na_2SO_4, filtered and solvent removed by rotary evaporation to yield Ndab.

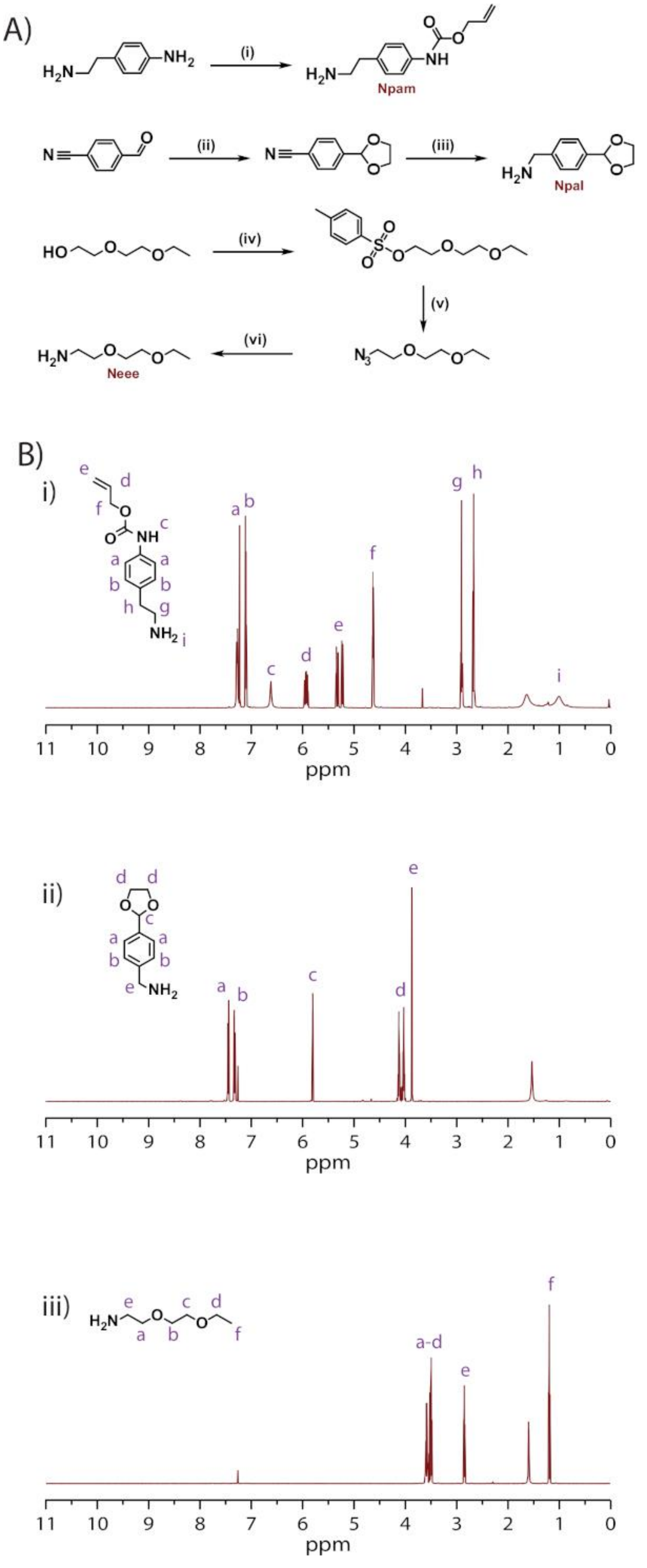
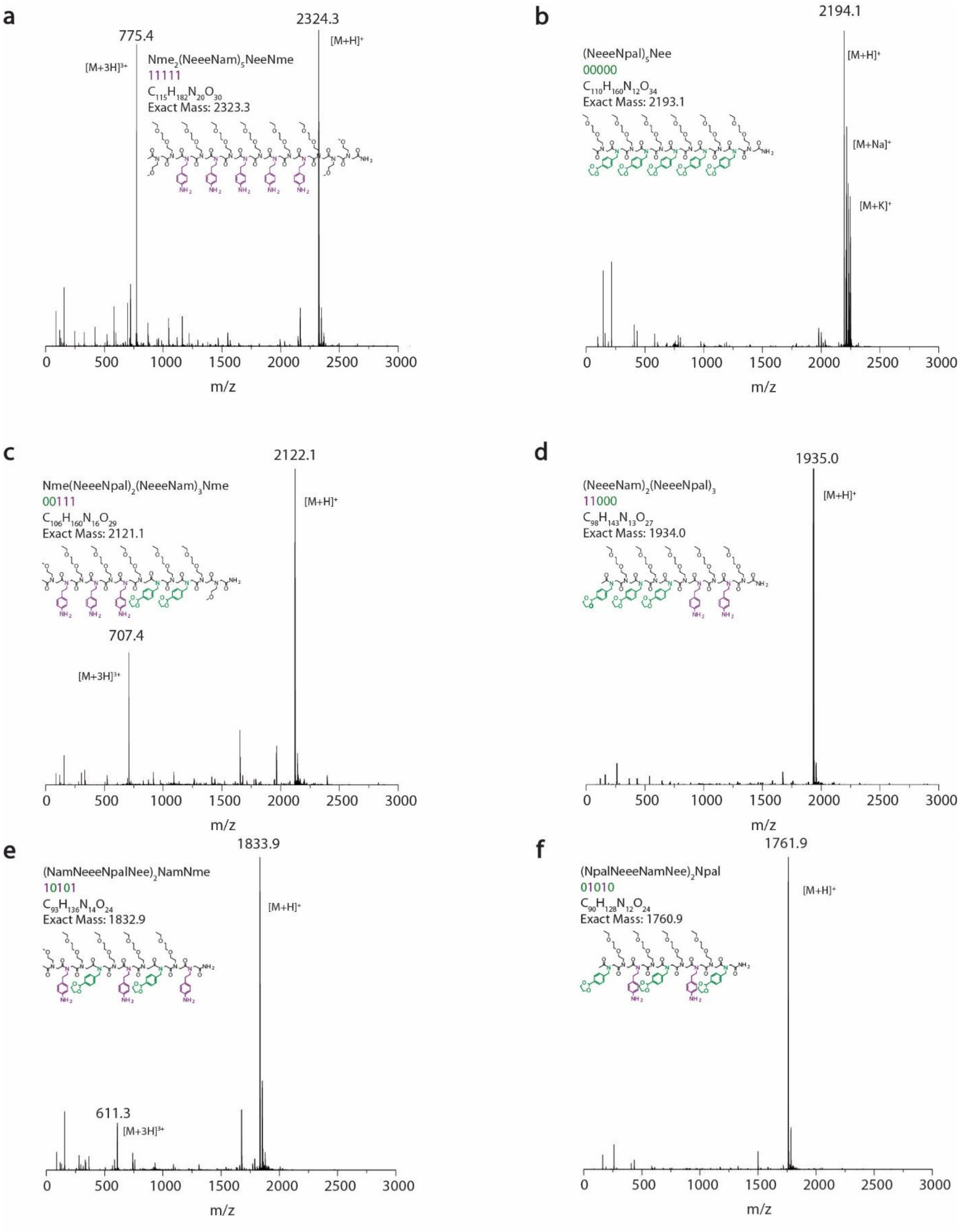

A)
(i)
Npam
(ii)
(iii)
Npal
(iv)
(v)
(vi)
Neee
B)
i)
a b
c
d
e
f
g h
i
ii)
a b
c
d
e
iii)
a-d
e
f
11 10 9 8 7 6 5 4 3 2 1 0
ppm

Figure 2.2 Monomer synthetic schemes and ^{1}H-NMR spectra. (A) Monomer synthetic schemes with reagents and conditions: (i) allyl chloroformate, 10% aqueous acetic acid, 1,4-dioxane, room temperature, overnight; (ii) ethylene glycol, toluene-p-sulfonic acid, toluene, reflux, overnight; (iii) $LiAlH_4$, anhydrous Et_2O, 0 °C for 4 h then room temperature for 12 h; (iv) tosyl chloride, THF, 0 °C; (v) NaN3, DMF, 60 °C, 36 h; (vi) triphenylphosphine, THF, overnight. (B) Monomer ^{1}H-NMR spectra (500 MHz, $CDCl_3$): (i) 4-(2-aminoethyl)-N-(allylcarbonyloxy)phenylamine (Npam); (ii) 4-(1,3-dioxacyclopent-2-yl)benzylamine (Npal); (iii) 2-(2-ethoxyethoxy)ethylamine (Neee).

2.4.3 Oligo(peptoid) Synthesis

Peptoids were synthesized *via* a sub-monomer approach to solid phase synthesis using Fmoc-Photolabile SS (0.1 mmol scale, 100-200 mesh, 1% DVB, Advanced ChemTech, Kentucky) as the resin. Syntheses were performed in an automated microwave synthesizer (Liberty Blue, CEM Corporation, North Carolina) adapted for peptoid synthesis. Resin was swelled at room temperature for 5 minutes with DMF before deprotection with 20% 4-methylpiperidine in DMF (v/v) for 30 s at 75°C and 90 s at 90°C. Subsequently, alternating additions of 1 mL of 1 M bromoacetic acid (coupled with 1 mL of 1 M *N*,*N′*-diisopropylcarbodiimide (DIC)) and 2.5 mL of 0.5 M primary amine monomer were repeated until desired sequence length was achieved. Primary amine monomers consisted of Neee as the principal spacer, an allyloxycarbonyl- (Alloc) protected amine (Npam protected, Nam unprotected), an ethylene acetal-protected aldehyde (Npal protected, Nal unprotected), and Nme as a mass label (Supplementary Table 1). The *N*-terminal of the oligomers were capped with 1 M acetic anhydride, activated by DIC. Fluorescein-functionalized oligomers were synthesized by omitting the acetic anhydride capping step and instead introducing a final coupling step by adding 1 mL of a 0.4 M 5,6-carboxyfluoroscein solution in DMF in addition to 1 mL 0.6 M HCTU in DMF and 1.5 mL of 0.8 M *N*,*N*-diisopropylethylamine in DMF. Oligo(peptoids) were subjected on-resin to 0.1 equivalents of tetrakis(triphenylphosphine) palladium(0) and 25 equivalents of phenylsilane per Alloc group in dry DCM for one hour to deprotect the Alloc-amines. The deprotection solution was filtered off, and peptoids were introduced to a fresh deprotection solution for another hour. Upon filtration of the second solution, the peptoids on-resin were submerged in 5 mL DMF and cleaved for 36 hours under irradiation at approximately 25 mW.cm^{-2} with 405 nm. Resin was separated from liberated peptoids with the use of a syringe filter. Solvent was removed under vacuum, and the oligomers were reconstituted in acetonitrile and further purified in preparative HPLC using a linear gradient of acetonitrile (MeCN) and water: (1) 30% MeCN, 0.1-2.1 min; (2) 30-95% MeCN, 2.1-16.1 min; (3) 95% MeCN, 16.1-23.1 min; (4) 95% MeCN, 23.1-26.1 min. Purified peptoids were lyophilized to yield off-white powder. Characterization of peptoid strands can be seen in Figures 2.3.-2.5.


a
775.4
2324.3
[M+3H]3+
[M+H]+
Nme2(NeeeNam)5NeeNme
11111
C115H182N20O30
Exact Mass: 2323.3
b
2194.1
(NeeeNpal)5Nee
00000
C110H160N12O34
Exact Mass: 2193.1
[M+H]+
[M+Na]+
[M+K]+
c
2122.1
Nme(NeeeNpal)2(NeeeNam)3Nme
00111
C106H160N16O29
Exact Mass: 2121.1
[M+H]+
707.4
[M+3H]3+
d
1935.0
(NeeeNam)2(NeeeNpal)3
11000
C98H143N13O27
Exact Mass: 1934.0
[M+H]+
e
1833.9
(NamNeeeNpalNee)2NamNme
10101
C93H136N14O24
Exact Mass: 1832.9
[M+H]+
611.3
[M+3H]3+
f
1761.9
(NpalNeeeNamNee)2Npal
01010
C90H128N12O24
Exact Mass: 1760.9
[M+H]+
NH2
0
500
1000
1500
2000
2500
3000
m/z

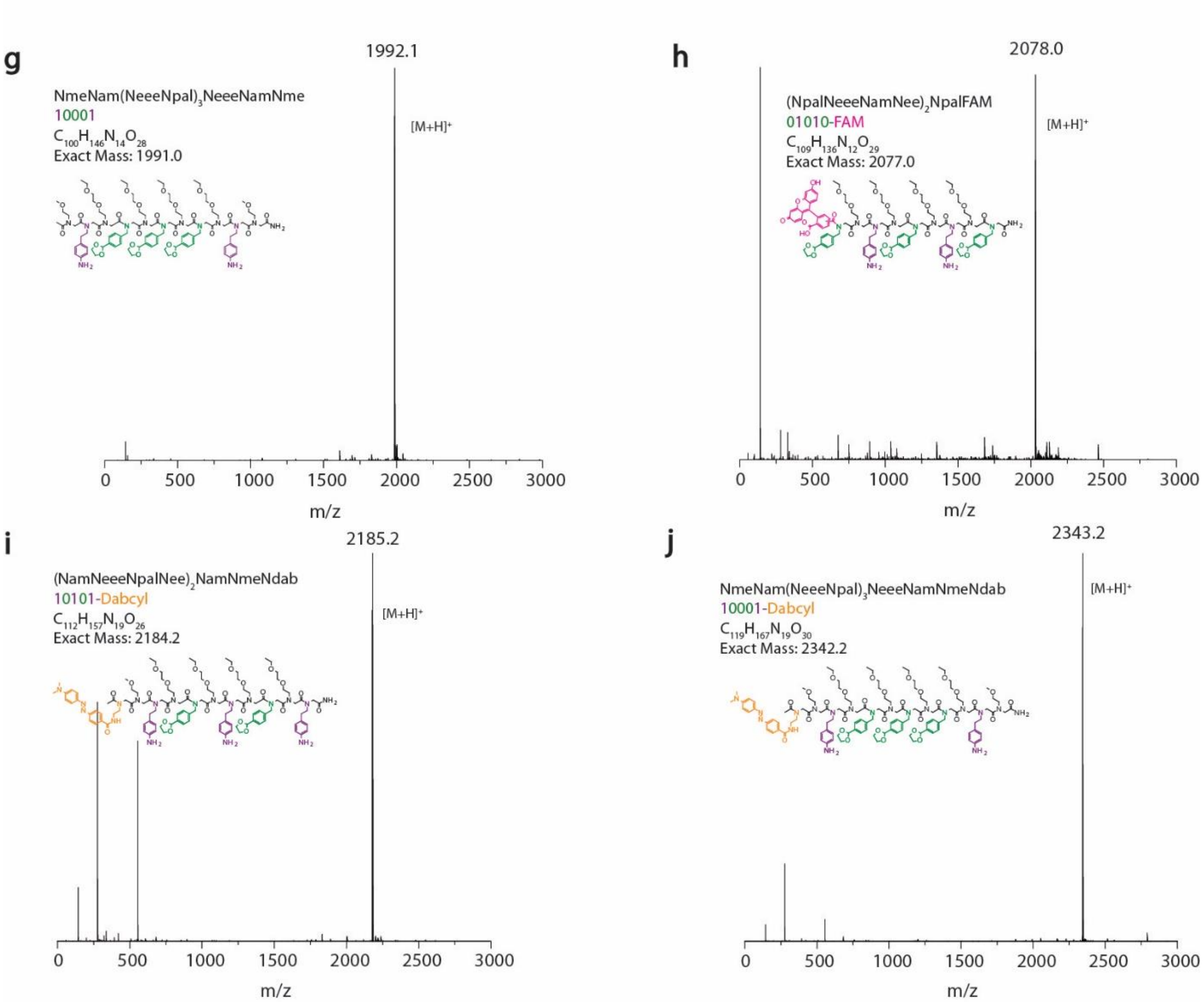
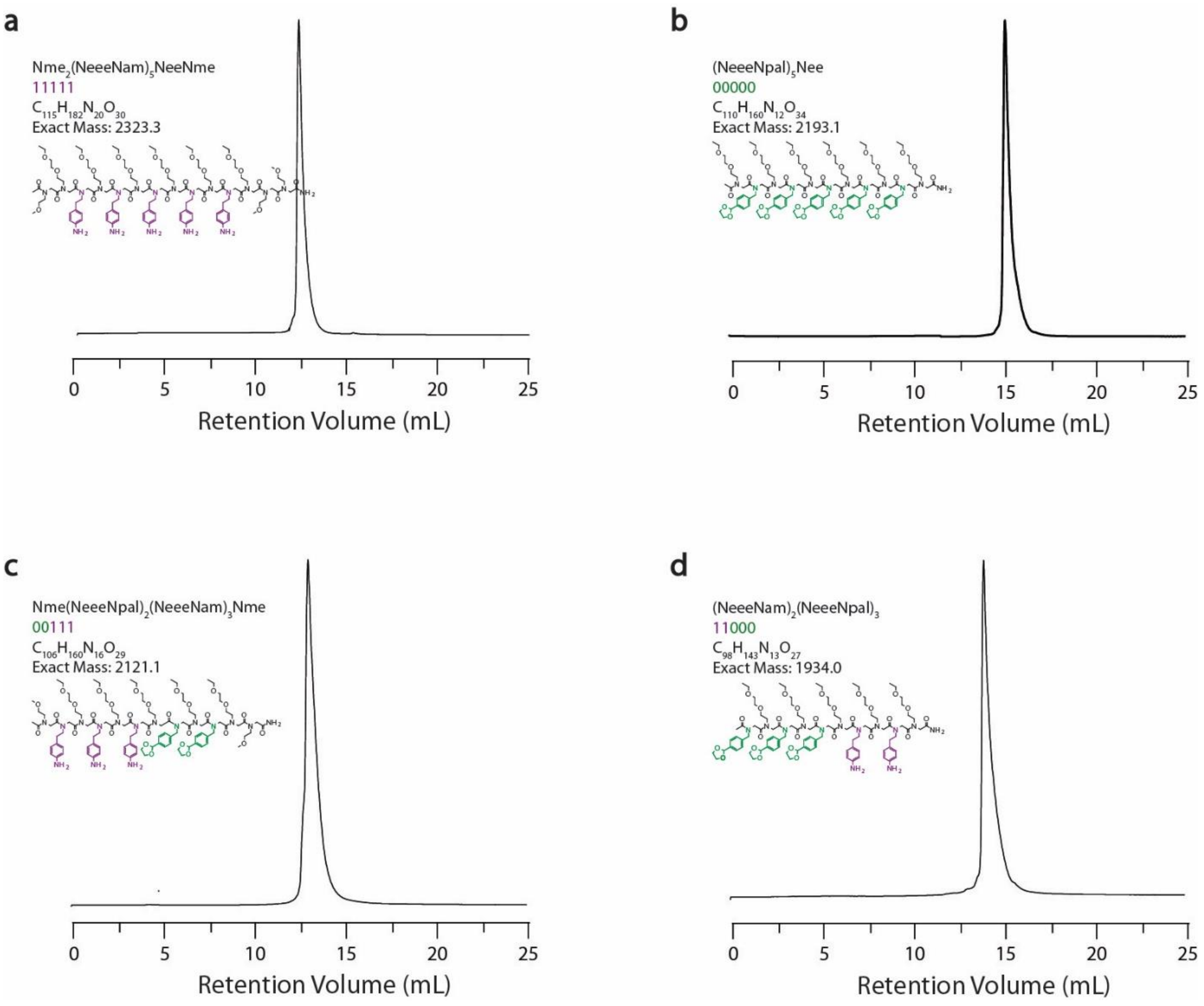


Figure 2.3 ESI mass spectra of purified peptoid sequences; a, Nme$_2$(NeeeNam)$_5$NeeNme; b, (NeeeNpal)$_5$Nee; c, Nme(NeeeNpal)$_2$(NeeeNam)$_3$Nme; d, (NeeeNam)$_2$(NeeeNpal)$_3$; e, (NamNeeeNpalNee)$_2$NamNme; f, (NpalNeeeNamNee)$_2$Npal; g, NmeNam(NeeeNpal)$_3$NeeeNamNme; h, (NpalNeeeNamNee)$_2$NpalFAM; i, (NamNeeeNpalNee)$_2$NamNmeNdab; j, NmeNam(NeeeNpal)$_3$NeeeNamNmeNdab.


a
Nme2(NeeeNam)5NeeNme
11111
C115H182N20O30
Exact Mass: 2323.3
NH2
Retention Volume (mL)
0
5
10
15
20
25
b
(NeeeNpal)5Nee
00000
C110H160N12O34
Exact Mass: 2193.1
NH2
Retention Volume (mL)
c
Nme(NeeeNpal)2(NeeeNam)3Nme
00111
C106H160N16O29
Exact Mass: 2121.1
NH2
Retention Volume (mL)
d
(NeeeNam)2(NeeeNpal)3
11000
C98H143N13O27
Exact Mass: 1934.0
NH2
Retention Volume (mL)

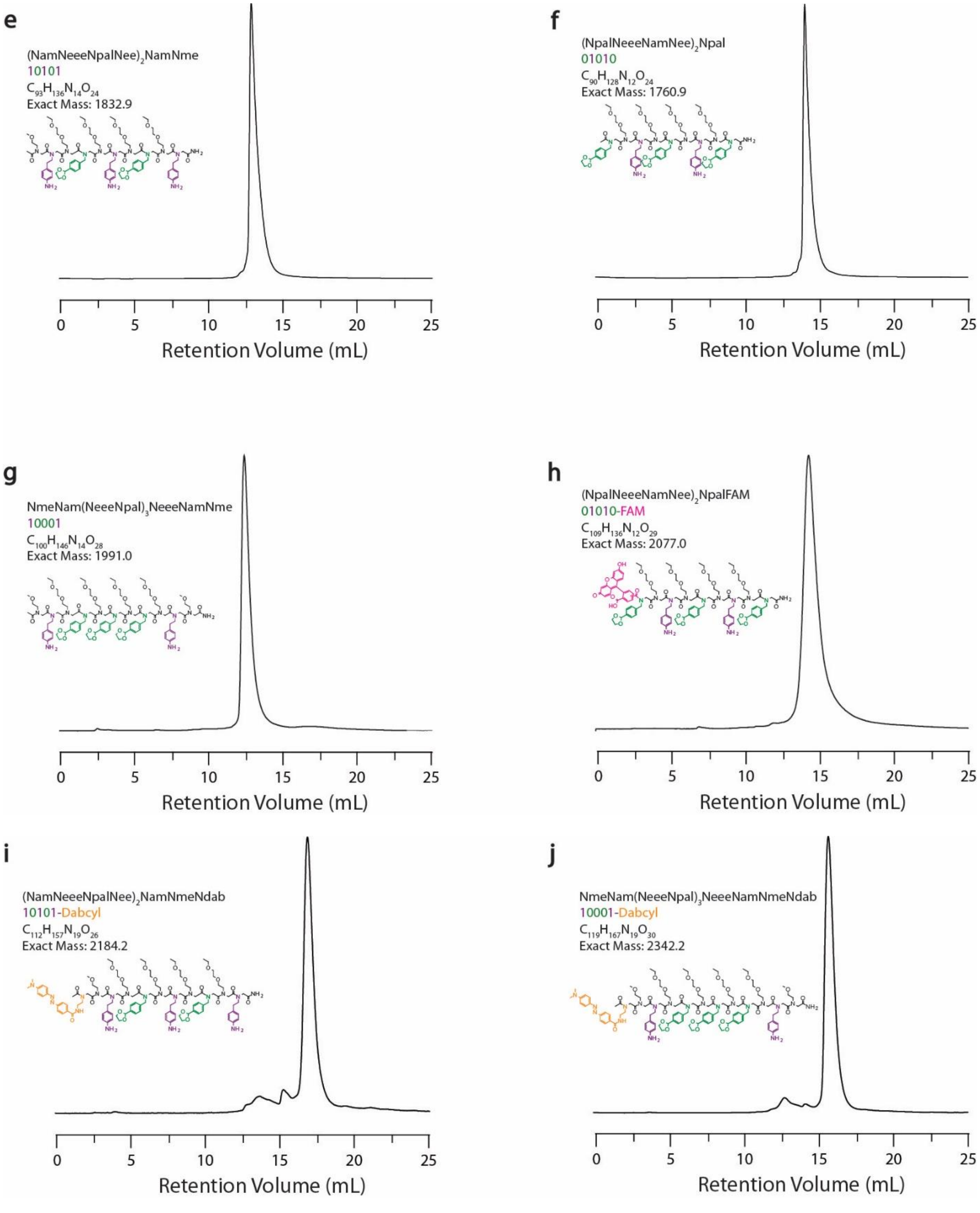
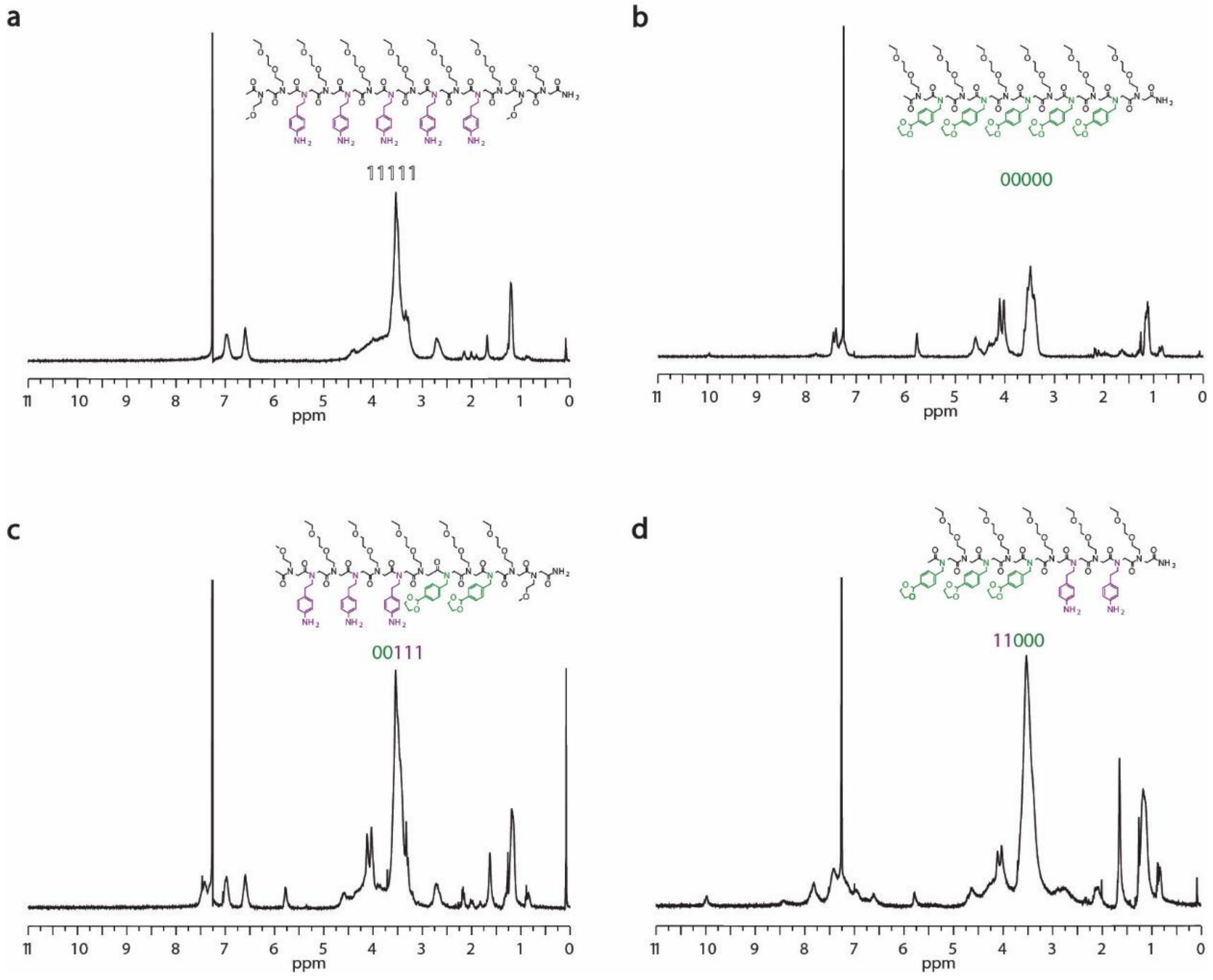


Figure 2.4 a-j, RP-HPLC traces of purified peptoid sequences. a, Nme$_2$(NeeeNam)$_5$NeeNme; b, (NeeeNpal)$_5$Nee; c, Nme(NeeeNpal)$_2$(NeeeNam)$_3$Nme; d, (NeeeNam)$_2$(NeeeNpal)$_3$; e, (NamNeeeNpalNee)$_2$NamNme; f, (NpalNeeeNamNee)$_2$Npal; g, NmeNam(NeeeNpal)$_3$NeeeNamNme; h, (NpalNeeeNamNee)$_2$NpalFAM; i, (NamNeeeNpalNee)$_2$NamNmeNdab; j, NmeNam(NeeeNpal)$_3$NeeeNamNmeNdab.


a
11111
11 10 9 8 7 6 5 4 3 2 1 0
ppm
b
00000
11 10 9 8 7 6 5 4 3 2 1 0
ppm
c
00111
11 10 9 8 7 6 5 4 3 2 1 0
ppm
d
11000
11 10 9 8 7 6 5 4 3 2 1 0
ppm
NH2

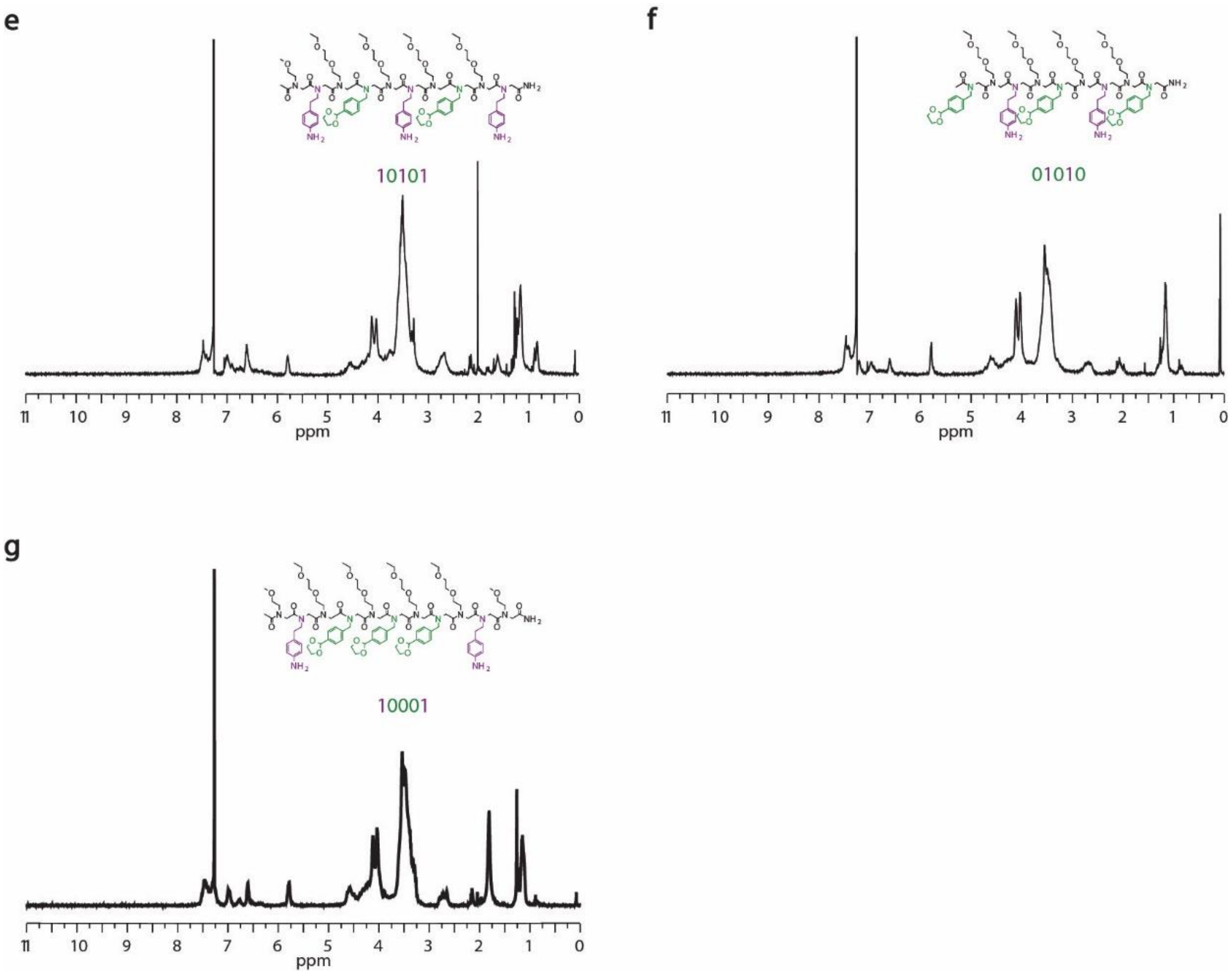


Figure 2.5 ^{1}H NMR spectrum (500 MHz, CDCl$_3$) of purified peptoid sequences.

a) **11111: δ (ppm): 6.85-7.05 (m, Ar), 6.45-6.70 (m, Ar), 3.00-4.53 (br, -N-C*H*$_2$-CO-, -O-C*H*$_2$-C*H*$_2$-O-, Ar-CH$_2$-C*H*$_2$-N-, -O-C*H*$_3$, -N-C*H*$_2$-C*H*$_2$-O-, -OC*H*$_2$-CH$_3$), 2.50-2.80 (br, Ar-C*H*$_2$-CH$_2$-N-), 1.58-1.73 (m, -CO-C*H*$_3$), 1.00-1.21 (m, -O-CH$_2$-C*H*$_3$)**

b) **00000: δ (ppm): 9.95-10.07 (m, Ar-C*H*O), 7.60-8.00 (m, Ar), 7.31-7.58(m, Ar), 5.76-5.80 (m, -Ar-C*H*-O$_2$-), 3.13-4.80 (m, Ar-C*H*$_2$-N-, -N-C*H*$_2$-C*H*$_2$-O-, -N-C*H*$_2$-CO-, -O-C*H*$_2$-C*H*$_2$-O-, -O-C*H*$_2$-CH$_3$, -(O-C*H*$_2$)$_2$-, Ar-C*H*$_2$-CH$_2$-N-), 2.06-2.18 (m, -CO-C*H*$_3$), 1.05-1.20 (m, -O-CH$_2$-C*H*$_3$)**

c) **00111: δ (ppm): 7.33-7.54 (m, Ar), 7.25-7.35 (m, Ar), 6.90-7.04 (m, Ar), 6.51-6.66 (m, Ar), 5.64-5.88 (m, -Ar-C*H*-O$_2$-), 3.13-4.86 (br, Ar-C*H*$_2$-N-, -O-C*H*$_2$-C*H*$_2$-O-, Ar-C*H*$_2$-CH$_2$-N-, -N-C*H*$_2$-C*H*$_2$-O-, -O-C*H*$_3$, -N-C*H*$_2$-CO-,-O-C*H*$_2$-CH$_3$, -(O-C*H*$_2$)$_2$-,), 2.50-2.84 (br, Ar-CH$_2$-C*H*$_2$-N-), 1.51-1.70 (m, -CO-C*H*$_3$), 0.99-1.39 (-O-CH$_2$-C*H*$_3$)**

d) **11000: δ (ppm): 9.96-10.01 (m, Ar-C*H*O), 7.63-7.97 (m, Ar), 7.33-7.54 (m, Ar), 6.85-7.02 (m, Ar), 6.46-6.70 (m, Ar), 5.68-5.85 (m, -Ar-C*H*-O$_2$-), 3.05-4.81 (br, Ar-C*H*$_2$-N-, -O-C*H*$_2$-C*H*$_2$-O-, Ar-C*H*$_2$-CH$_2$-N-, -N-C*H*$_2$-C*H*$_2$-O-, -O-C*H*$_3$, -N-C*H*$_2$-CO-,-O-C*H*$_2$-CH$_3$, -(O-C*H*$_2$)$_2$-), 2.52-2.99 (br, Ar-CH$_2$-C*H*$_2$-N-), 1.50-1.72 (m, -CO-C*H*$_3$), 0.96-1.31 (m, -O-CH$_2$-C*H*$_3$)**

e) 10101: δ (ppm): 7.36-7.52 (m, Ar), 7.24-7.34 (m, Ar), 6.88-7.08 (m, Ar), 6.49-6.66 (m, Ar), 5.72-5.86 (m, -Ar-C*H*-O$_2$-), 3.17-4.88 (br, Ar-C*H*$_2$-N-, -O-C*H*$_2$-C*H*$_2$-O-, Ar-C*H*$_2$-CH$_2$-N-, -N-C*H*$_2$-C*H*$_2$-O-, -O-C*H*$_3$, -N-C*H*$_2$-CO-,-O-C*H*$_2$-CH$_3$, -(O-C*H*$_2$)$_2$-), 2.54-2.87 (br, Ar-CH$_2$-C*H*$_2$-N-), 1.50-1.62 (m, -CO-C*H*$_3$), 0.99-1.33 (m, -O-CH$_2$-C*H*$_3$)

f) 01010: δ (ppm): 7.34-7.53 (m, Ar), 7.10-7.34 (m, Ar), 6.90-7.07 (m, Ar), 6.51-6.66 (m, Ar), 5.66-5.92 (m, -Ar-C*H*-O$_2$-), 3.15-5.18 (br, Ar-C*H*$_2$-N-, -O-C*H*$_2$-C*H*$_2$-O-, Ar-C*H*$_2$-CH$_2$-N-, -N-C*H*$_2$-C*H*$_2$-O-, -O-C*H*$_3$, -N-C*H*$_2$-CO-,-O-C*H*$_2$-CH$_3$, -(O-C*H*$_2$)$_2$-), 2.51-2.94 (br, Ar-CH$_2$-C*H*$_2$-N-), 1.50-1.62 (m, -CO-C*H*$_3$), 0.99-1.35 (m, -O-CH$_2$-C*H*$_3$)

g) 10001: δ (ppm): 7.32-7.55 (m, Ar), 7.21-7.34 (m, Ar), 6.91-7.01 (m, Ar), 6.55-6.63 (m, Ar), 5.70-5.83 (m, -Ar-C*H*-O$_2$-), 3.21-4.70 (br, Ar-C*H*$_2$-N-, -O-C*H*$_2$-C*H*$_2$-O-, Ar-C*H*$_2$-CH$_2$-N-, -N-C*H*$_2$-C*H*$_2$-O-, -O-C*H*$_3$, -N-C*H*$_2$-CO-,-O-C*H*$_2$-CH$_3$, -(O-C*H*$_2$)$_2$-), 2.58-2.82 (br, Ar-CH$_2$-C*H*$_2$-N-), 1.66-1.90 (m, -CO-C*H*$_3$), 1.05-1.32 (m, -O-CH$_2$-C*H*$_3$)

2.4.4 Peptoid Hybridization

Conventional Single-step Method

A vial was charged with 20 µL of 10 mM solutions of each of the desired oligo(peptoid) sequences in acetonitrile, 0.2 eq. of Sc(OTf)$_3$ per potential imine bond from a 10 mM stock solution, and sufficient solvent to afford 200 µL total of a 2% (v/v) of water/acetonitrile solution. The mixture stirred gently for two hours at 70°C.

Dissociation/Extraction/Annealing Method

Dissociation of peptoid sequences was achieved by adding a rare earth metal triflate (either 1.5 eq. of Sc(OTf)$_3$, Yb(OTf)$_3$, or Lu(OTf)$_3$, or 3.0 eq. of Y(OTf)$_3$ or Sm(OTf)$_3$ per potential imine bond) from a 10 mM stock solution to a 200 µL 2% (v/v) of water/acetonitrile solution containing 0.2 µmol of each desired oligo(peptoid) sequence and heating at 70°C. After heating for two hours, 200 µL of chloroform and 2 mL of water were added to the vial, followed by gentle shaking. The mixture was allowed to stand and, upon complete phase separation[30], the organic layer was extracted and stirred in a new vial at 70°C for two hours for oligomer annealing.

Sequence Identification Method

Sequence identification hybridizations were similarly performed by adding 15 µL of a 10 mM solution of ‘message’ peptoid to a vial containing 200 µL of a 2% (v/v) of water/acetonitrile solution, 0.2 µmol of each peptoid strand constituting the known library (01010, 11000, and 11111) and 0.225 µmol of $Sc(OTf)_3$. As before, this solution was heated at 70°C for two hours, then diluted with 200 µL of chloroform. After heating, 2 mL of water was added and the solution was gently shaken. Upon isolation of the chloroform layer, annealing was carried out at 70°C for two hours.

2.4.5 Förster Resonance Energy Transfer (FRET)

Quantification of hybridization selectivity was achieved *via* the dissociation/extraction/annealing of 1 mM hybridization mixtures initially consisting of distinct combinations of 01010-FAM, 10101, 10001, 10101-DABCYL, and 10001-DABCYL strands in stoichiometric ratios and 1.5 eq. $Sc(OTf)_3$ in 200 µL of a 2% (v/v) of water/acetonitrile solution. After annealing at 70°C for six hours and room temperature for one day, 200 µL of a pH 9 bicarbonate-carbonate buffer solution was added to each hybridization mixture. The hybridized solutions were further diluted to 1.2 mL with acetonitrile, then the fluorescence of the resultant solutions was recorded with a Horiba Quanta Master fluorimeter by exciting at 495 nm and measuring the fluorescence intensity at 525 nm. Triplicates of each solution were tested and normalized to a control mixture of 01010-FAM that was subject to the same dissociation/extraction/annealing process. Results are reported with standard error.

2.4.5 Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

Post-dissociation/extraction solutions of hybridized 11111×00000, 00111×11000, or 10101×01010 (initially 1 mM in 200 µL 2% (v/v) of water/acetonitrile and either 1.5 µmol of $Sc(OTf)_3$, $Yb(OTf)_3$, or $Lu(OTf)_3$, or 3.0 µmol of $Y(OTf)_3$ or $Sm(OTf)_3$) in $CHCl_3$ were dried under a steady stream of nitrogen and reconstituted in 1 mL of 2% nitric acid (aq). Samples were diluted 4 × 10^6-fold with *Milli-Q* water and the residual scandium concentrations were determined by ICP-MS using a Perkin-Elmer NexION 2000. Experiments on each solution were performed in triplicate and results are reported with standard errors.

2.5 Results and Discussion

2.5.1 Dynamic Covalent Assembly of Encoded Molecular Ladders

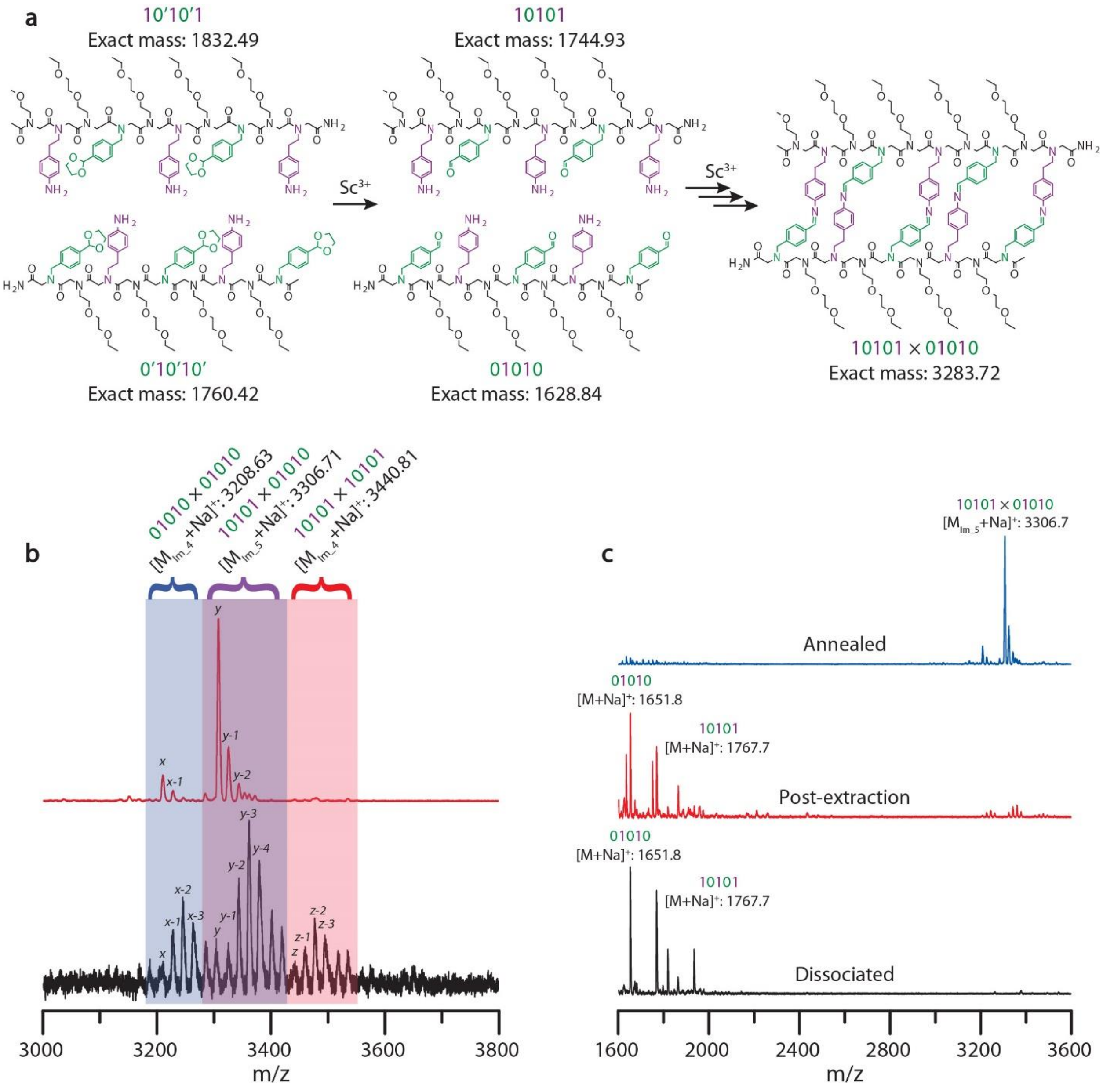

a
10’10’1
Exact mass: 1832.49
10101
Exact mass: 1744.93
Sc3+
Sc3+
0’10’10’
Exact mass: 1760.42
01010
Exact mass: 1628.84
10101 × 01010
Exact mass: 3283.72
b
01010 × 01010
[M_Im_4+Na]+: 3208.63
10101 × 01010
[M_Im_5+Na]+: 3306.71
10101 × 10101
[M_Im_4+Na]+: 3440.81
y
x
x-1
y-1
y-2
y-3
y-4
y-2
x-2
x-1
x-3
y-1
y
x
z-2
z-3
z-1
z
3000
3200
3400
3600
3800
m/z
c
10101 × 01010
[M_Im_5+Na]+: 3306.7
Annealed
01010
[M+Na]+: 1651.8
10101
[M+Na]+: 1767.7
Post-extraction
01010
[M+Na]+: 1651.8
10101
[M+Na]+: 1767.7
Dissociated
1600
2000
2400
2800
3200
3600
m/z

Figure 2.6 Dynamic covalent assembly of encoded molecular ladders. a, Structures of complementary, mass-labeled precursor peptoids and desired, in-registry molecular ladder product. b, MALDI mass spectra of molecular ladder reaction mixtures generated either after a single-step, deprotection and direct assembly process (bottom, black), or *via* a stepwise, dissociation/extraction/annealing assembly process (top, red). Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs (e.g., *y* = in-registry, 5-rung 10101×01010 molecular ladder, *y*-1 = out-of-registry, 4-rung 10101×01010 molecular ladder, etc.). c, MALDI mass spectra of the reaction mixture initially containing 10101 and 01010 sequences and a high loading of $Sc(OTf)_3$ (bottom, black), immediately after aqueous extraction (middle, red), and after annealing at 70°C for 2 hours (top, blue). Expected exact masses: $[M_{01010}+Na]^+$ = 1651.83; $[M_{10101}+Na]^+$ = 1767.92; $[M_{01010\times01010}+Na]^+$ (x) = 3208.63; $[M_{10101\times01010}+Na]^+$ (y) = 3306.71; $[M_{10101\times10101}+Na]^+$ (z) = 3440.81.

Peptoids (i.e., oligo(*N*-substituted glycine)s)[30] were employed here as the precursor oligomers owing to their ready synthetic accessibility *via* Zuckerman's sub-monomer solid phase synthetic method[31]. Moreover, the peptoid backbone is inherently neutral, achiral, and lacks hydrogen bond donor sites[32], ensuring that oligomer hybridization selectivity is dominated by covalent interactions between pendant groups. These oligomers were synthesized as binary sequences of amine and aldehyde pendant groups (denoted here as '1' and '0', respectively) alternating with inert spacer residues such that, owing to the zig-zag 'Σ-strand' conformation adopted by the constituent peptoid chains, the reactive groups were presented on the same side of the peptoid backbone[33]. The spacer residues also increased solubility of both the precursor oligomers and hybridized molecular ladders, preventing precipitation during the self-assembly process. To enable facile identification of the hybridization product species by mass spectrometry, sequences were mass-tagged by adding extra, inert spacer residues to the peptoid ends during synthesis. Additionally, each imine bond generated by an amine/aldehyde condensation reaction reduces the molar mass of the resultant molecular ladder by 18 g/mol owing to the loss of a water molecule, enabling in- and out-of-registry species to be readily distinguished by mass. As the oligomer sequences bore multiple, covalently coreactive functional group types, ethylene acetals were used as aldehyde protecting groups to preclude premature reaction between coreactive pendant groups during synthesis and purification, and which could then be deprotected *in situ*, allowing the dynamic covalent assembly reaction to proceed. Monomer residue structures are as shown in Figure 2.2. and purified, acetal-protected peptoid sequences were characterized as shown in Figures 2.3., 2.4., and 2.5..

Sequence-selective, dynamic covalent oligomer hybridization was initially examined using a model system consisting of two complementary peptoid sequences, each bearing alternating amine- and aldehyde-based reactive pendant groups (i.e., sequences 10101, mass-labeled with an additional methoxyethylamine residue at its *N*-terminal end, and 01010, see Figure 2.6.a). Our group has previously demonstrated the utilization of scandium(III) triflate at low loadings as a dual-role, Lewis acidic catalyst both to effect the *in situ* deprotection of oligomer-borne, acetal-protected aldehydes and to catalyze transimination and imine metathesis reactions for the rearrangement of imine bonds generated by amine/aldehyde condensation reactions, thereby successfully yielding molecular ladders with up to eight rungs *via* the dynamic covalent assembly of diblock, amine- and aldehyde-bearing precursor oligomers. Consequently, we applied the previously-determined reaction conditions (i.e., 0.2 equivalents of $Sc(OTf)_3$ for 2 hours at 70°C)[34] to the model system examined here.

For this system, the anticipated, thermodynamically-favored product should emerge from the cross-reaction between the 10101 sequence and its 01010 oligomeric complement (i.e., 10101×01010) to selectively afford the formation of an in-registry molecular ladder with five imine-based rungs (denoted here as Im_5). As determined by MALDI mass spectrometry (Figure 2.6.b, bottom spectrum), the applied reaction conditions did yield complete acetal deprotection; however, these conditions resulted in the non-selective, kinetically-trapped generation of multiple in- and out-of-registry molecular ladder species resulting from the incomplete coreaction of two

01010 (01010×01010) and two 10101 (10101×10101) sequences, both of which can yield molecular ladders with a maximum of four imine rungs (i.e., Im_4), in addition to those from the 10101×01010 cross-reaction.

Our previous work on the Sc(III)-catalyzed dynamic covalent assembly of imine-based molecular ladders revealed that the assembly reaction proceeds *via* a hybrid mechanism, where amine- and aldehyde-bearing oligomers initially associate and bind through condensation reactions at any point along the backbone to form out of registry molecular ladders, whereupon transimination and imine metathesis reactions proceed to shuffle the imine bonds until the molecular ladders come into registry[35]. Whereas this mechanism enabled the annealing of out-of-registry, intermediate species for simple systems composed of homo- or diblock oligomers, more complex sequences where the reactive residue type switches between amine and aldehyde pendant groups multiple times impede inter-strand bond shuffling as imine rearrangement reactions can no longer occur exclusively between adjacent residues. Rather, rearrangement reactions for such sequences would need to proceed between relatively remote functionalities to enable inter-strand shuffling. This registry mechanism is dissimilar to that commonly employed for DNA assembly where, instead of inter-strand shuffling, dissociated DNA sequences selectively hybridize with their complements upon gradual cooling from raised temperature to afford in-registry double helices at low error rates[8]. Pioneering work by the Lehn group on the catalysis of imine rearrangement reactions demonstrated that, at moderate concentrations, Lewis acidic rare earth metal triflates do act to catalyze imine connectivity rearrangement, but at raised concentrations they shift the amine/aldehyde condensation reaction equilibrium towards the unpaired reactants[36]. Analogous to the thermal cycling commonly employed for nucleic acid melting and annealing, raising and lowering the scandium triflate concentration was thus used to mitigate kinetic trapping by enabling quantitative dissociation and subsequent, sequence-selective assembly of the 10101 and 01010 peptoid sequences to afford the desired 10101×01010 molecular ladder product (see Figure 2.6.b, top spectrum).

This process is detailed in Figure 2.6.c where MALDI mass spectra collected on reaction mixture aliquots after each process step are presented. The original reaction mixture initially contained a stoichiometric amount of ethylene acetal-protected 10101 and 01010 sequences and a high loading (i.e., 1.5 equivalents per amine/aldehyde pair) of $Sc(OTf)_3$. Under these reaction conditions, the acetals groups were quantitatively deprotected to reveal amine-reactive aldehydes; nevertheless, the peptoid strands remained dissociated, evident from the presence of deprotected single strands at 1651.8 and 1767.7 m/z and virtual absence of paired species from 3200-3600 m/z, owing to the influence of $Sc(OTf)_3$ on the reaction equilibrium (Figure 2.6.c, bottom black spectrum). This reaction mixture was subject to an aqueous extraction to selectively extract a fraction of the $Sc(OTf)_3$, reducing its relative loading. The residual scandium concentration was determined by inductively coupled plasma mass spectrometry (ICP-MS), revealing that 0.055 ± 0.002 equivalents of Sc^{3+} remained post-extraction, approaching the optimal concentration for transimination catalysis reported previously[36]. Characterization of the reaction mixture by mass spectrometry immediately after extraction indicated that the unpaired single strands remained dominant, although non-negligible amounts of intermediate hybridized species were observable (Figure 2.6.c, middle red spectrum). Owing to the chemical and molecular weight similarities between in- and out-of-registry ladders, ionization efficiencies in MALDI mass spectrometry are presumed sufficiently similar to permit relative concentrations of the individual species to be estimated[35,37,38]. Heating for two hours at 70°C to anneal the system yielded the in-registry 10101×01010 molecular ladder as the primary reaction product (Figure 2.6.c, top blue spectrum).

Small peaks attributable to 01010×01010 ladder species remained after annealing owing to a slight stoichiometric imbalance that arose during the extraction step where some of the 10101 sequence was lost in the aqueous extraction. Additionally, trace amounts of species at +18 m/z (i.e., 4-rung ladders) were attributable to the residual Sc^{3+} affecting the reaction equilibrium rather than kinetically-trapped, out-of-registry products. Notably, when the system was annealed at room temperature, fully in-registry 10101×01010 ladder species were observed by MALDI mass spectrometry within one day and emerged as the major peak within seven days (Figure 2.7.) indicating that, although raised temperatures increase reaction rates as expected, the self-assembly process does proceed under ambient conditions. Gel permeation chromatography (GPC), calibrated using low dispersity polystyrene standards, confirmed formation of the 10101×01010 molecular ladder as the predominant product (43.8%) by comparing elution volumes of inert, Alloc-protected single strands to hybridization mixtures (Figure 2.8.a-d). Notably, this yield is lower than previously determined yields for imine rung-based molecular ladders assembled from homoblock oligomeric sequences[39] as these oligomers are incapable of participating in intramolecular amine/aldehyde condensation reactions, thus eliminating a potential mechanism for side product generation. Peak fractions, collected during GPC purification, were characterized by MALDI mass spectrometry to identify the fraction constituents (Figure 2.8.e). The hybridization product was preceded by several minor peaks at higher molecular weights, attributable to multimeric complexes, and proceeded by single-stranded species (Figure 2.8.f and 2.8.g), including unreacted 10101 and 01010 strands and those with either one intramolecular imine bond to yield macrocycles, or two intramolecular imine bonds to potentially afford hairpin-like structures. The occurrence of a binding event was further verified with diffusion ordered spectroscopy (DOSY) NMR, which demonstrated a decrease in the diffusion coefficient after extraction and annealing of an initially dissociated mixture containing 10101 and 01010 strands (Figure 2.9.).

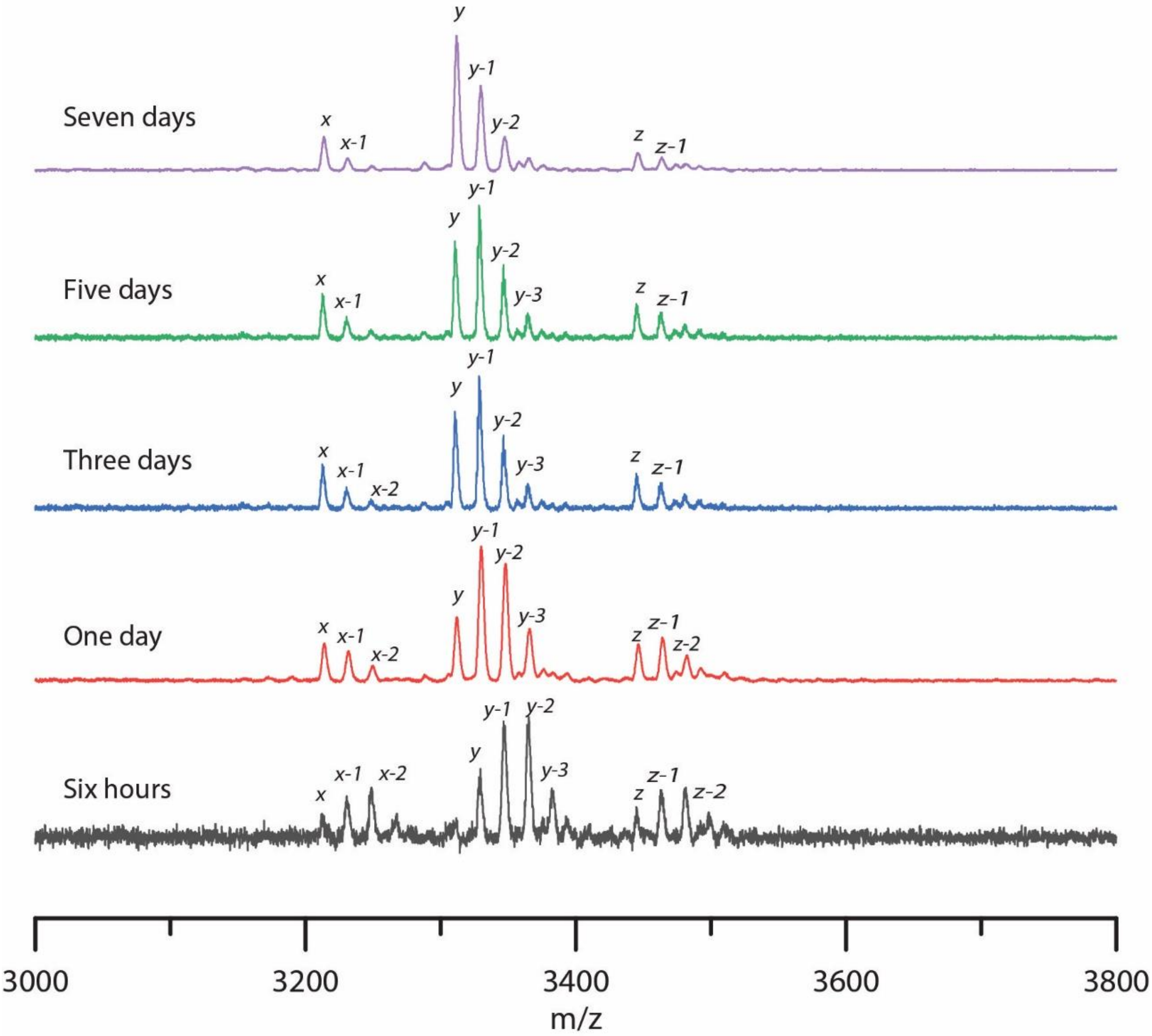


Figure 2.7 MALDI mass spectra of a 10101×01010 molecular ladder reaction mixture annealed at room temperature over a seven-day period. Expected exact masses: $[M_{01010\times01010}+Na]^+$ (*x*) = 3208.63; $[M_{10101\times01010}+Na]^+$ (*y*) = 3306.71; $[M_{10101\times10101}+Na]^+$ (*z*) = 3440.81. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs (e.g., *y* = in-registry, 5-rung 10101×01010 molecular ladder, *y*-1 = out-of-registry, 4-rung 10101×01010 molecular ladder, etc.).

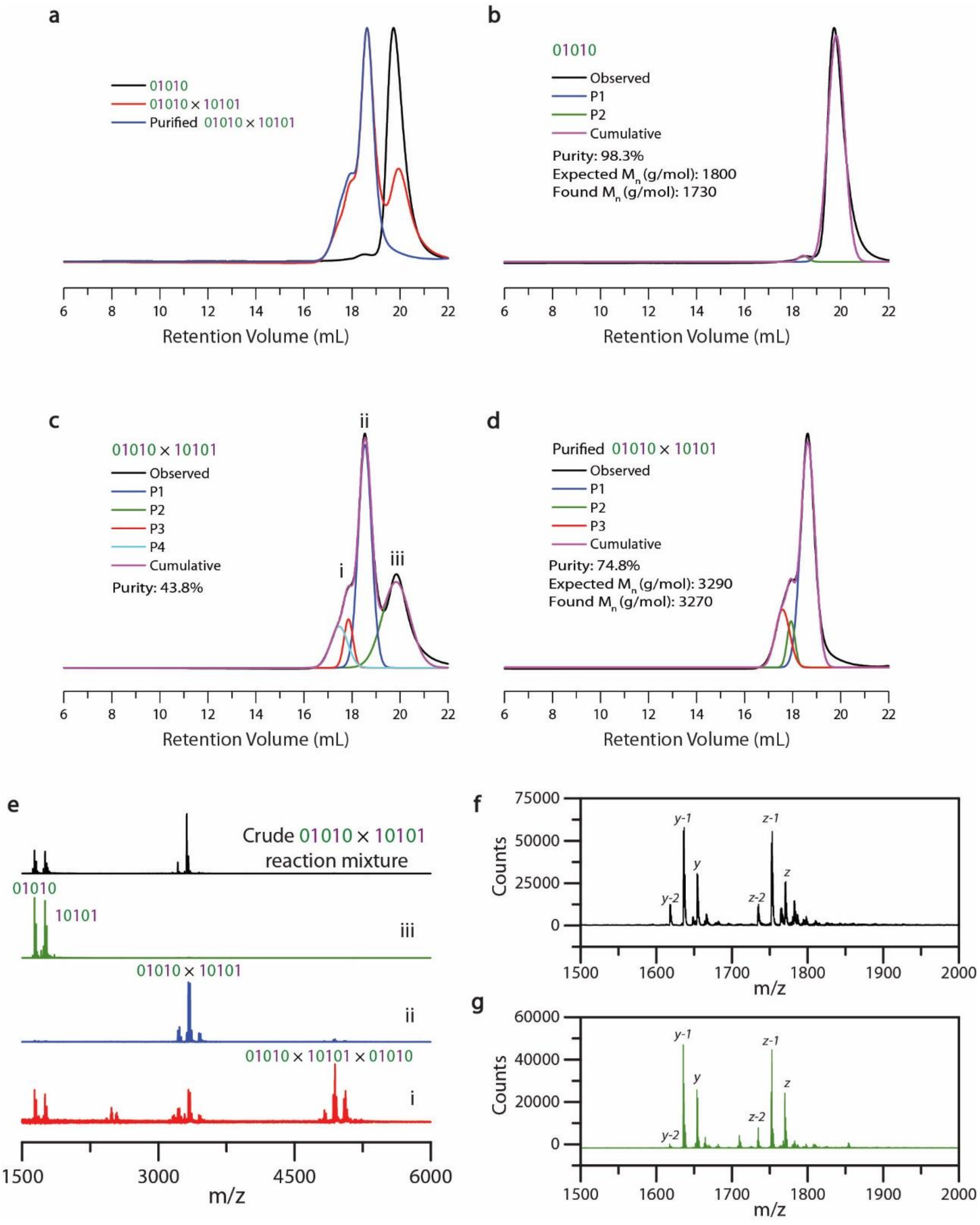


Figure 2.8 a, Superimposed GPC traces of an Alloc-protected, inert single strand mixture, a 10101×01010 molecular ladder mixture after performing the dissociation/extraction/annealing method, and a post-purification 10101×01010 molecular ladder. Each trace was deconvoluted (b, c, and d) by fitting Gaussian functions to simulated peaks with OriginPro. Upon subtracting the baseline, each peak over 0.5% of the max

intensity was utilized to fit the spectra. e, MALDI mass spectra of the different fractions (i, ii, and iii) collected from a 10101×01010 molecular ladder mixture after GPC purification. The peaks in the molecular weight region between 4750 and 5250 correspond to 01010×01010×01010, 01010×01010×10101, and 01010×10101×10101 triple ladder species. MALDI mass spectra expanded to highlight the m/z range expected for single-stranded oligomeric species of f, the crude 01010×10101 reaction mixture, and g, GPC fraction iii. Expected exact masses: $[M_{01010}+Na]^+$ (*y*) = 1651.83; $[M_{10101}+Na]^+$ (*z*) = 1767.92. Peaks at multiples of -18 m/z values are attributable to single-stranded species incorporating intramolecular imine bonds (e.g., *y* = 01010, *y*-1 = 01010 with one intramolecular imine bond, *y*-2 = 01010 with two intramolecular imine bonds, etc.).

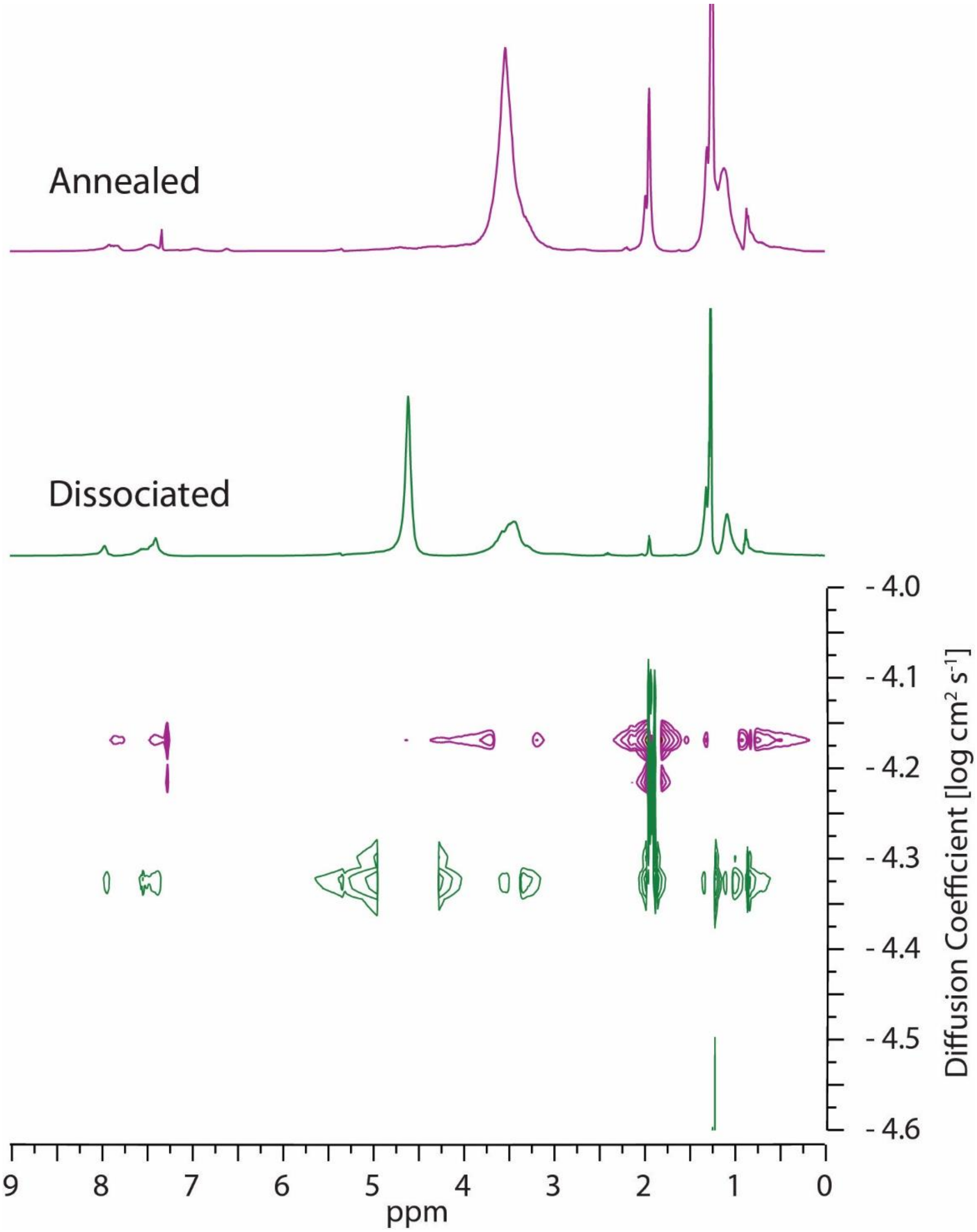


Figure 2.9 DOSY NMR spectra of a 10101×01010 molecular ladder mixture upon dissociation with 1.5 eq of scandium triflate (green) and subsequent binding upon extracting the scandium triflate and annealing at 70°C (purple).

To examine the effect of Sc^{3+} concentration on molecular ladder dissociation, varying amounts of scandium triflate were added to post-annealed, in-registry 10101×01010 reaction mixtures which were then heated at 60°C for six hours, left to stand at room temperature overnight to ensure system equilibration, and characterized by mass spectrometry (Figure 2.10.). At 0.20 equivalents of $Sc(OTf)_3$ (matching the Sc^{3+} concentration employed for the single-pot, single-step assembly method), the desired in-registry molecular ladder remained the major product; however, peaks at multiples of +18 m/z, attributable to ladders species with fewer rungs, were more apparent, suggesting a significant shift in the reaction equilibrium towards the amine and aldehyde reactants. At 0.30 equivalents of Sc^{3+} equivalents and above, the extent of ladder dissociation progressively increases and peaks attributable to dimerized species of mismatched, non-complementary strands (i.e., those consisting of 01010×01010 and 10101×10101 sequences) become more prevalent.

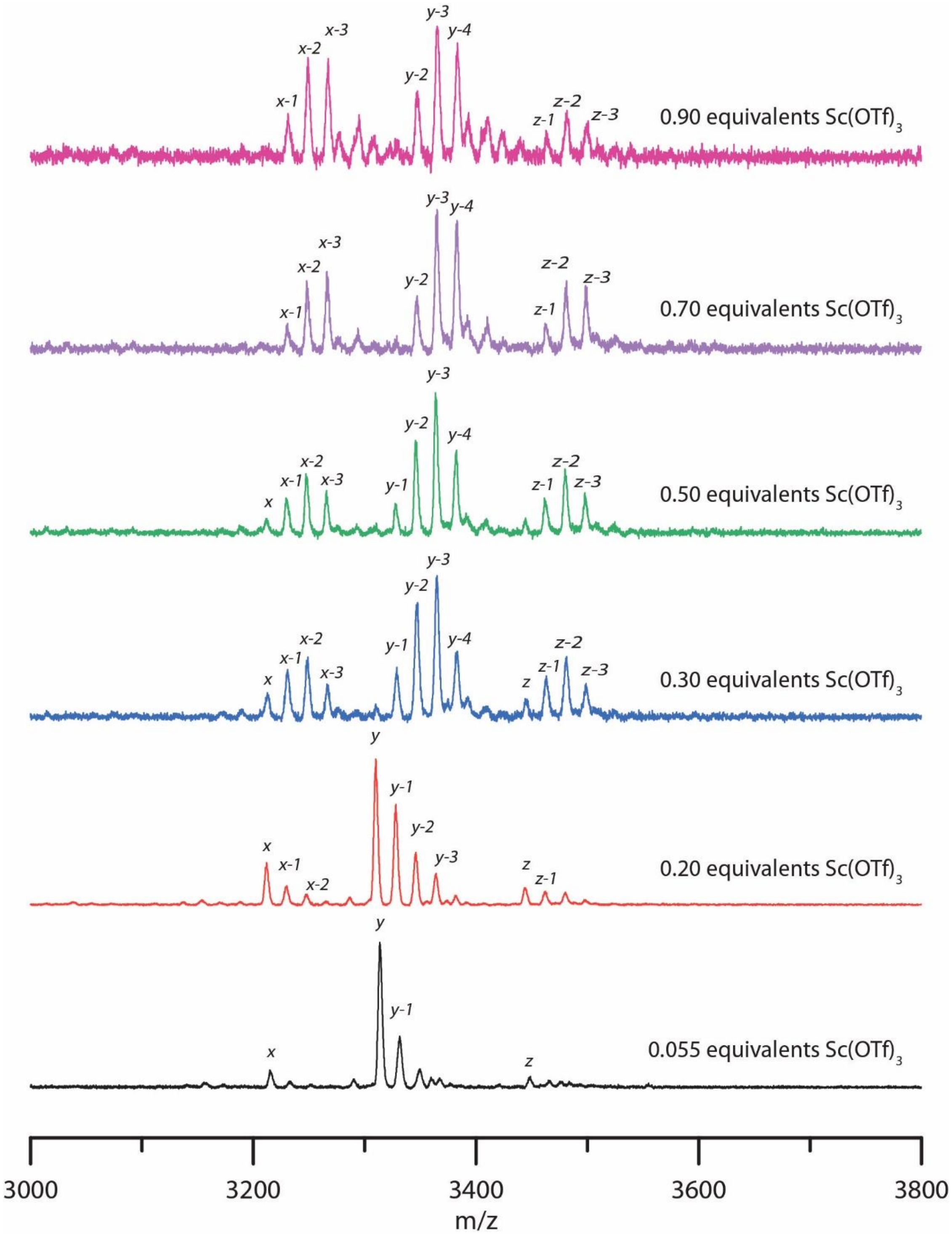


Figure 2.10 MALDI mass spectra of 10101×01010 molecular ladders treated with additional Sc(OTf)$_3$ post-annealing and allowed to equilibrate. Expected exact masses: [M$_{01010\times01010}$+Na]$^+$ (*x*) = 3208.63; [M$_{10101\times01010}$+Na]$^+$ (*y*) = 3306.71; [M$_{10101\times10101}$+Na]$^+$ (*z*) = 3440.81. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs (e.g., *y* = in-registry, 5-rung 10101×01010 molecular ladder, *y*-1 = out-of-registry, 4-rung 10101×01010 molecular ladder, etc.)

To examine the robustness of this dissociation/extraction/annealing assembly approach to alternative, Lewis acidic multi-role reagents, reaction mixtures of 10101 and 01010 oligomeric sequences were formulated with one of a library of rare-earth metal triflates with varying Lewis acidity, including $Yb(OTf)_3$, $Lu(OTf)_3$, $Y(OTf)_3$, and $Sm(OTf)_3$ (Figure 2.11.). Whereas ytterbium triflate is more Lewis acidic than lutetium triflate, the annealing period required for in-registry 10101×01010 molecular ladder equilibrium was at least three days at 70°C for the ytterbium triflate-containing reaction mixture while overnight at 70°C was sufficient to yield in-registry ladders for the mixture formulated with lutetium triflate. The residual rare-earth metal triflate in the reaction mixtures post-extraction was assessed by ICP-MS and revealed that a greater concentration of lutetium triflate remained in the organic layer than that of ytterbium triflate (0.014±0.0008 eq. and 0.009±0.0004 eq. respectively, indicating that the higher residual concentration of Lu^{3+} more than compensated for its lower Lewis acidity than Yb^{3+}. Notably, again owing to their lower Lewis acidity, three equivalents of yttrium triflate and samarium triflate (i.e., twice the required equivalents of Sc^{3+}, Yb^{3+}, and Lu^{3+}) were necessary to fully deprotect and dissociate the oligomeric strands. After extraction, in-registry molecular ladder equilibrium was achieved after seven days at 70°C for yttrium triflate-containing solution, while the reaction mixture incorporating samarium triflate required an additional two days for complete ladder assembly. Here, ICP-MS revealed residual, post-extraction concentrations of 0.025±0.001 eq. and 0.059±0.003 eq. for Y^{3+} and Sm^{3+}, respectively).

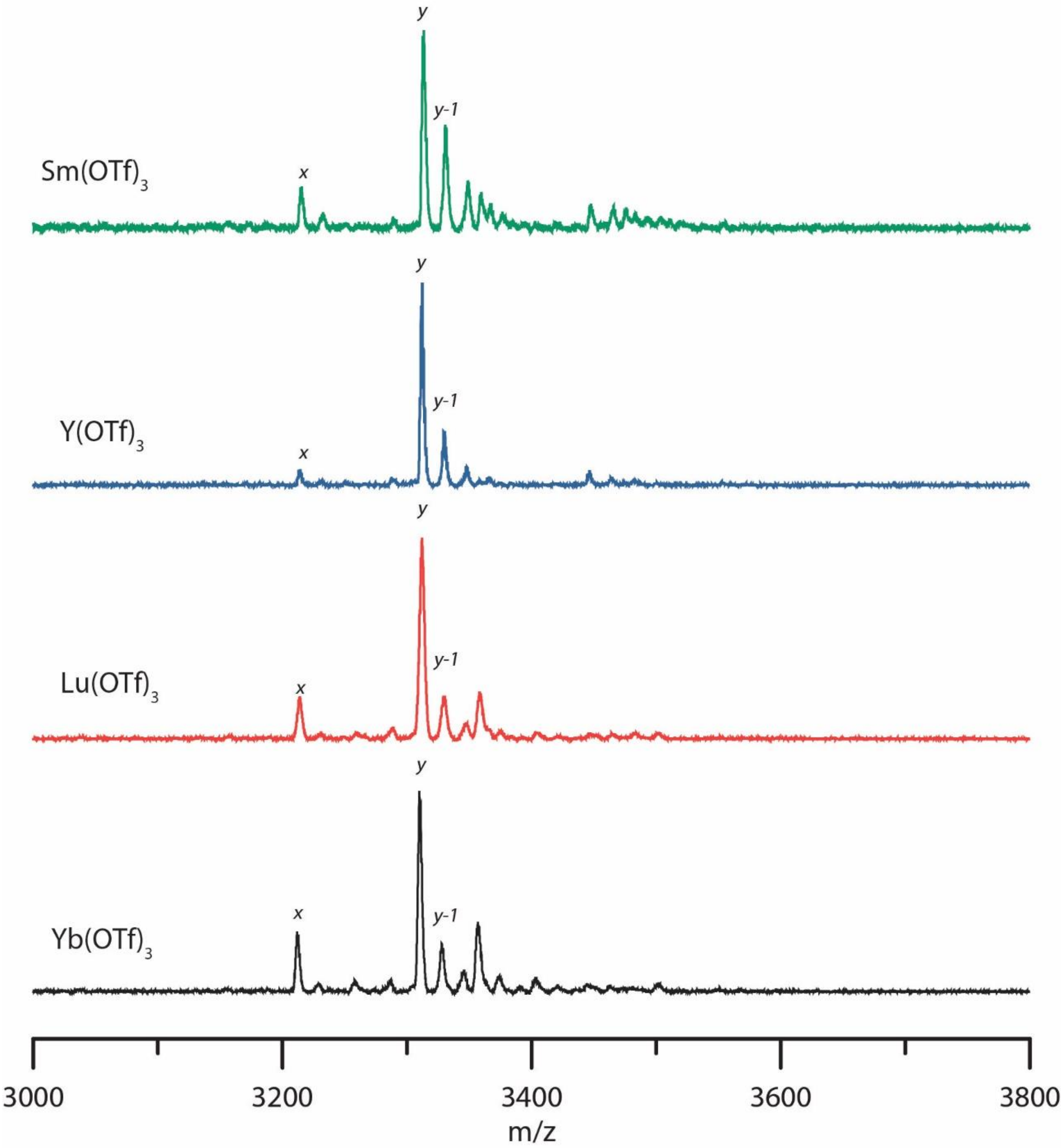


Figure 2.11 MALDI mass spectra of 10101×01010 molecular ladders *via* the dissociation/extraction/ annealing process challenged against a library of rare-earth metal triflates. Expected exact masses: $[M_{01010\times01010}+Na]^+$ (*x*) = 3208.63; $[M_{10101\times01010}+Na]^+$ (*y*) = 3306.71. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs (e.g., *y* = in-registry, 5-rung 10101×01010 molecular ladder, *y*-1 = out-of-registry, 4-rung 10101×01010 molecular ladder, etc.)

2.5.2 Competitive Dynamic Covalent Assembly

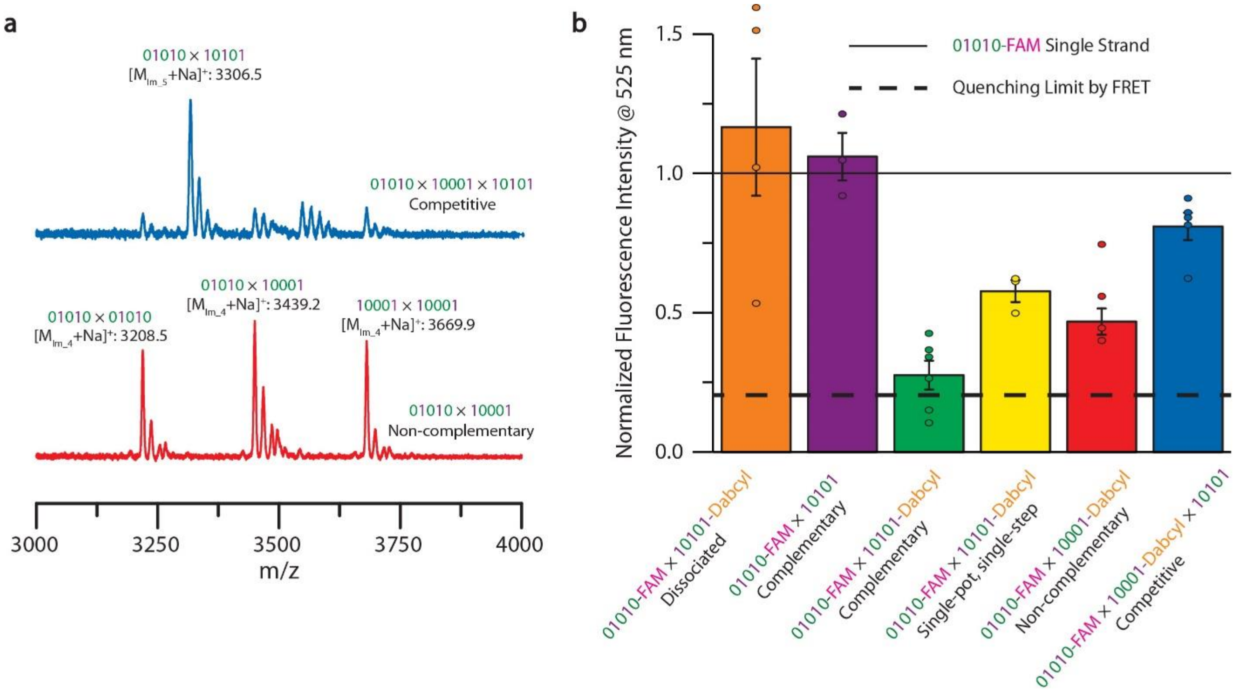


Figure 2.12 Competitive dynamic covalent assembly. a, MALDI mass spectra of reaction mixtures generated after applying the dissociation/extraction/annealing process to non-complementary 01010 and 10001 sequences (bottom, red) and to a ternary mixture of a competitive environment consisting of a complementary sequence and a non-complementary sequence *via* the dissociation/extraction/annealing process (top, blue). Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs. b, Sequence-selective hybridization assessed by FRET and normalized to the fluorescence intensity of a 01010-FAM sequence single strand. Assembly solutions were excited at 495 nm and fluorescence emission was measured at 425 nm. Error bars represent standard error. The horizontal solid and dashed lines 1 and 0.2, respectively, represent the unquenched 01010-FAM control and the reported FAM/DABCYL fluorescence quenching efficiency of 80%[40], respectively.

The single reactant pair mismatch discrimination of our dynamic covalent self-assembly process was examined using combinations of peptoid strands, including the complementary pair of 10101 and 01010, and a 10001 sequence that is non-complementary with the other two. Importantly, the 10001 and 01010 sequences contain only a single mismatch located at the third residue to preclude any overhangs that might otherwise occur from influencing the 10001×01010 hybrid stability. Subjecting a non-complementary mixture of 10001 and 01010 strands in a 1:1 stoichiometric ratio to the dissociation/extraction/annealing reaction conditions yielded three cohorts of dimerized ladder species, 10001×10001, 01010×01010, and 01010×10001 (Figure 2.12.a, bottom red spectrum), demonstrating poor hybridization selectivity. In contrast, upon inclusion of a stoichiometric amount of the 10101 sequence to the initial reaction mixture to afford a ternary sequence system where two of the peptoid strands are complementary (i.e., 10101 and 01010) results in the preferential generation of a complementary and in-registry 10101×01010 molecular ladder as the major mass spectrum peak (Figure 2.12.a, top blue spectrum).

Additional quantification of the hybridization selectivity was performed using peptoid strands functionalized with 5,6-carboxyfluorescein (FAM) and 4-((4-(dimethylamino)phenyl)azo)benzoyl (DABCYL) groups as a Förster resonance energy transfer (FRET) fluorophore and fluorescence quencher pair, respectively. These groups were selected for FRET studies owing to their ready incorporation as oligomer end groups and their stability to resin cleavage conditions. Here, the oligomer sequence 01010 was labeled with FAM (denoted as 01010-FAM) and the sequences 10001 and 10101 were labeled with DABCYL (10001-DABCYL and 10101-DABCYL, respectively). All assays were normalized to a non-quenched fluorescence intensity ceiling, determined using a solution of 01010-FAM as a single-strand negative control. A control solution incorporating 01010-FAM and 10101-DABCYL, fully dissociated with 1.5 eq. of scandium triflate, was assessed to confirm the absence of fluorescence quenching for un-hybridized strands, while another solution of 01010-FAM×10101 (i.e., omitting a quencher) was examined to confirm that the hybridization of the FAM-bearing strand with its DABCYL-free complementary sequence does not negatively affect fluorescence (see Figure 2.12.b, orange and purple bars, respectively). Upon treatment with the dissociation/extraction/annealing process, the 01010-FAM×10101-DABCYL reaction mixture afforded a normalized fluorescence intensity of 0.275±0.050 (i.e., a 72.5±5.0% reduction in fluorescence, see Figure 2.12.b, green bar), a value approaching the reported FAM/DABCYL fluorescence quenching efficiency of 80%[40]. Conversely, solutions of both the same 01010-FAM and 10101-DABCYL oligomer pair after treatment with the single-pot, single-step assembly method, and the non-complementary 01010-FAM and 10001-DABCYL sequences after the dissociation/extraction/annealing process yielded moderate reductions in fluorescence intensity (Figure 2.12.b, yellow and red bars, respectively); however, the extent of fluorescence quenching for these systems was curtailed by the generation not only of hybridized species bearing both FAM and DABCYL but also of hybrids exclusively bearing either FAM or DABCYL groups (i.e., equivalent to the hybridized species between non-complementary strands apparent in Figure 2.5.b, bottom black, and Fig. 2.12.a, bottom red spectra, respectively). Finally, hybridization selectivity was examined using a ternary system, equivalent to that described above, employing stoichiometric amounts of the sequences 01010-FAM, 10001-DABCYL, and quencher-free 10101. After the dissociation/extraction/annealing process, the normalized fluorescence intensity was determined to be 0.810±0.055 (Fig. 2.12.b, blue bar), demonstrating the preferential hybridization of complementary sequences enabled by this process to selectively afford the 10101×01010-FAM molecular ladder bearing 5 imine-based rungs (Fig. 2.12.b, green bar) over a non-complementary, 4 imine-rung 10001-DABCYL×01010-FAM molecular ladder (Fig. 2.12.b, red bar).

2.5.3 Sequence-selective Hybridization

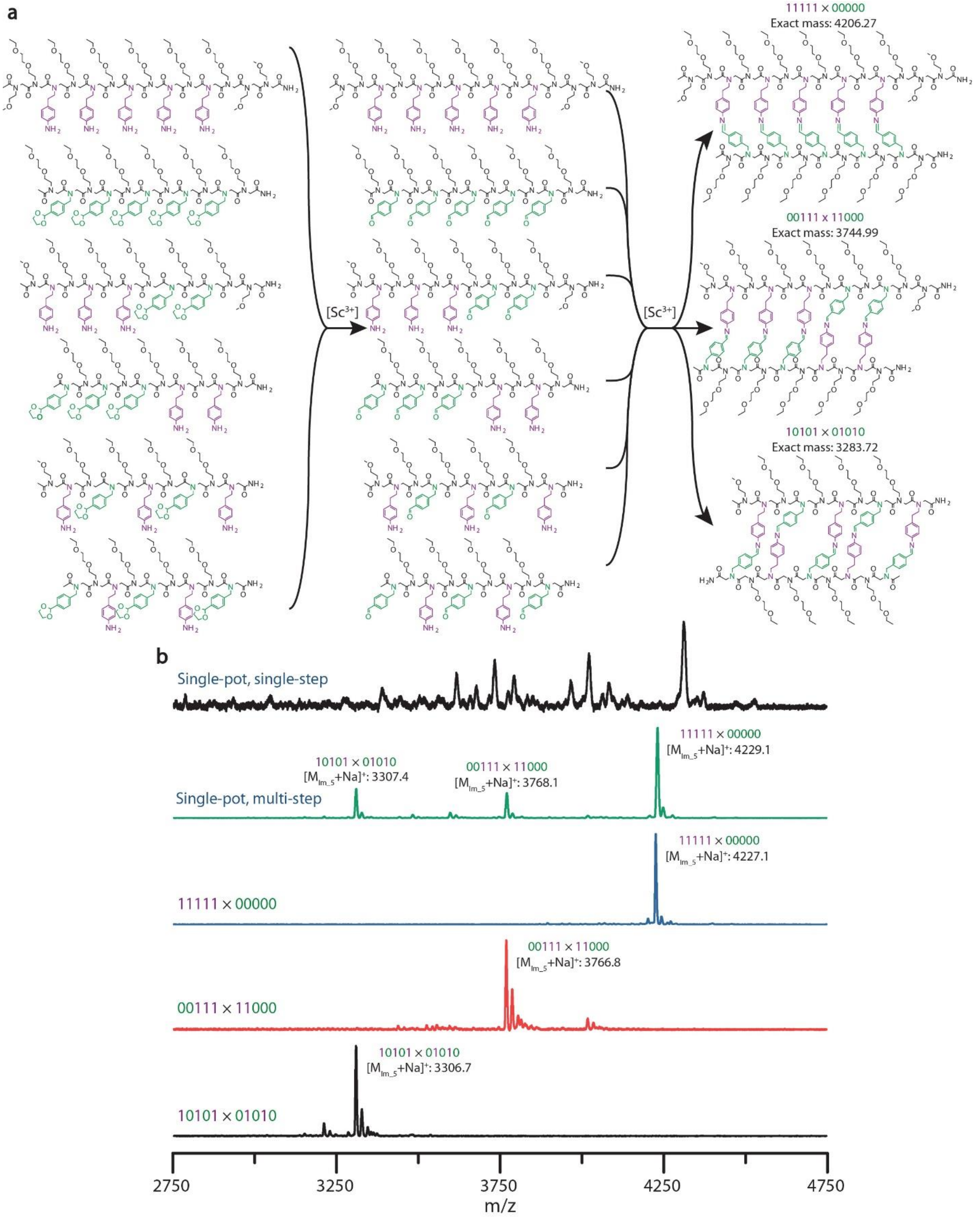


Figure 2.13 Sequence-selective hybridization. a, Structures of mass-labeled, ethylene acetal-protected precursor peptoid sequences, deprotected, dissociated sequences upon treatment with $Sc(OTf)_3$, and in-registry molecular ladders upon [Sc^{3+}] extraction and annealing. b, MALDI mass spectra of individual encoded molecular ladders assembled *via* the dissociation/extraction/annealing process, including 10101×01010 (bottom, black), 00111×11000 (second from bottom, red), 11111×00000 (middle, red), and a single-pot solution of all six oligomers to yield three in-registry molecular ladders (second from top, green). A single-pot solution of the six oligomers after the single-step, deprotection and direct assembly process (top, black) is shown for comparison. Expected exact masses: $[M_{10101\times01010}+Na]^+ = 3306.71$; $[M_{00111\times11000}+Na]^+ = 3767.93$; $[M_{11111\times00000}+Na]^+ = 4229.25$. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs.

The assembly of sophisticated, multimeric macromolecular complexes requires the selective and concurrent hybridization of multiple information-encoded components[7–10]. To examine the potential for this dynamic covalent assembly system to achieve concurrent hybridization selectivity, three unique pairs of mass-labeled, complementary peptoids sequences (shown in Figure 2.13.a) were initially each subjected to the stepwise, dissociation/extraction/annealing assembly process described above. The mass spectra of each reaction mixture demonstrate the successful assembly of the target 10101×01010, 11100×00011, and 11111×00000 molecular ladders from their respective precursor strands (see Figure 2.13.b, bottom three black, red, and blue spectra, respectively); for comparison, mass spectra of the 00111×11000 and 11111×00000 molecular ladder reaction mixtures using the previously-determined, single-step reaction conditions are shown in Figure 2.14. Notably, assembly of the precursor strands in either a parallel or anti-parallel orientation can afford in-registry molecular ladders for both 10101×01010 and 11111×00000 systems; however, the in-registry 00111×11000 ladder product can only be achieved through parallel-oriented strands. Whereas hybridization of 11111 and 00000 homo-oligomers achieves registry under either reaction conditions (Figure 2.14.a), the 00111×11000 system results in poor hybridization selectivity under the direct deprotection and assembly reaction conditions (Figure 2.14.b, bottom black spectrum), while the dissociation/extraction/annealing process selectively affords the in-registry, five-rung molecular ladder resulting from the hybridization of the 00111 strand with its 11000 complement in a parallel orientation (Figure 2.14.b, top red spectrum).

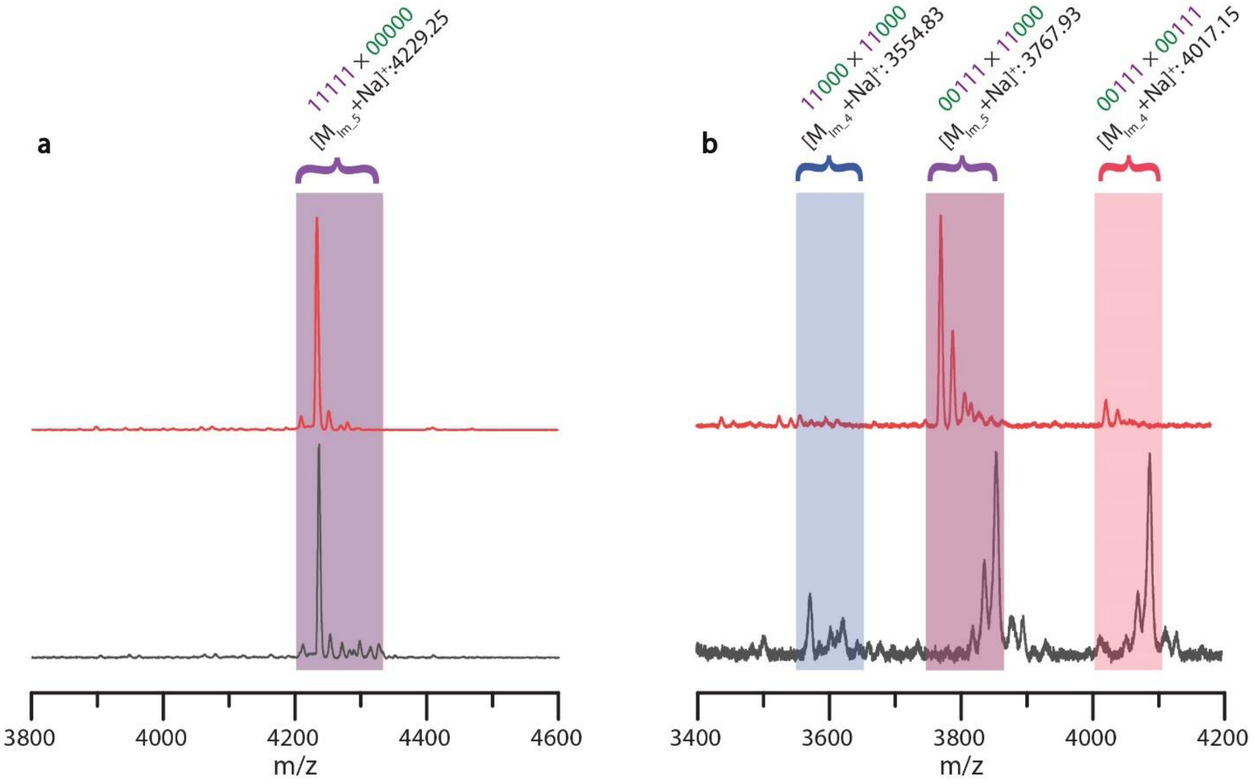


Figure 2.14 MALDI mass spectra of molecular ladder reaction mixtures utilizing either the previously-determined, single-pot, single-step approach (bottom, black), or *via* the dissociation/extraction/ annealing process (top, red) of a, 11111×00000, and b, 00111×11000.

Concurrent hybridization selectivity for these three sets of complementary peptoid pairs was subsequently examined by allowing equimolar amounts of the six oligomeric sequences (i.e., Figure 2.13.a) to react simultaneously in a single pot reaction mixture. Here, excluding out-of-registry and multimeric reaction products, this system of six individual strands could potentially yield 19 unique duplexes (as the 11111 and 00000 sequences cannot self-hybridize); nevertheless, application of the dissociation/extraction/annealing assembly process enabled each strand to preferentially hybridize with its residue sequence complement, resulting in the selective assembly of three in-registry molecular ladder products, including 10101×01010, parallel-oriented 00111×11000, and 11111×00000 (Figure 2.13.b, second from top green spectrum). In contrast, the single-step, deprotection and direct assembly reaction conditions again impeded hybridization selectivity of this system, yielding a non-equilibrium mixture of multiple, unidentified dimeric species (Figure 2.13.b, top black spectrum), highlighting the extraordinary capacity for the dissociation/extraction/annealing process to alleviate the kinetic trapping that commonly accompanies dynamic covalent assembly (Figure 2.15.). The full range of this spectrum (m/z of 1000 to 10000, see Figure 12.16.) again revealed small peaks attributable to single-stranded species, including both unreacted strands and those with either one or two intramolecular imine bonds, while multimeric complexes at higher molecular weights were not observed. Moreover, the concurrent assembly of 10101, 01010, 00111, and 11000 sequences exclusively affords 10101×01010 and parallel-oriented 00111×11000 ladders rather than the non-complementary cross-species 10101×11000 and 01010×00111, both of which could potentially form five imine bonds, confirming that pendant group sequence dictates the hybridization selectivity rather than simply maximizing the number of interstrand imine bonds, and strongly suggesting an unconvoluted, ladder-like conformation.

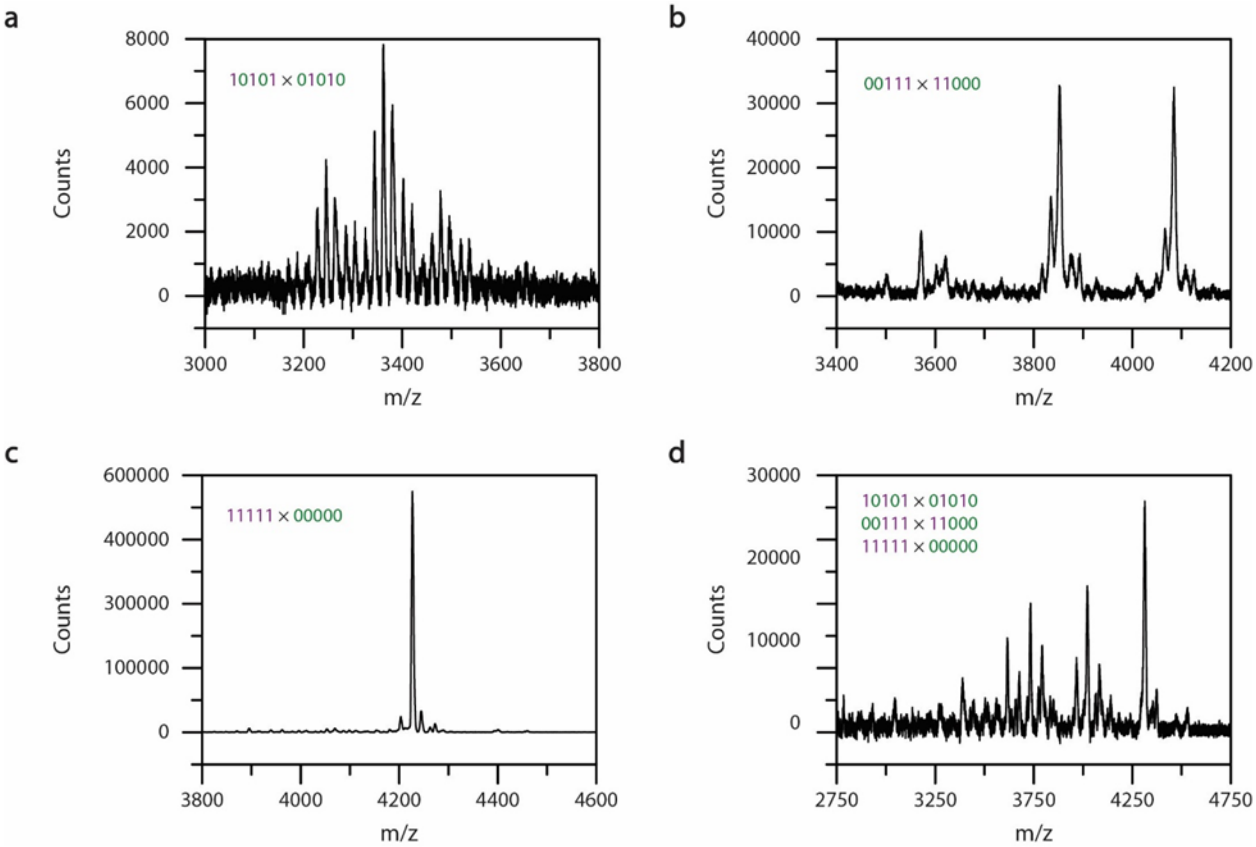


Figure 2.15 MALDI mass spectra, with counts, of molecular ladder reaction mixtures utilizing the previously-determined, single-pot, single-step approach of a, 10101×01010, b, 00111×11000, c, 11111×00000, and d, all six single strands simultaneously.

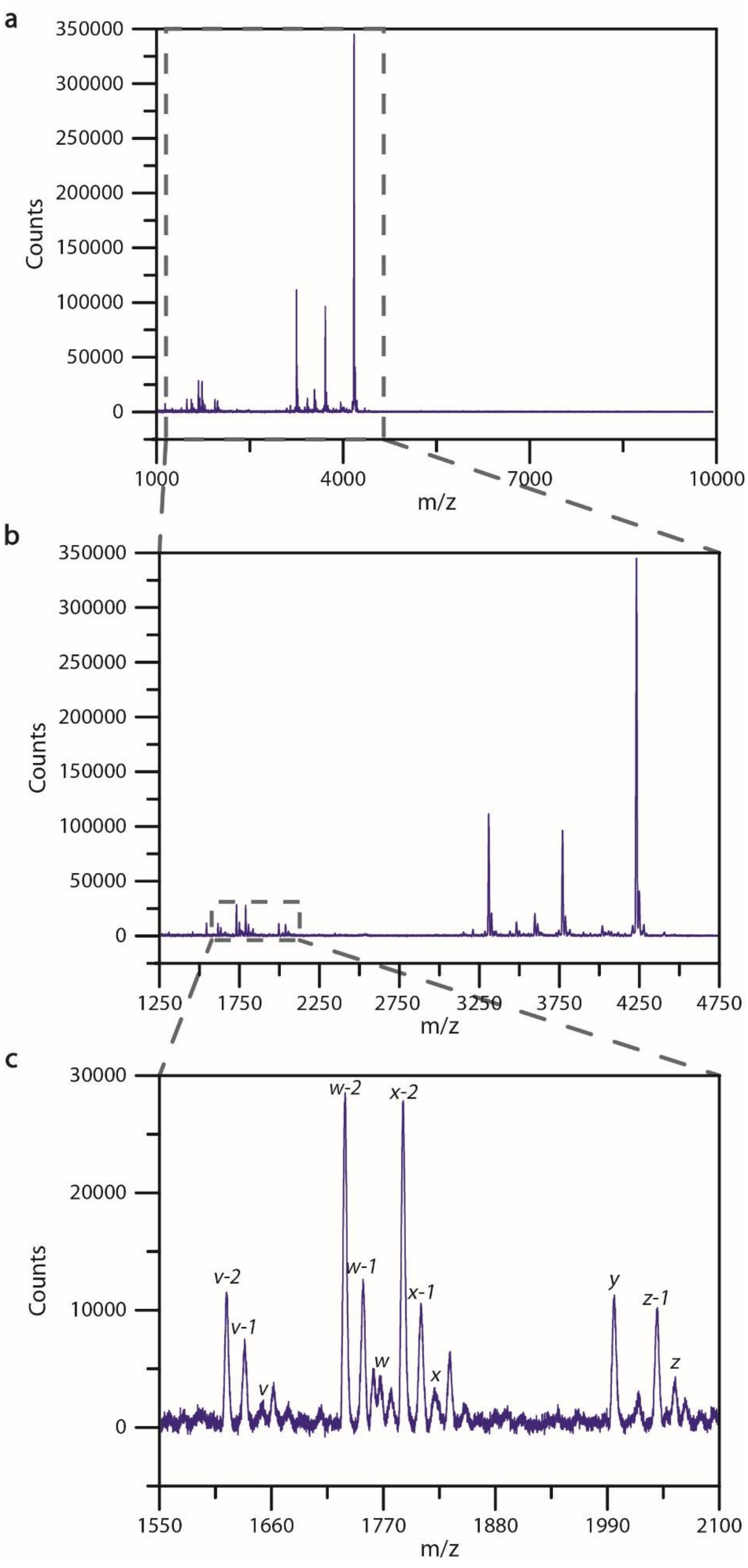

a
350000
300000
250000
200000
150000
100000
50000
0
Counts
1000
4000
7000
10000
m/z
b
350000
300000
250000
200000
150000
100000
50000
0
Counts
1250
1750
2250
2750
3250
3750
4250
4750
m/z
c
30000
20000
10000
0
Counts
v-2
v-1
v
w-2
w-1
w
x-2
x-1
x
y
z-1
z
1550
1660
1770
1880
1990
2100
m/z

Figure 2.16 MALDI mass spectrum of a single-pot solution of six oligomers, including 10101, 01010, 00111, 11000, 11111, and 00000, assembled *via* the dissociation/extraction/annealing process. a, m/z 1000-10000 (full range), b, m/z 1250-4750 (expected range for dimeric and single-stranded species), and c, m/z 1550-2100 (expected range for single-stranded species). Expected exact masses: $[M_{01010}+Na]^+$ (*v*) = 1651.83; $[M_{10101}+Na]^+$ (*w*) = 1767.92; $[M_{11000}+Na]^+$ (*x*) = 1824.93; $[M_{00000}+Na]^+$ (*y*) = 1995.97; $[M_{00111}+Na]^+$ (*z*) = 2056.09. Peaks at multiples of -18 m/z values are attributable to single-stranded species incorporating intramolecular imine bonds (e.g., *v* = 01010, *v*-1 = 01010 with one intramolecular imine bond, *v*-2 = 01010 with two intramolecular imine bonds, etc.).

2.4.4 Information Storage and Retrieval

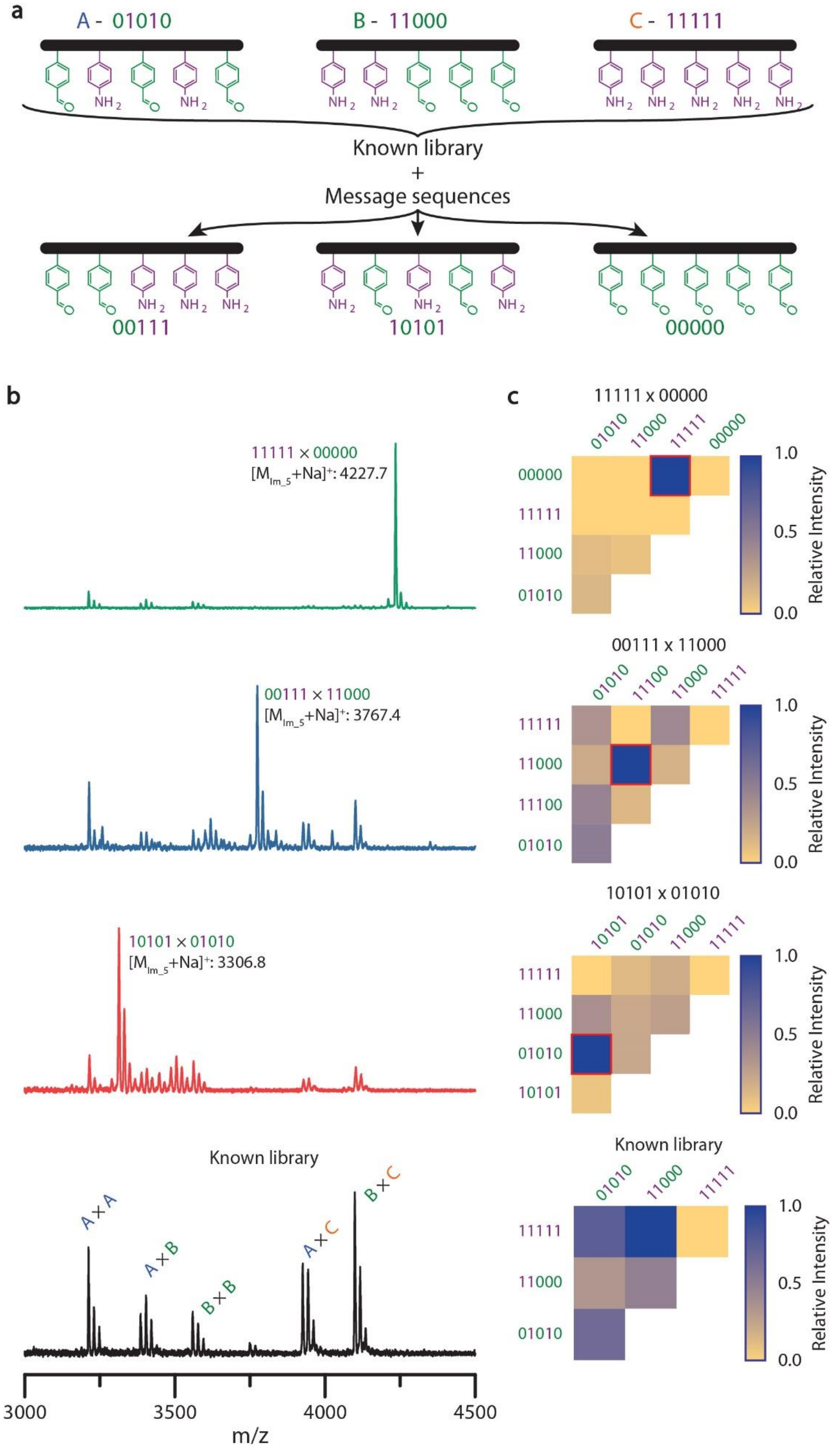


Figure 2.17 Dynamic covalent information storage and retrieval. a, Schematic of a known library composed of three unique, mass-labeled oligomeric sequences A (01010), B (11000), and C (11111), and individual messages sequences used to challenge the library. b, MALDI mass spectrum of the known sequence library treated to the dissociation/extraction/annealing assembly process in the absence of a message strand (black, bottom), resulting in non-specific binding among the library strands, yielding all five possible sets of dimeric ladder species with varying numbers of rungs. Introduction of a message sequence to the known library and treatment of the mixture to the multi-step assembly process yielded MALDI mass spectra with dominant peaks attributable to the in-registry hybridization product of the message strand with its library complement (10101 message, second from bottom, red; 00111 message, second from top, blue; 00000 message, top, green). c, Hybridization specificity of message strands with their library complements assessed by normalized MALDI mass spectrum intensities. Possible dimeric ladder species are shown; red boxes indicate expected hybrids between message and complementary library strands. Peaks at multiples of +18 m/z values are attributable to ladders species with fewer rungs.

Inspired by the facile nucleic acid sequence identification provided by contemporary DNA microarray technology[41,42] and the hybridization selectivity of the encoded sequences demonstrated above, we were motivated to explore the versatility of this model system as a potential information retrieval mechanism. Here, messages in the form of binary-encoded peptoid sequences are challenged against a known library of unpaired, mass-labeled oligomers. Selective hybridization of a message peptoid with its complementary sequence should afford a molecular ladder with a pre-determined mass such that the message can be identified through mass spectrometry on the reaction mixture. To validate this concept, the mass-labeled, acetal-protected peptoid oligomers 01010, 11000, and 11111 were used in equimolar amounts as a known library of potential complementary sequences (Figure 2.17.a). Initially, treatment of this library to the dissociation/extraction/annealing assembly process in the absence of a message strand resulted in non-specific binding among the library strands, yielding five sets of dimeric ladder species with varying numbers of rungs (see Figure 2.17.b, bottom black spectrum). Upon addition of a sub-stoichiometric amount of a 00111 message strand to a fresh sequence library and application of the dissociation/extraction/annealing process, characterization of the reaction mixture by mass spectrometry revealed a dominant new peak attributable to a selectively hybridized, in-registry ladder composed of the message oligomer and its complementary strand from the known library (Figure 2.17.b, second from top blue spectrum), successfully demonstrating the capacity of this system to convey information despite the presence of competitive binding species in excess. Similar results were obtained for 01010 and 00000 message strands when challenged against the oligomer library (see Figure 2.17.b, second from bottom red and top green spectra, respectively). To determine the detection limit for this message identification approach, varying sub-stoichiometric amounts of a 10101 message strand were added to the known sequence library described above (see Figure 2.18.a) and, after treatment with the dissociation/extraction/annealing process, the mixtures were characterized again by mass spectrometry. Whereas peaks attributable to the 10101×01010 ladder product were readily apparent in the mass spectrum at 0.2 equivalents of the message strand, addition of 0.75 equivalents of the message were necessary for a peak attributable the in-registry ladder product generated from the message strand and its complementary sequence to become dominant in the mass spectrum (Figure 2.18.b and Table 2.1.).

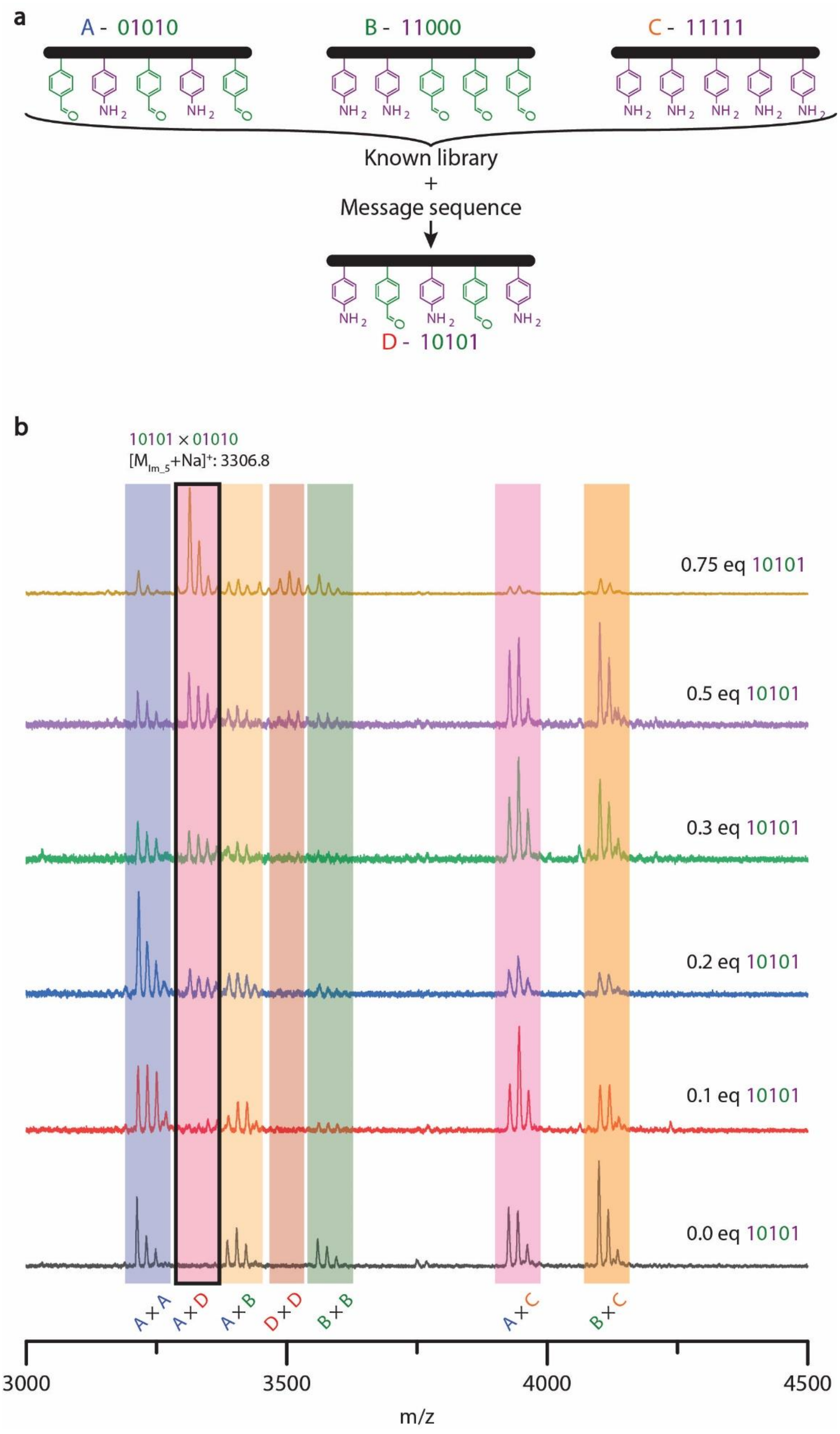

a
A - 01010
B - 11000
C - 11111
NH2
Known library
+
Message sequence
D - 10101
b
10101 × 01010
[M_Im_5+Na]+: 3306.8
0.75 eq 10101
0.5 eq 10101
0.3 eq 10101
0.2 eq 10101
0.1 eq 10101
0.0 eq 10101
A+A
A+D
A+B
D+D
B+B
A+C
B+C
3000
3500
4000
4500
m/z

Figure 2.18 Dynamic covalent information storage and retrieval with varying equivalents of the message strand. a, Schematic of a known library composed of three unique, mass-labeled oligomeric sequences A, B, and C, and individual message sequence used to challenge the library, D. b, MALDI mass spectrum of the known sequence library treated to increasing equivalents of the message strand.

Table 2.1 Hybridization specificity of message strands with their library complements assessed by normalized MALDI mass spectrum intensities.

Known Library		**11111 x 00000**	
Sequence Combination	Normalized Intensity	Sequence Combination	Normalized Intensity
01010 x 01010	0.65	01010 x 01010	0.14
01010 x 11000	0.48	01010 x 11000	0.11
01010 x 11111	0.75	01010 x 00000	-
11000 x 11000	0.33	01010 x 11111	-
11000 x 11111	1.00	11000 x 11000	0.08
		11000 x 00000	-
		11000 x 11111	-
		11111 x 00000	1.00

00111 x 11000		**10101 x 01010**	
Sequence Combination	Normalized Intensity	Sequence Combination	Normalized Intensity
01010 x 01010	0.51	01010 x 01010	0.23
01010 x 11000	0.21	01010 x 10101	1.00
01010 x 00111	0.45	01010 x 11000	0.23
01010 x 11111	0.30	01010 x 11111	0.12
11000 x 11000	0.17	10101 x 11000	0.36
11000 x 00111	1.00	10101 x 10101	0.07
00111 x 00111	-	11000 x 11000	0.27
11000 x 11111	0.42	11000 x 11111	0.20

2.6 Conclusions

The ability to direct the self-assembly of oligomeric strands based on their residue sequence and mediated by dynamic covalent interactions as demonstrated here is a crucial step towards the fabrication of complex, unimolecular constructs from modest, synthetically-accessible precursors. By employing a step-wise process technique involving $Sc(OTf)_3$ as a multi-role reagent to deprotect acetal-protected aldehyde groups, affect the equilibrium of the amine/aldehyde condensation reaction, and catalyze the rearrangement of inter-strand imine bonds, the kinetic trapping of non-equilibrium species that often prevails in dynamic covalent assembly systems is mitigated, allowing the system to converge towards thermodynamic equilibrium through sequence-selective assembly of molecular ladders from oligo(peptoid)s. By employing a dissociation/annealing assembly process, other dynamic covalent interactions could be utilized to mediate selective sequence assembly (e.g., boronic acid/diol condensation by titration, Diels-Alder cycloaddition by thermal annealing, etc.), similarly mimicking that of nucleic acids. Although this study involved molecular ladders bearing covalent rungs, this multi-step approach to dynamic covalent assembly process may also be useful for other applications in which the alleviation or elimination of kinetic trapping is critical. Specifically, this process will provide significantly improved synthetic access to robust, complex covalent nanostructures such as molecular cages and crystalline, porous polymer networks.

2.7 References

1. Lehn, J.-M. Perspectives in Supramolecular Chemistry-From Molecular Recognition towards Molecular Information Processing and Self-organization. *Angew. Chem. Int. Ed.* **29,** 1304–1319 (1990).
2. Lutz, J.-F., Ouchi, M., Liu, D. R. & Sawamoto, M. Sequence-controlled polymers. *Science* **341,** 1238149 (2013).
3. Chen, Z. *et al.* Programmable design of orthogonal protein heterodimers. *Nature* **565,** 106–111 (2018).
4. Coppock, M. B., Miller, J. R. & Williams, M. E. Assembly of a trifunctional artificial peptide into an anti-parallel duplex with three Cu(II) cross-links. *Inorg. Chem.* **50,** 949–955 (2011).
5. Gong, B. Molecular duplexes with encoded sequences and stabilities. *Acc. Chem. Res.* **45,** 2077–2087 (2012).
6. Yang, Y. & Chen, C.-F. in *Hydrogen Bonded Supramolecular Structures* (eds. Li, Z.-T. & Wu, L.-Z.) 115–136 (Springer Berlin Heidelberg, 2015).
7. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99,** 237–247 (1982).
8. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).
9. Tørring, T., Voigt, N. V., Nangreave, J., Yan, H. & Gothelf, K. V. DNA origami: A quantum leap for self-assembly of complex structures. *Chem. Soc. Rev.* **40,** 5636–5646 (2011).
10. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485,** 623–626 (2012).

11. Clausen-Schaumann, H., Rief, M., Tolksdorf, C. & Gaub, H. E. Mechanical stability of single DNA molecules. *Biophys. J.* **78,** 1997–2007 (2000).
12. Nielsen, P. E. Peptide nucleic acid. A molecule with two identities. *Acc. Chem. Res.* **32,** 624–630 (1999).
13. Eriksson, M. & Nielsen, P. E. PNA-nucleic acid complexes. Structure, stability and dynamics. *Q. Rev. Biophsics* **29,** 369–394 (1996).
14. Zhang, L., Peritz, A. & Meggers, E. A simple glycol nucleic acid. *J. Am. Chem. Soc.* **127,** 4174–4175 (2005).
15. Schweitzer, B. A. & Kool, E. T. Aromatic Nonpolar Nucleosides as Hydrophobic Isosteres of Pyrimidine and Purine Nucleosides. *J. Org. Chem.* **59,** 7238–7242 (1994).
16. Piccirilli, J. A., Krauch, T., Moroney, S. E. & Benner, S. A. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* **343,** 33–37 (1990).
17. Abdallah, H. O. *et al.* Stabilisation of self-assembled DNA crystals by triplex-directed photo-cross-linking. *Chem. Commun.* **52,** 8014–8017 (2016).
18. Takasugi, M. *et al.* Sequence-specific photo-induced cross-linking of the two strands of double-helical DNA by a psoralen covalently linked to a triple helix-forming oligonucleotide. *Proc. Natl. Acad. Sci. U.S.A.* **88,** 5602–5606 (1991).
19. Ren, F., Day, K. J. & Hartley, C. S. Two- and three-tiered stacked architectures by covalent assembly. *Angew. Chem. Int. Ed.* **55,** 8620–8623 (2016).
20. Tozawa, T. *et al.* Porous organic cages. *Nat. Mater.* **8,** 973–978 (2009).
21. Huang, N., Wang, P. & Jiang, D. Covalent organic frameworks: A materials platform for structural and functional designs. *Nat. Rev. Mater.* **1,** 1–19 (2016).
22. Reuther, J. F. *et al.* Dynamic covalent chemistry enables formation of antimicrobial peptide quaternary assemblies in a completely abiotic manner. *Nat. Chem.* **10,** 45–50 (2018).
23. Reuther, J. F., Dahlhauser, S. D. & Anslyn, E. V. Tunable Orthogonal Reversible Covalent (TORC) Bonds: Dynamic Chemical Control over Molecular Assembly. *Angew. Chem. Int. Ed.* **58,** 74–85 (2019).
24. Hebel, M. *et al.* Sequence Programming with Dynamic Boronic Acid/Catechol Binary Codes. *J. Am. Chem. Soc.* **141,** 14026–14031 (2019).
25. Strom, K. R., Szostak, J. W. & Prywes, N. Transfer of Sequence Information and Replication of Diimine Duplexes. *J. Org. Chem.* **84,** 3754–3761 (2019).
26. Dunn, M. F., Wei, T., Zuckermann, R. N. & Scott, T. F. Aqueous dynamic covalent assembly of molecular ladders and grids bearing boronate ester rungs. *Polym. Chem.* **10,** 2337–2343 (2019).
27. Elliott, E. L., Hartley, C. S. & Moore, J. S. Covalent ladder formation becomes kinetically trapped beyond four rungs. *Chem. Commun.* **47,** 5028–5030 (2011).
28. Hartley, C. S., Elliott, E. L. & Moore, J. S. Covalent assembly of molecular ladders. *J. Am. Chem. Soc.* **129,** 4512–4513 (2007).
29. Mattia, E. & Otto, S. Supramolecular Systems Chemistry. *Nat. Nanotechnol.* **10,** 111–119 (2015).
30. Zuckermann, R. N. Peptoid origins. *Biopolymers* **96,** 545–555 (2011).
31. Zuckermann, R. N., Kerr, J. M., Moosf, W. H. & Kent, S. B. H. Efficient Method for the Preparation of Peptoids [Oligo(N-substituted glycines)] by Submonomer Solid-Phase Synthesis. *J. Am. Chem. Soc.* **114,** 10646–10647 (1992).

32. Sun, J. & Zuckermann, R. N. Peptoid polymers: A highly designable bioinspired material. *ACS Nano* **7,** 4715–4732 (2013).
33. Edison, J. R. *et al.* Conformations of peptoids in nanosheets result from the interplay of backbone energetics and intermolecular interactions. *Proc. Natl. Acad. Sci. U.S.A.* **115,** 5647–5651 (2018).
34. Wei, T., Furgal, J. C. & Scott, T. F. In situ deprotection and dynamic covalent assembly using a dual role catalyst. *Chem. Commun.* **53,** 3874–3877 (2017).
35. Wei, T., Furgal, J. C., Jung, J. H. & Scott, T. F. Long, self-assembled molecular ladders by cooperative dynamic covalent reactions. *Polym. Chem.* **8,** 520–527 (2017).
36. Giuseppone, N., Schmitt, J. L., Schwartz, E. & Lehn, J. M. Scandium(III) catalysis of transimination reactions. Independent and constitutionally coupled reversible processes. *J. Am. Chem. Soc.* **127,** 5528–5539 (2005).
37. Meng, W. *et al.* An autonomous molecular assembler for programmable chemical synthesis. *Nat. Chem.* **8,** 542–548 (2016).
38. Furgal, J. C., Goodson, T. & Laine, R. M. D_{5h} $[PhSiO_{1.5}]_{10}$ synthesis via F^- catalyzed rearrangement of $[PhSiO_{1.5}]_n$. An experimental/computational analysis of likely reaction pathways. *Dalt. Trans.* **45,** 1025–1039 (2016).
39. Wei, T., Jung, J. H. & Scott, T. F. Dynamic covalent assembly of peptoid-based ladder oligomers by vernier templating. *J. Am. Chem. Soc.* **137,** 16196–16202 (2015).
40. Marras, S. A. E., Kramer, F. R. & Tyagi, S. Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res.* **30,** (2002).
41. Botstein, D., Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 14863–14868 (1998).
42. Pollack, J. R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23,** 41–46 (1999).

Chapter 3 Sequence-directed Dynamic Covalent Assembly of Base-4-encoded Oligomers

3.1 Original Publication Information

Leguizamon, S.C.*, Dunn, M.F.*, Scott, T.F. Sequence-directed Dynamic Covalent Assembly of Base-4-encoded Oligomers. Chem. Commun., 2020, **56**, 7817-7820

**Authors contributed equally*

Modifications have been made to the original document in order to adapt the content to the proper format.

3.2 Abstract

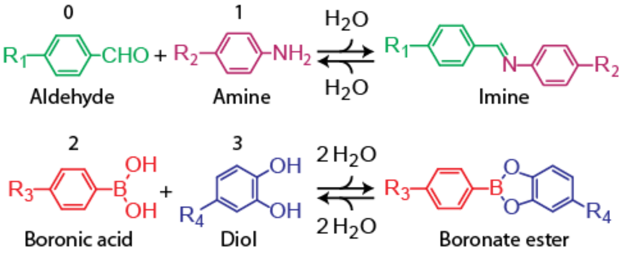
As the primary information-bearing biomacromolecule, DNA is encoded in base-4, where each pendant group site can be occupied by any one of four canonical nucleobases. Moreover, the double and triple hydrogen bonding between the adenine–thymine and guanine–cytosine base pairs enables the selective hybridization of complementary DNA sequences. Here, we examine the use of two dynamic covalent interactions, between boronic acids and diols to afford boronate esters and between amines and aldehydes to afford Schiff base imines, to mediate the sequence-selective assembly of molecular ladders and grids from base-4-encoded oligo(peptoid)s. Upon confirmation of reaction compatibility and orthogonality of the reactions used, a base-4, dynamic covalent self-assembly system, analogous to the hydrogen bond-based hybridization of nucleic acids, was implemented. Molecular ladders and 3×3 grids were assembled from oligo(peptoid) strands bearing sequences of the dynamic covalent pendant groups. The utilization of concurrent, orthogonal dynamic covalent interactions makes possible the formation of structures with the high

level of complexity to be used in molecular electronics, nanomechanical devices and the field of nanomedicine.

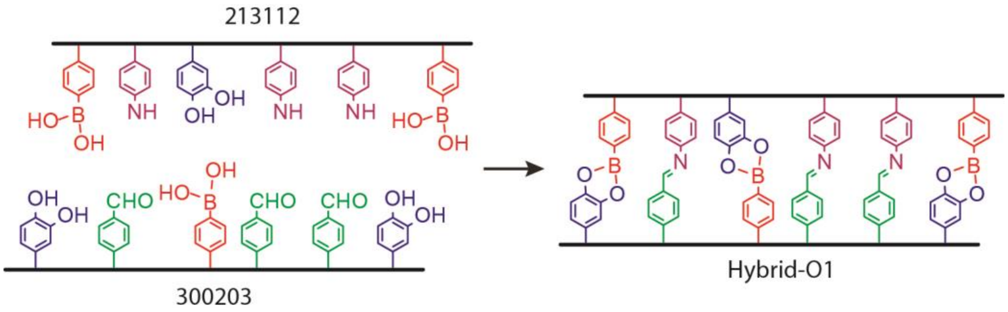
3.3 Introduction

Information in DNA is encoded in the nucleobase sequence affixed as pendant groups on a deoxyribophosphate polymer backbone. The double and triple hydrogen bonding between paired nucleobase residues of adenine–thymine and guanine–cytosine, respectively, enable complementary DNA sequences to selectively self-assemble, yielding molecular ladders encoded in base-4.[1] Owing to this sequence-selective self-assembly, DNA has emerged as a versatile nanoconstruction medium where, upon thermal melting and annealing, mixtures of designed DNA strands can hybridize to afford complex, multi-dimensional structures.[2–4] Nevertheless, despite the success of DNA-based nanotechnology, reliance on interstrand hydrogen bonds and a sugar-phosphate backbone can compromise the mechanical, thermal, and chemical stability of the resultant structures.[5,6]

Mimicking the sequence-selective hybridization of DNA in robust, abiotic systems, the self-assembly of molecular ladders incorporating covalent bond rungs from complementary oligomeric precursor strands bearing covalently coreactive pendant groups has been previously described.[7–9] Employing dynamic covalent interactions to mediate these self-assembly processes provides a mechanism for interstrand connectivity rearrangement, necessary for error correction and enable the annealing of intermediate, out-of-registry species to yield the thermodynamically favored product while maintaining the strength of covalent bonds. The systems examined to date have exclusively employed a single reactant pair such that residue sequences can only be encoded in binary; in contrast, mimicking the base-4 encoding of DNA requires the use of dual concurrent and orthogonal dynamic covalent reactant pairs. Several orthogonal dynamic covalent interactions have been described.[10–12] In particular, imine- and boronate ester-yielding condensation reactions have been employed concurrently to synthesize macrocycles, molecular cages, and dendritic nanostructures.[13–15] Here, we demonstrate an abiotic, DNA-mimetic dynamic covalent system encoded in base-4 by demonstrating the sequence-directed self-assembly of molecular ladders and grids from oligomers bearing two covalently coreactive pendant group pairs whose condensation reactions are orthogonal but proceed under the same reaction conditions (Scheme 3.1. and Scheme 3.2.).



Scheme 3.1 pH-mediated dynamic covalent condensation reactions. The addition of amines (0) to aldehydes (1) yields imines, while the reaction of boronic acids (2) with diols (3) affords boronate esters.



Scheme 3.2 The assembly of a base-4 molecular ladder.

3.4 Experimental

3.4.1. General Experimental Procedure

^{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.

3.4.2. 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 (Figure 3.1.). The dynamic covalent functional group consisted of the prepared dopamine (acetonide) (Nace), 4-(2-aminoethyl)-*N*-(*tert*-butoxycarbonyl)phenylamine (Nbam), 4-(2-aminoethyl)-N-(allylcarbonyloxy)phenylamine (Npam), 4-(1,3-dioxacyclopent-2-yl)benzylamine (Npal), 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 in Chapter 2. 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. Npam and Npal were synthesized as outlined in Chapter 2.

Nace was synthesized according to the following approach (Scheme 3.3.).

The acetonide-protected dopamine monomer for peptoid synthesis was synthesized via a three-step process. The synthesis of the Tfa-dopamine proceeded by treatment of dopamine hydrochloride (21.3 g) with methyl trifluoroacetate (23 mL) in 250 mL of methanol in the presence of triethylamine (64 mL) overnight at room temperature. The solvent was removed by rotary evaporation, and the pH of the resulting solid was adjusted to ~1 with 1 N HCl solution. A liquid extraction was then performed with ethyl acetate followed by water washes. The product was dried with sodium sulfate and rotary evaporation was used to remove the solvent. Subsequently, Tfa-dopamine (5 g) was refluxed with 2,2-dimethoxypropane (10 mL) and a catalytic amount of p-toluenesulfonic acid (172 mg) in toluene. The reaction was outfitted with a Soxhlet extractor filled with granular anhydrous CaCl2. The reaction was refluxed for 2 hours and then allowed to cool to room temperature before being filtered with a short silica gel column washed with dichloromethane (DCM). The solvent was then removed from the product by rotary evaporation before recrystallization in hexanes to afford Tfa-dopamine (acetonide). The acetonide-protected dopamine was obtained by hydrolysis of Tfa-dopamine (acetonide) in THF and an aqueous lithium hydroxide solution (2 equivalents), followed by a liquid-liquid extraction with ethyl acetate. The ethyl acetate was removed by rotary evaporation before the monomer was dried under high vacuum. The 1H-NMR of dopamine(acetonide) in δ (ppm) relative to residual solvent (CDCl3 = 7.24) is shown in Figure 3.2.

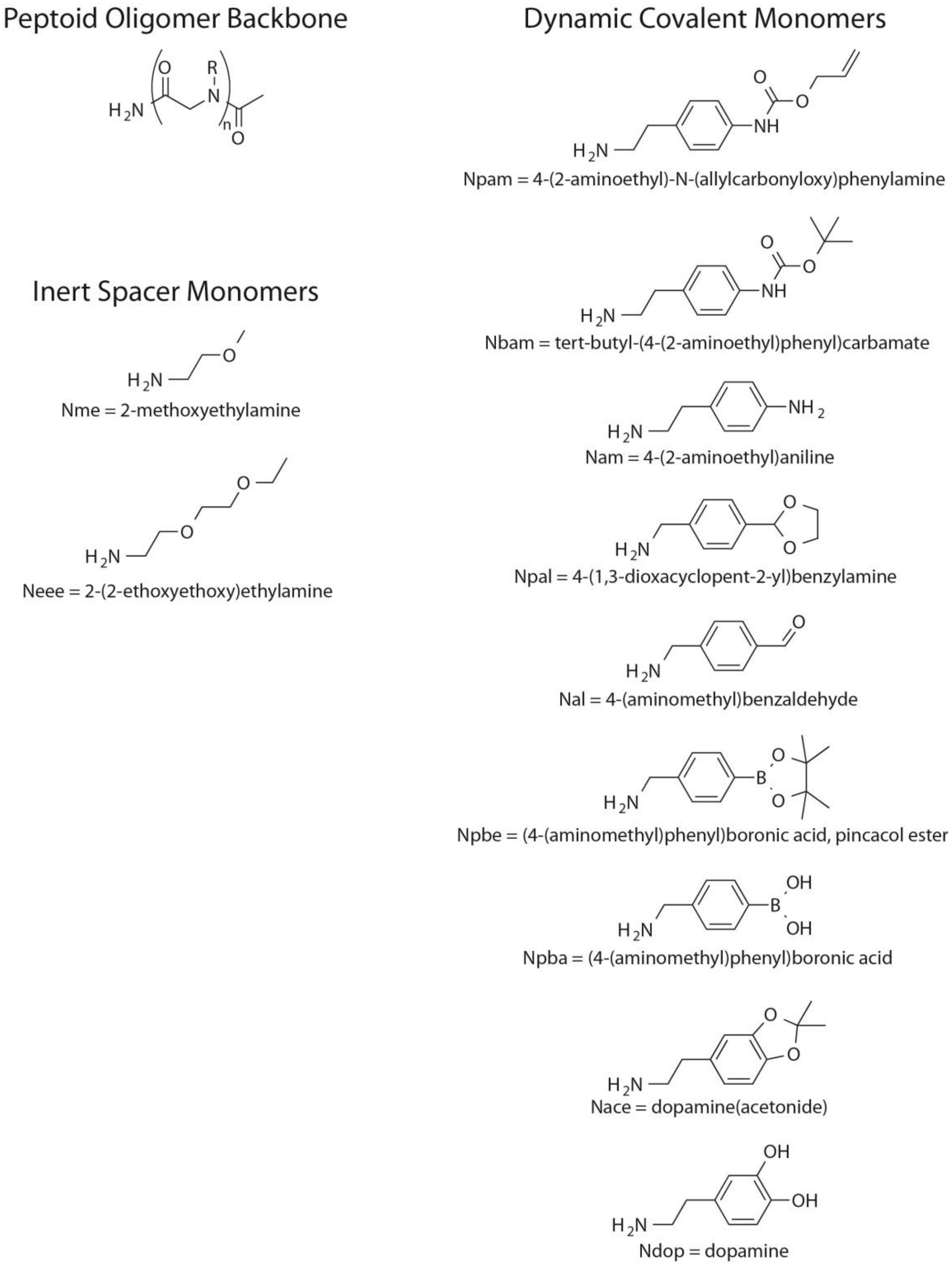
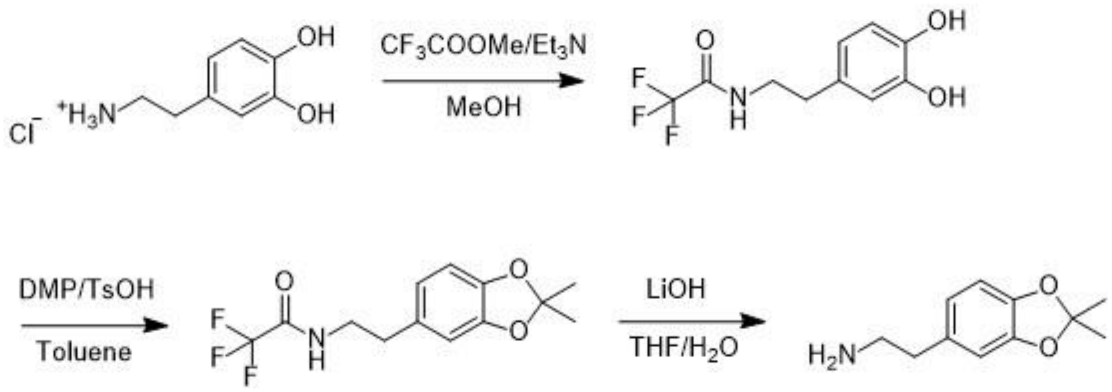

Peptoid Oligomer Backbone
Dynamic Covalent Monomers
Inert Spacer Monomers
Npam = 4-(2-aminoethyl)-N-(allylcarbonyloxy)phenylamine
Nbam = tert-butyl-(4-(2-aminoethyl)phenyl)carbamate
Nme = 2-methoxyethylamine
Nam = 4-(2-aminoethyl)aniline
Npal = 4-(1,3-dioxacyclopent-2-yl)benzylamine
Neee = 2-(2-ethoxyethoxy)ethylamine
Nal = 4-(aminomethyl)benzaldehyde
Npbe = (4-(aminomethyl)phenyl)boronic acid, pincacol ester
Npba = (4-(aminomethyl)phenyl)boronic acid
Nace = dopamine(acetonide)
Ndop = dopamine

Figure 3.1 Primary amine monomers used throughout this chapter.


CF3COOMe/Et3N
MeOH
DMP/TsOH
Toluene
LiOH
THF/H2O

Scheme 3.3 Reaction pathway for the formation of Nace.

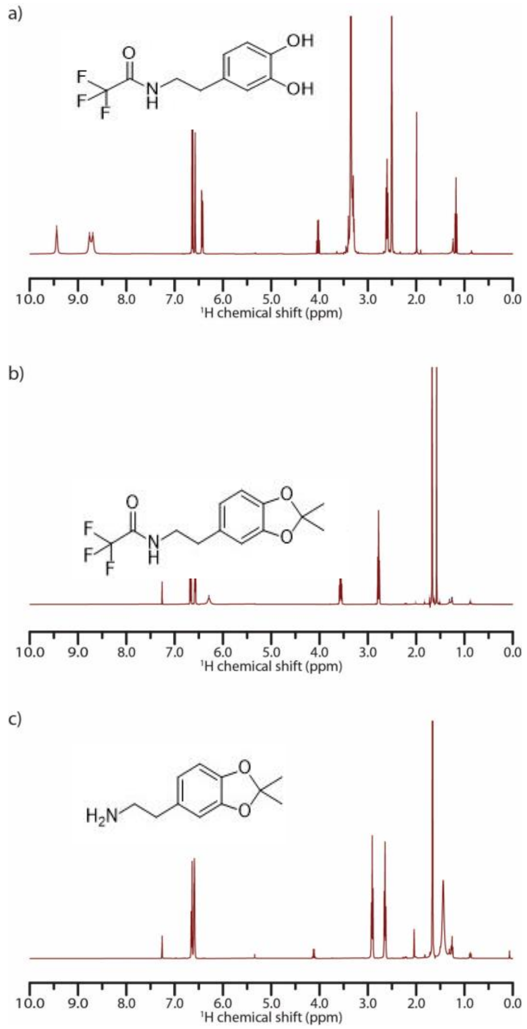


Figure 3.2 ^{1}H NMR spectra for the Nace and each intermediate product.

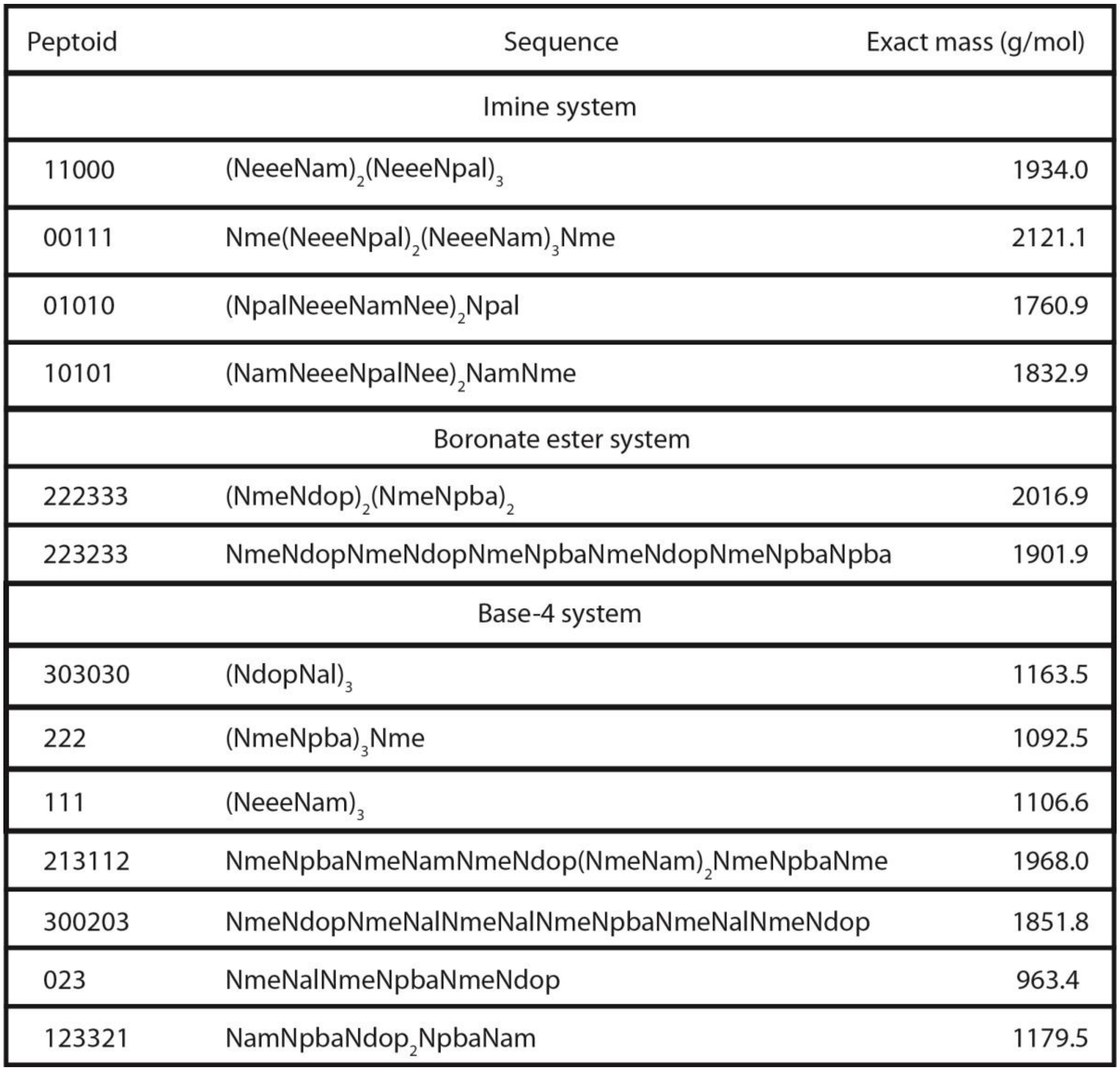
3.4.3. Peptoid Deprotection and Purification

Peptoids bearing Alloc-protected amines (imine only system) were deprotected via an adaptation of the approach in Chapter 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). 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.

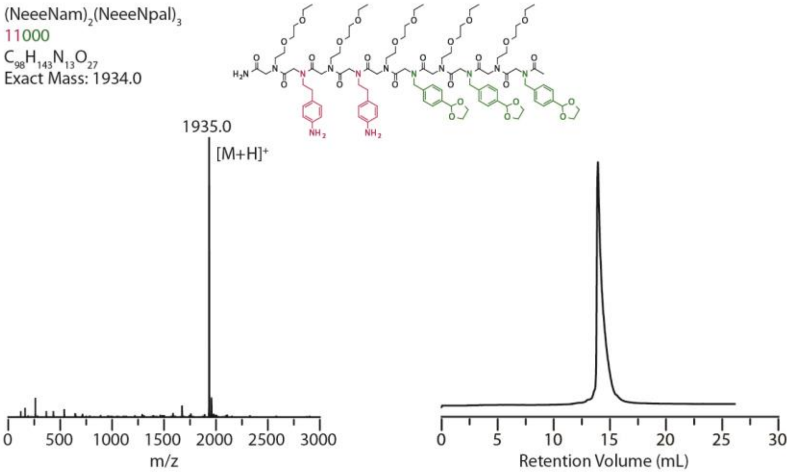
The oligopeptoids were purified by preparative scale RP-HPLC using a linear gradient of acetonitrile and water. Major peaks were collected and fractions were combined before utilizing electrospray ionization (ESI) to confirm the identity of each strand. Fractions of the peptoids were combined and lyophilized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

Table 3.1 Nomenclature of peptoids used in this chapter, 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)



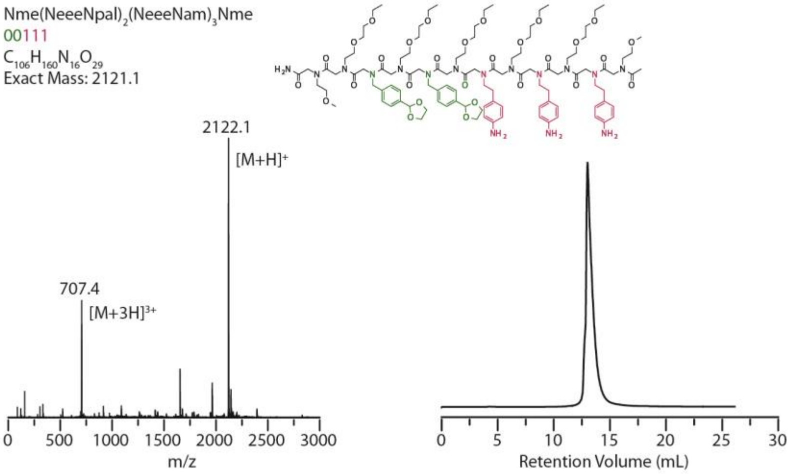
(NeeeNam)$_2$(NeeeNpal)$_3$

11000

$C_{98}H_{143}N_{13}O_{27}$

Exact Mass: 1934.0

(b)



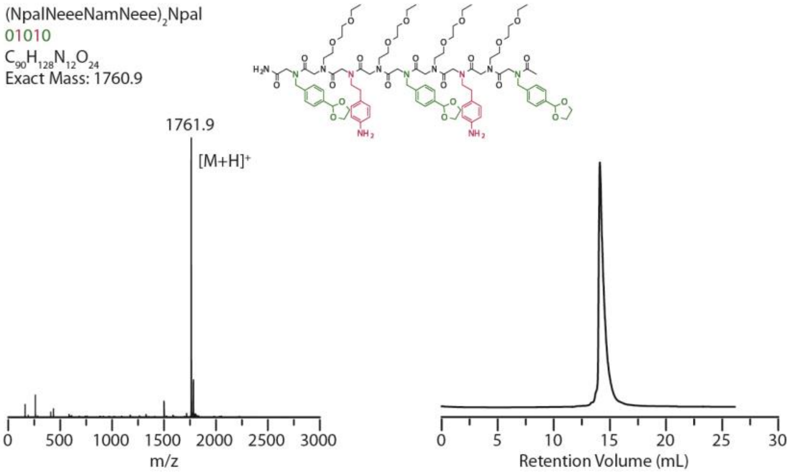
Nme(NeeeNpal)$_2$(NeeeNam)$_3$Nme

00111

$C_{106}H_{160}N_{16}O_{29}$

Exact Mass: 2121.1

(c)



(NpalNeeeNamNeee)$_2$Npal

01010

$C_{90}H_{128}N_{12}O_{24}$

Exact Mass: 1760.9

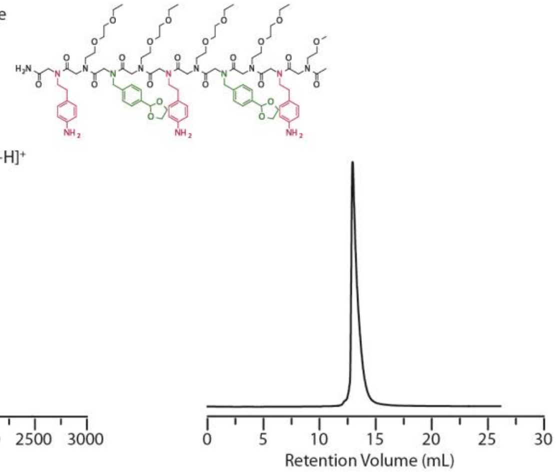
(d)

(NamNeeeNpalNeee)$_2$NamNme

10101

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

Exact Mass: 1832.9



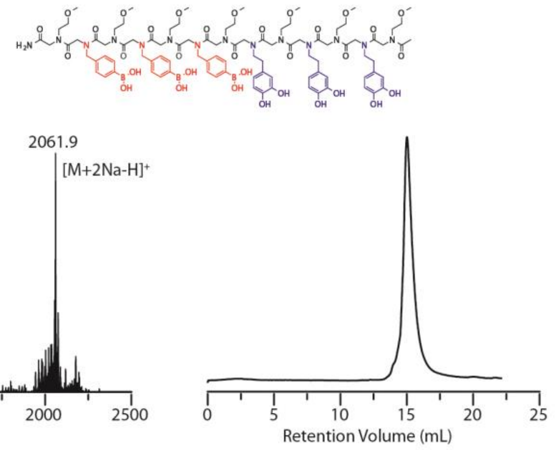
(e)

(NmeNpba)$_3$(NmeNdop)$_3$Nme

222333

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

Exact Mass: 2016.9



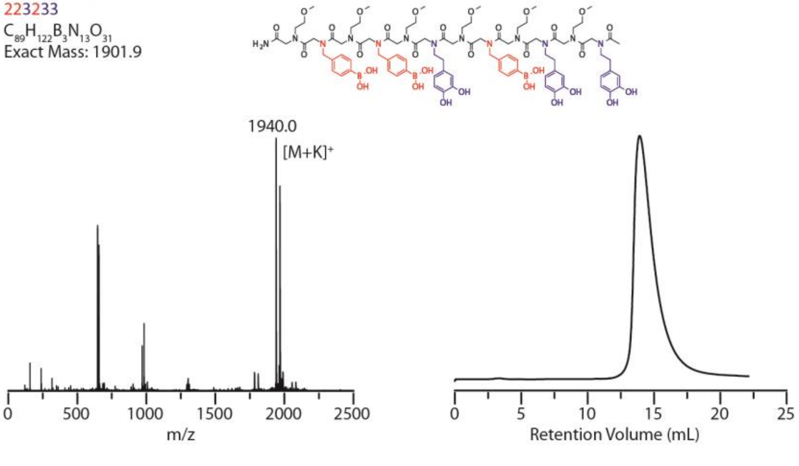
(f)

(NmeNpba)$_2$NmeNdopNmeNpba(NmeNdop)$_2$

223233

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

Exact Mass: 1901.9



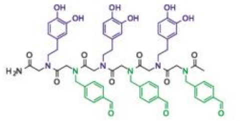
(g)

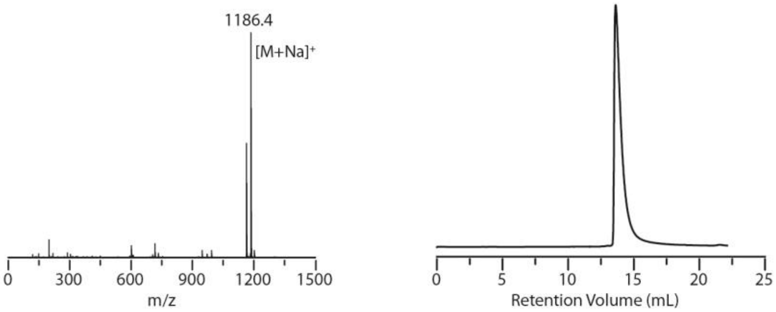
(NdopNal)$_3$

303030

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

Exact Mass: 1163.5





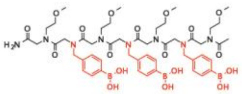
(h)

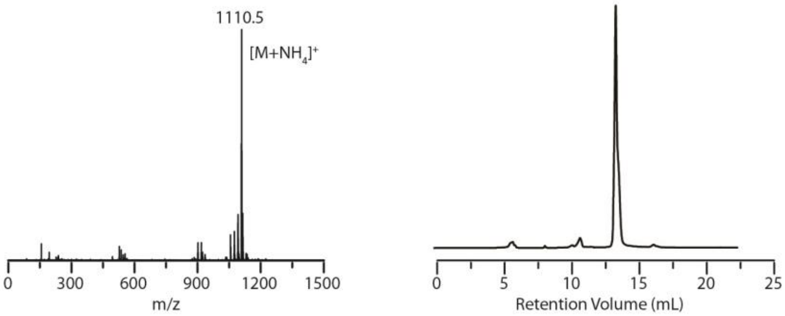
(NmeNpba)$_3$Nme

222

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

Exact Mass: 1092.5





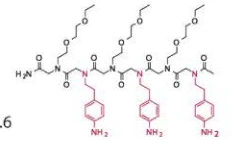
(i)

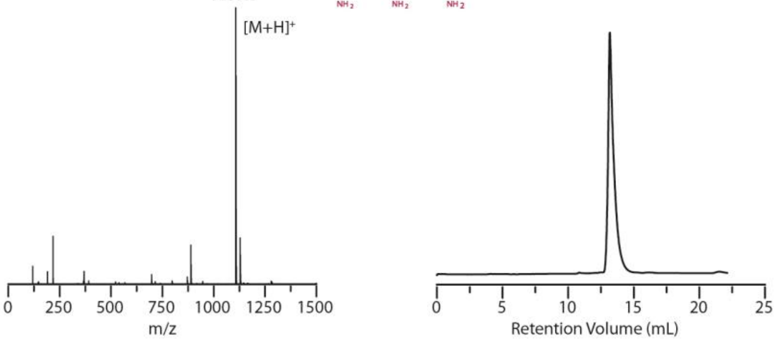
(NeeeNam)$_3$

111

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

Exact Mass: 1106.6





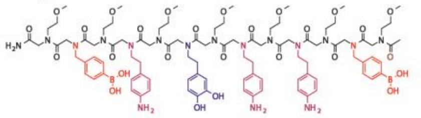
(j)

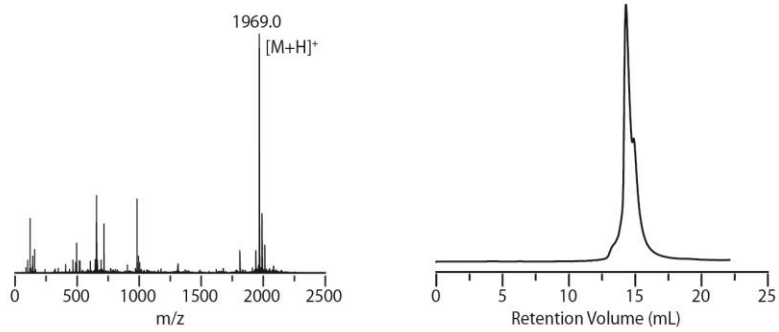
NmeNpbaNmeNamNmeNdop(NmeNam)$_2$NmeNpbaNme

213112

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

Exact Mass: 1968.0





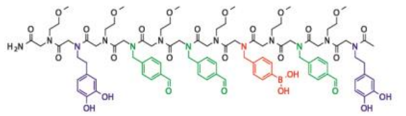
(k)

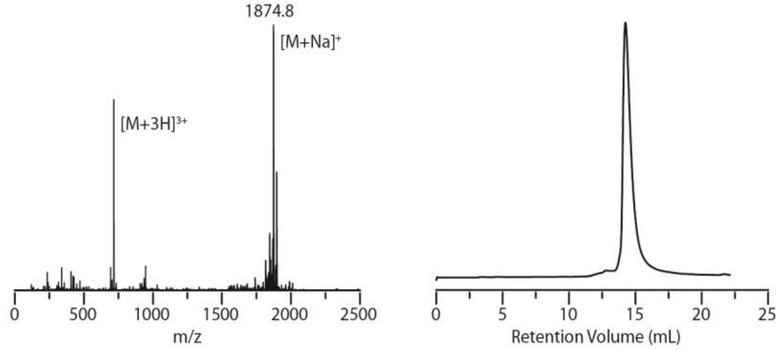
NmeNdop(NmeNal)$_2$NmeNpbaNmeNalNmeNdop

300203

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

Exact Mass: 1851.8





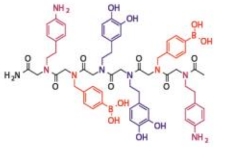
(l)

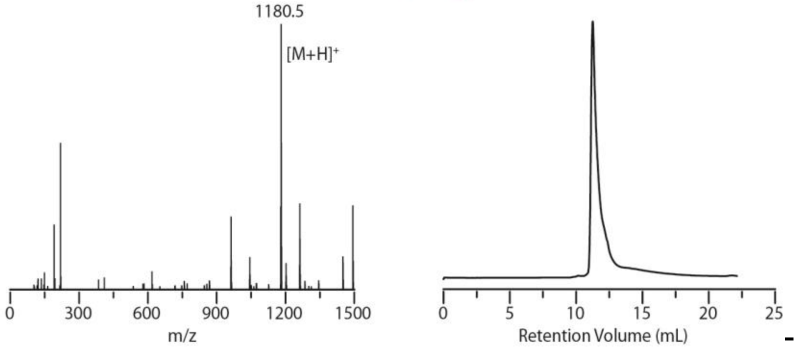
NamNpbaNdop$_2$NpbaNam

123321

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

Exact Mass: 1179.5





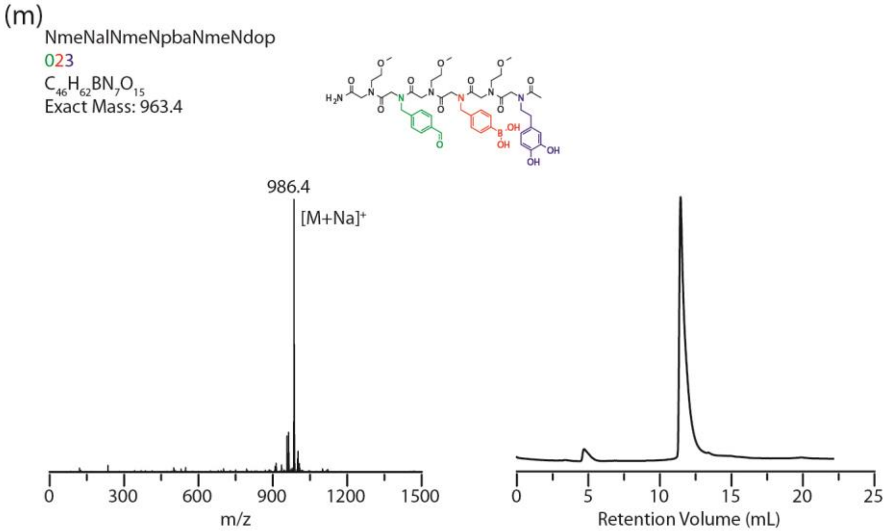


Figure 3.3 ESI mass spectra and corresponding analytical HPLC traces of peptoids synthesized for this study.

3.4.4. 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.

3.4.5. 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.

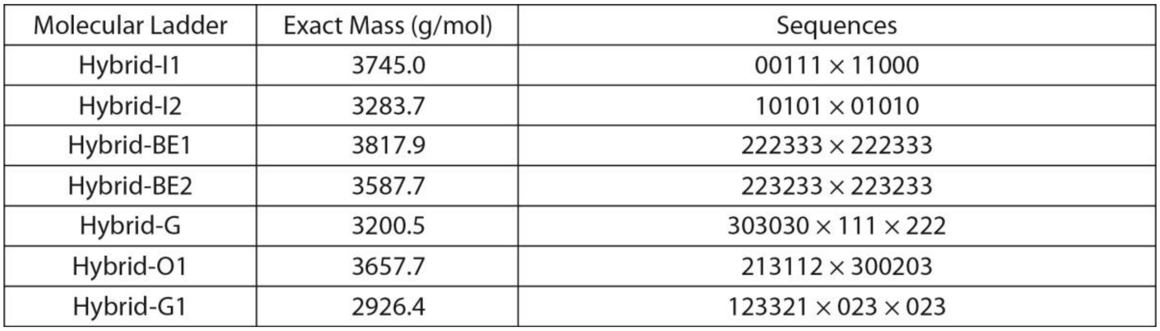
3.4.6. Orthogonality Examination with Model Compounds

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 ^{1}H NMR and ^{11}B NMR spectra were collected. The ^{11}B NMR spectra were collected in a quartz NMR tube.

3.4.7. 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.

Table 3.2 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

3.5. Results and Discussion

3.5.1. Base-2 Sequence Selective Assembly

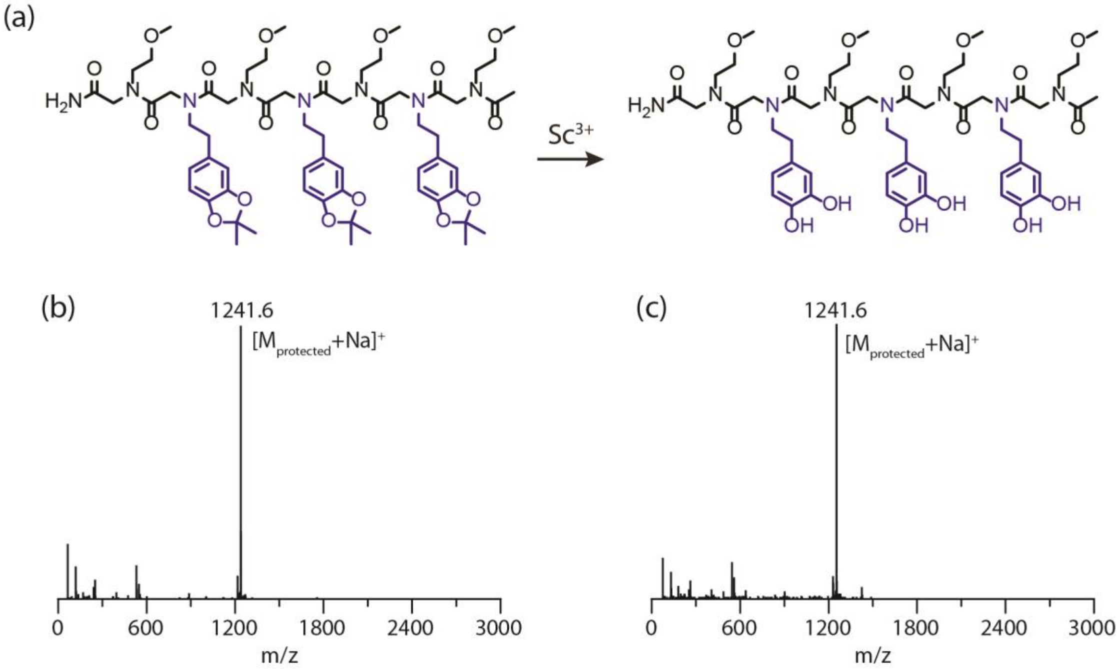


Figure 3.4 Attempted deprotection of an oligomer bearing acetonide-protected diol groups. (a) Hypothesized reaction scheme depicting the deprotection of the acetonide-protected catechol pendant groups on a peptoid by Sc^{3+}. Up to 0.33 equivalents of scandium triflate proved unsuccessful, as shown by 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.

To date, studies examining the dynamic covalent assembly of molecular ladders from single-stranded oligomeric precursors have exclusively employed one pair of pendant groups whose co-reaction proceeds under dissimilar reaction conditions.[8,9,16] To ensure that coreaction of dual reactant pairs affixed to base-4-encoded oligomers would proceed concurrently, we first determined reaction conditions under which both boronic acid/catechol and amine/aldehyde condensation and rearrangement reactions would mediate the assembly of molecular ladders from binary-encoded strands. Sequences bearing multiple co-reactive functional group types necessitate *in situ* functional group deprotection to preclude premature reaction of the precursors, anticipated to yield cross-linked, intractable material. We recently described an approach for the information-directed assembly of peptoid (i.e., oligo(*N*-substituted glycine)) strands bearing amine and acetal-protected aldehyde pendant groups whereby high concentrations of a Lewis acidic rare earth metal triflate facilitated the simultaneous *in situ* aldehyde deprotection and, by shifting the amine/aldehyde reaction equilibrium, dissociation of strands; subsequent extraction of the Lewis acid enabled the imine-yielding condensation reaction and consequent hybridization of information-bearing ladders to proceed.[17] Other work has described acetonide-protected diol deprotection upon Lewis acid treatment,[18,19] suggesting that the reaction conditions previously employed to deprotect acetal-protected aldehydes could similarly effect *in situ* deprotection of acetonide-protected catechol groups. Nevertheless, attempts to deprotect oligomers bearing acetonide-protected diols using scandium triflate (a strong Lewis acid)[20] proved unsuccessful, where no deprotection was observed for a peptoid bearing three acetonide-protected catechol pendant groups even at high Lewis acid concentrations (up to 0.33 equivalents of scandium triflate per functional group, an amount exceeding that required for quantitative acetal deprotection,[18] see Figure 3.4.).

This unsuccessful attempt at Lewis acid-mediated *in situ* deprotection prompted the examination of an alternative approach to affect the reaction equilibria of the boronic acid/catechol and amine/aldehyde condensation reactions by acidifying the reaction mixture with trifluoroacetic acid (TFA) to dissociate the oligomeric sequences, then raising the mixture pH by adding a base to effect ladder assembly. Here, peptoid oligomers were synthesized bearing acetal-protected aldehyde and amine pendant groups (denoted as ‘0’ and ‘1’, respectively). Synthesized peptoids bearing 11000 and 00111 sequences were dissolved in acetonitrile and mixed with TFA to simultaneously deprotect the acetal-protected aldehyde and shift the amine/aldehyde condensation reaction equilibrium to favor dissociated strands. Chloroform was then added, the mixture was adjusted to a pH of 14 by addition of aqueous sodium hydroxide and, upon thorough mixing, washing, and phase separation, the organic layer revealed the formation of in-registry 00111×11000 molecular ladder species (Hybrid-I1, Figure 3.5.a). This approach was similarly employed to afford 10101×01010 molecular ladders from 10101 and 01010 precursor strands (Figure 3.5.a).

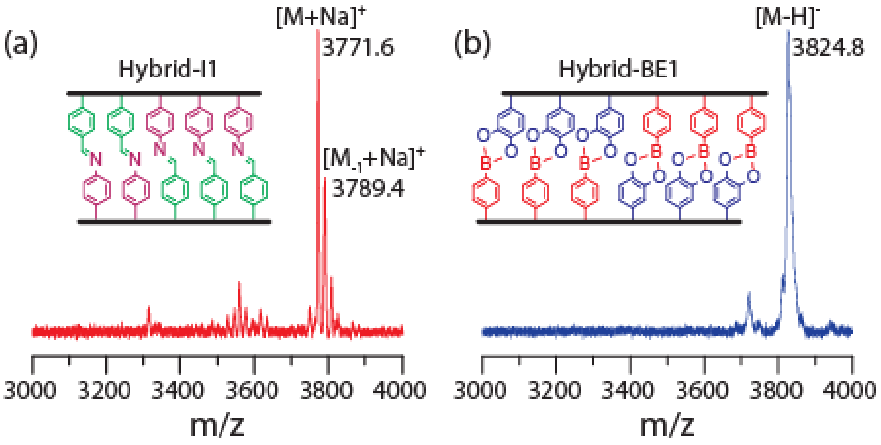


Figure 3.5 Dynamic covalent assembly of information-bearing molecular ladders. Idealized schematics and MALDI-TOF mass spectra of molecular ladders bearing (a) imine and (b) boronate ester rungs. Expected exact masses: $[M_{Hybrid\text{-}I1}+Na]^+ = 3768.0$; $[M_{Hybrid\text{-}BE1}-H]^- = 3816.9$. For imine-bearing ladders, peaks at multiples of +18 m/z values (e.g., M_{-1}) are attributable to ladders species with progressively fewer rungs.

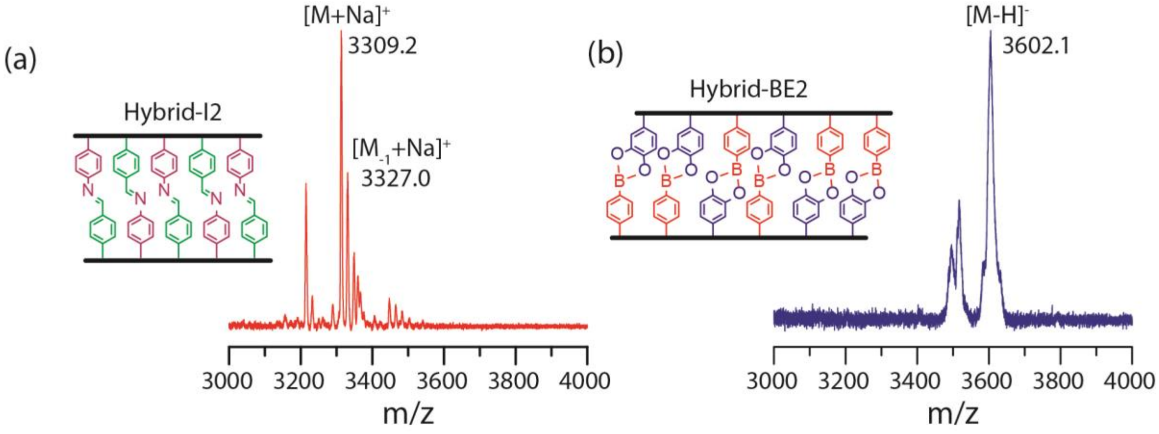


Figure 3.6 Sequence-selective hybridization of molecular ladders bearing (a) imine and (b) boronate ester rungs. Expected exact masses: $[M_{Hybrid\text{-}I2}+Na]^+ = 3306.7$; $[M_{Hybrid\text{-}BE2}-H]^- =$

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.

Self-complementary peptoid oligomers bearing boronic acid and catechol pendant groups (denoted as ‘2’ and ‘3’, respectively) were synthesized and also subjected to sequential acidic and alkaline reaction conditions to dissociate and hybridize these sequences. As with the amine/aldehyde system, the boronic acid/catechol condensation reaction equilibrium is sensitive to pH, where the binding constant between phenylboronic acid and catechol increases from 150 M^{-1} at a pH of 6.5 to 3300 M^{-1} at pH 8.5.[21] For these oligomers, TFA was employed to effect oligopeptoid cleavage and deprotection, ensuring acidic conditions to prevent premature hybridization. Solutions of peptoid sequences were added to an aqueous alkaline solution with a pH of 9 to yield the desired, in-registry molecular ladders 222333×222333 (Hybrid-BE1, Figure 3.5.b) and 223233×223233 (Figure 3.6.b). Out-of-registry species were not identified in the mass spectra of the reaction mixtures examined.

3.5.2. Confirmation of Dynamic Covalent Interactions Orthogonality

Having successfully demonstrated hybridization of binary oligomeric sequences to afford molecular ladders bearing imine- and boronate ester-based rungs by adjusting the reaction mixture pH, the orthogonality of boronic acid/catechol and amine/aldehyde condensation reactions was examined. Here, pairs of model compounds bearing the functional groups employed as dynamic covalent reactant pairs, including catechol, phenylboronic acid, aniline, and benzaldehyde, were mixed in deuterated acetonitrile and the resultant solutions were characterized using ^{11}B NMR (Figure 3.7.) and ^{1}H NMR (Figure 3.8.). Of the reaction mixtures examined, ^{11}B NMR spectrum indicated that phenylboronic acid exclusively reacted with catechol owing to the emergence of a new peak only for the phenylboronic acid/catechol reaction mixture. Similarly, ^{1}H NMR revealed co-reaction between benzaldehyde and aniline, evidenced by the emergence of a peak at 8.58 ppm attributable to imine generation, while neither compound reacted with phenylboronic acid or catechol (Figure 3.8.), confirming the orthogonality of these condensation reactions.

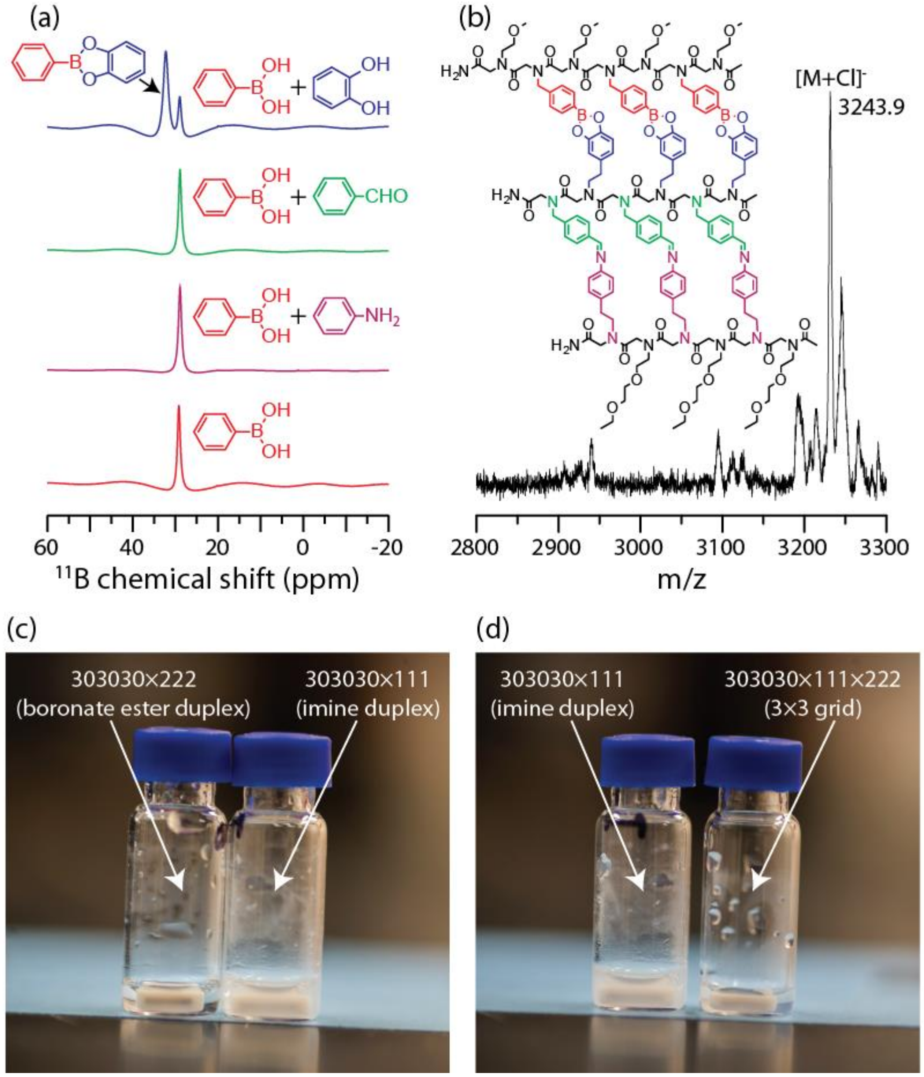


Figure 3.7 Orthogonality between imine and boronate ester systems. (a) ^{11}B-NMR spectra highlighting the specificity of boronic acids to diols. (b) Structure and MALDI-TOF spectra of a 3×3 molecular grid assembled from the concurrent addition of a core peptoid strand with aldehyde and catechol functionality and complementary, exterior sequences bearing amine- and boronic acid-based pendant groups. Expected exact mass: $[M+Cl]^-$ = 3235.5. Digital image capturing the alternative pathways for a 3×3 molecular grid *via* (c) the sequential addition of exterior strands and (d) the final 303030×111×222 molecular grid reaction solution by way of a 303030×111 intermediate.

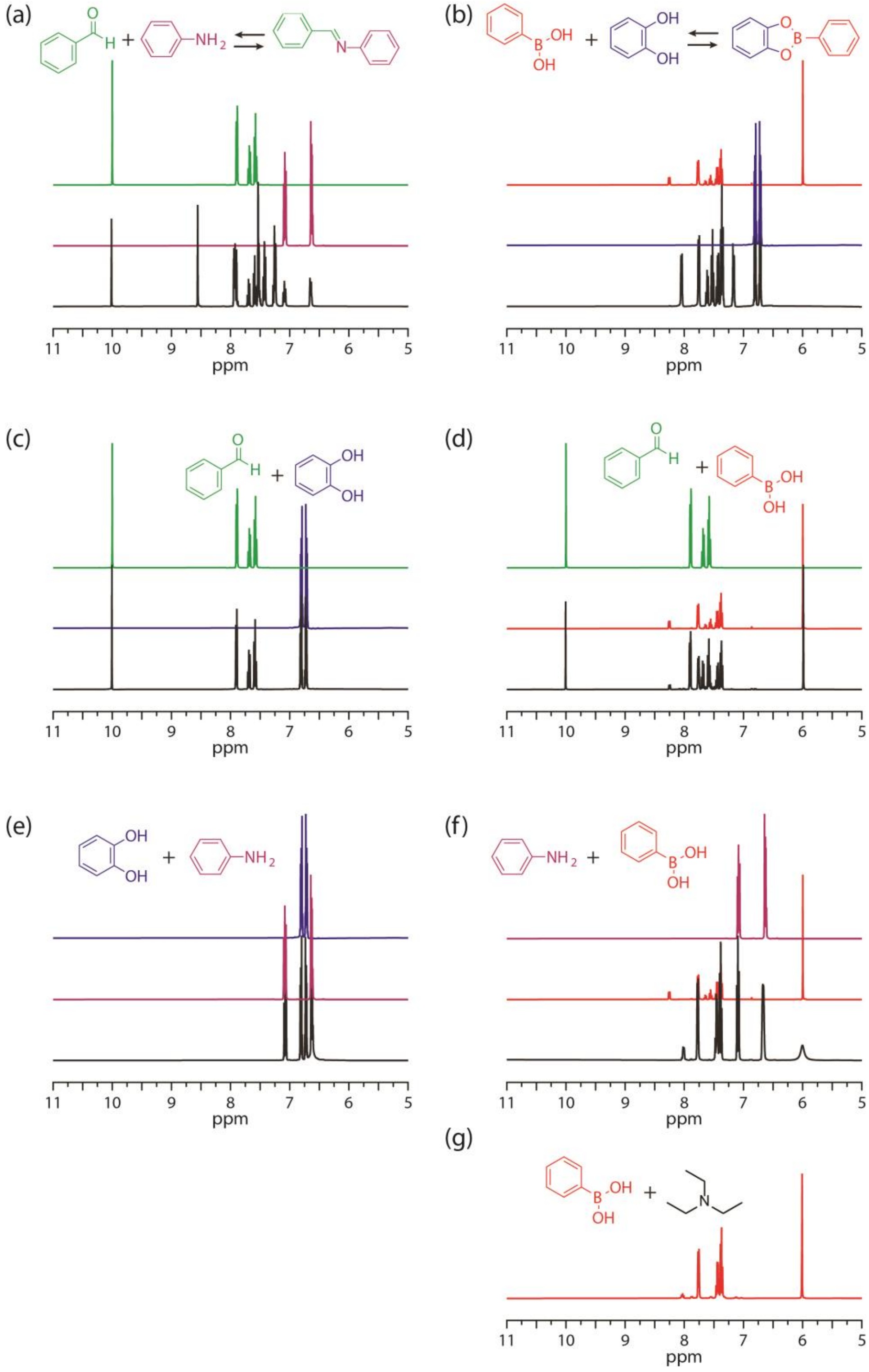
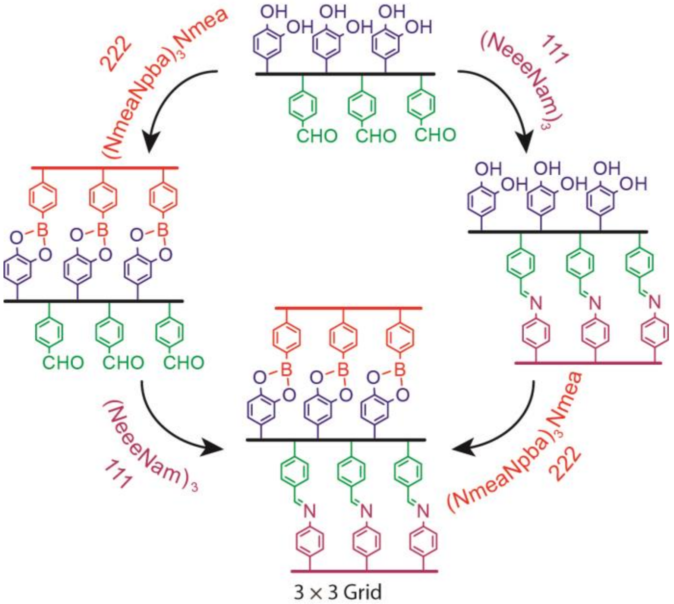


Figure 3.8 Proton nuclear magnetic resonance (H^1 NMR) spectra of cross reactions between model compounds for the dynamic covalent pendant groups. (g) ^{1}H NMR spectra of a phenylboronic acid and triethylamine solution demonstrating the peak shifts observed in the

^{1}H-NMR spectra of the reaction between phenylboronic acid and aniline are caused by the basicity of aniline rather than a coupling reaction.

3.5.3. Base-4 Grid Formation by Sequential Dynamic Covalent Interactions

The use of both boronic acid/catechol and amine/aldehyde reactions to mediate the dynamic covalent assembly of base-4-encoded peptoid oligomers was initially explored via the assembly of a molecular grid bearing both imine- and boronate ester-based rungs. A 3×3 molecular grid (i.e., three oligomeric strands linked by three dynamic covalent interactions per oligomer pair) was assembled from a core peptoid oligomer bearing six covalently reactive pendant groups and flanked by two peptoids each bearing three reactive pendant groups that complement the core. By exploiting the 'Σ-strand' conformation adopted by peptoids, where the pendant groups of adjacent residues are presented on opposite sides of the backbone,[22] the precursor oligomers were designed such that the imine- and boronate ester-based rungs of the assembled grid were organized on opposing flanks of the core peptoid sequence. This core strand incorporated alternating catechol and aldehyde pendant groups, whereas the flanking strands were composed of alternating reactive pendant groups, either amine- or boronic acid-based, respectively, and spacer residues to match the reactive pendant group spacing between the flanking and core strands. To begin, stock solutions of the core sequence (303030) and the amine- (111) and boronic acid-bearing (222) flanking strands in acetonitrile/water were each added at a 2:1:1 stoichiometric ratio to a pH 9 aqueous solution containing a catalytic amount of scandium triflate and gently stirred overnight. The resultant, clear hybridization solution was characterized and the target 3×3 molecular grid was identified by mass spectrometry (Figure 3.7.b), confirming successful dynamic covalent self-assembly using the two orthogonal condensation reactions. Interestingly, whereas the addition of the sequence 303030 to the alkaline aqueous solution yielded a cloudy mixture which persisted upon addition of the sequence 111, the reaction mixture rapidly became clear upon addition of sequence 222 prior to 111, (Figure 3.7.c). To examine the influence of the sequence addition order on the 3×3 molecular grid fabrication, the process for grid assembly was performed *via* two alternative, sequential assembly routes (Scheme 3.4.).



Scheme 3.4 Alternate pathways for the assembly of a base-4 molecular grid *via* the sequential addition of oligomeric sequences bearing boronic acid or amine pendant groups.

Two alkaline aqueous solutions containing sequence 303030 and either sequence 111 or 222 were allowed to react overnight with a catalytic amount of scandium triflate to form intermediate ladder species. As before, the 303030×111 and 303030×222 reaction mixtures were cloudy and clear solutions, respectively. MALDI-TOF mass spectrometry confirmed the formation of the target 303030×111 and 303030×222 molecular ladders (Figure 3.9.); however, misaligned or out-of-registry ladder species were also identified in the cloudy 303030×111 mixture, whereas the clear 303030×222 mixture contained exclusively in-registry duplexes. The final molecular grid was generated by addition of the remaining complementary sequences and again mixing overnight. Although no visible change was observed upon addition of the sequence 111 to the clear 303030×222 mixture, addition of sequence 222 to the cloudy 303030×111 reaction mixture yielded a transparent solution (Figure 3.7.d), suggesting that the boronate ester functionality strongly influences aqueous solubility of the generated structures.

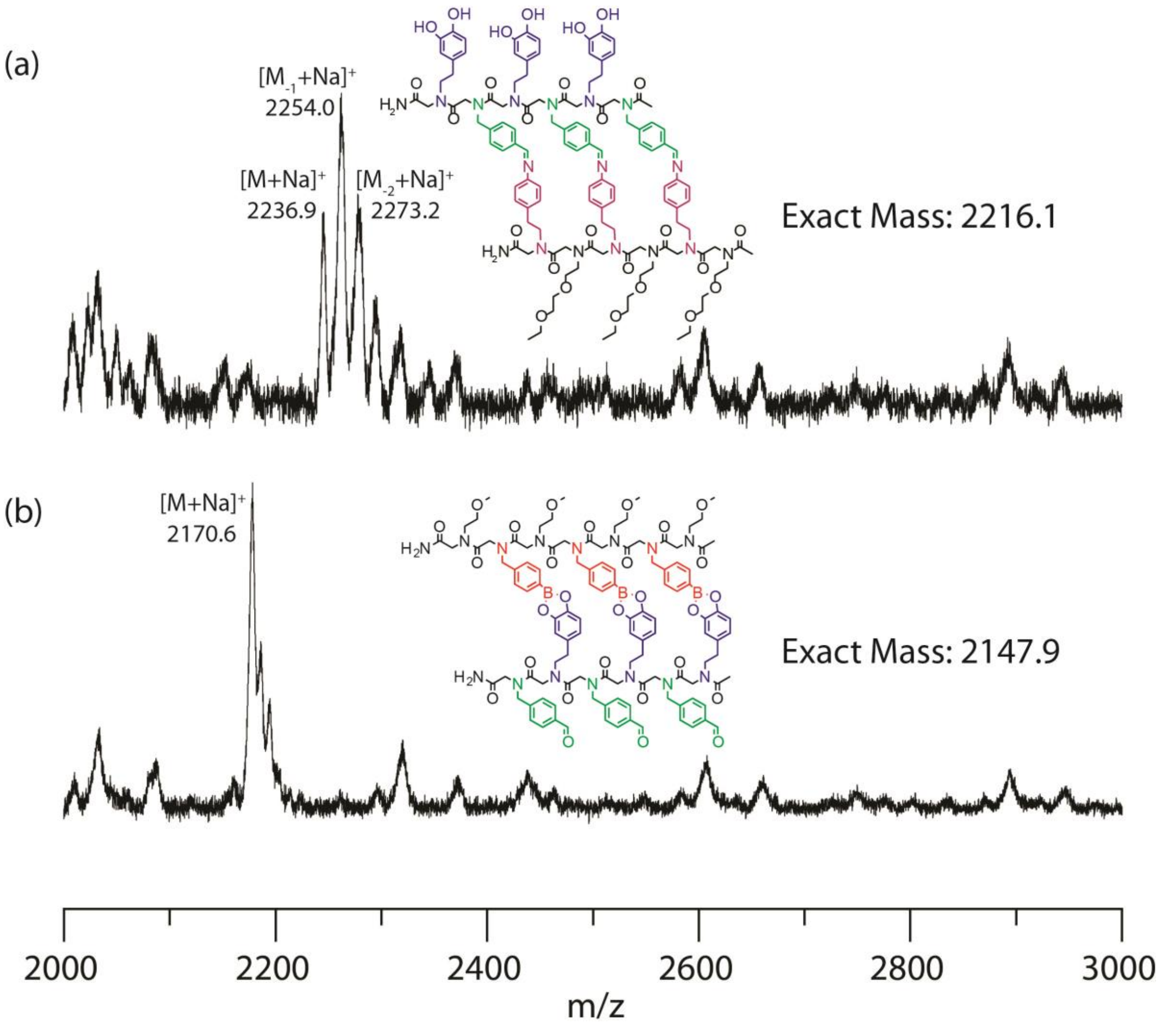


Figure 3.9 Positive mode MALDI-TOF spectra of the two duplexes formed with the catechol and aldehyde core and either the (a) amine- or (b) boronic acid-bearing flanking strands.

3.5.4. Base-4 Self-Assembly

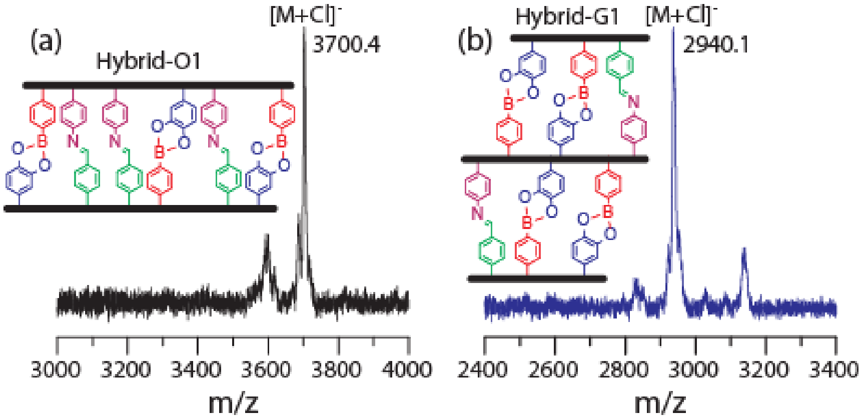


Figure 3.10 Positive mode MALDI-TOF spectra of the two duplexes formed with the catechol and aldehyde core and either the (a) amine- or (b) boronic acid-bearing flanking strands.

Having realized dynamic covalent assembly employing dual concurrent, orthogonal dynamic covalent interactions, we evaluated the ability of the orthogonal reaction strategy to direct the selective hybridization of information-encoded species. To this end, the oligomeric sequences 213112 and 300203 were synthesized as complementary sequences designed to afford a six-rung molecular ladder. To enable facile identification of assembled species by mass spectrometry, sequence 213112 was synthesized with an additional inert spacer residue prior to acetylation. Overnight incubation of sequences 213112 and 300203 at a 1:1 stoichiometric ratio in alkaline aqueous solution afforded the target molecular ladder (Hybrid-O1), as identified by MALDI-TOF mass spectrometry, with no evidence of non-complementary dimeric (Figure 3.10.a) or multimeric species (Figure 3.11.). Notably, parallel orientation of the 213112 and 300203 strands would preclude the generation of an imine and a boronate ester owing to misalignment of the reactive pendant groups; thus, the absence of a peak at +54 m/z (i.e., the sum of +18 and +36 to account for the unreacted imine and boronate ester precursors, respectively) suggests that the assembled molecular ladder consisted of antiparallel-orientated strands. In contrast to the 1:1 stoichiometric system, incubation of sequence 213112 with a 50% excess of 300203 yielded both dimeric and trimeric (consisting of one 213112 strand and two 300203 strands) species as identified by MALDI-TOF mass spectrometry (Figure 3.11.). To further explore sequence-selective hybridization using concurrent and orthogonal dynamic covalent interactions, a 3×3 molecular grid was designed in which the hybridization of the flanking strands is directed by the information encoded in the core strand. Here, again owing to the Σ-strand conformation, the spacer-free oligo(peptoid) sequence 123321 displays the sequence 132 on opposite sides of the backbone. Hybridization of this core strand with two stoichiometric equivalents of the sequence 320 afforded the base-4, 3×3 molecular grid, Hybrid-G1. Characterization of the reaction solution by MALDI-TOF mass spectrometry confirmed the formation of the molecular grid (Figure 3.10.b). Unfortunately, a non-negligible amount of the cross-product 123321×123321×320 ([M]=3160.4 m/z) was also observed, potentially attributable to sequence brevity and consequent lack of mismatched dynamic covalent reactant pairs.

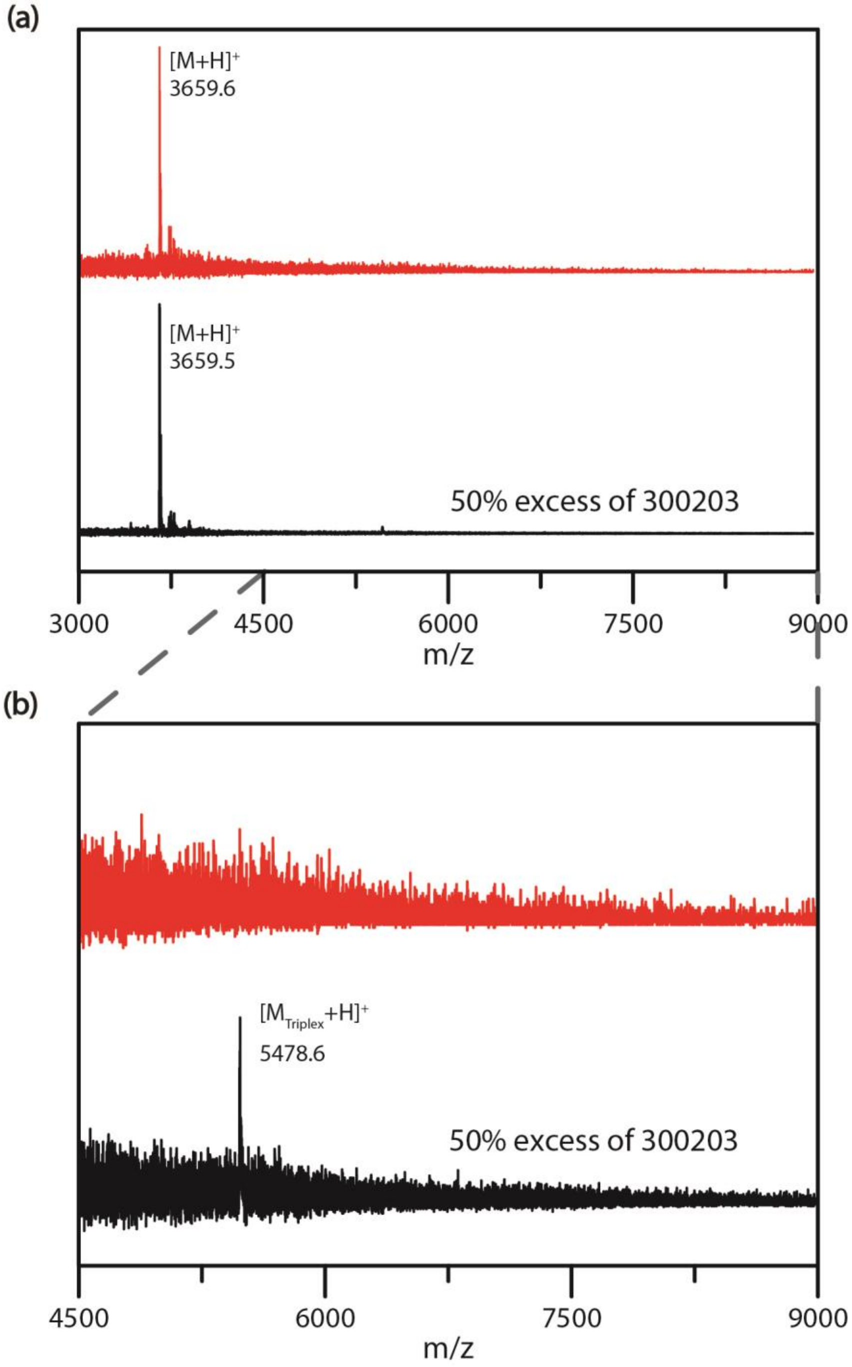


Figure 3.11 (a,b) Positive mode MALDI-TOF spectra of Hybrid-O1 with a 1:1 stoichiometric ratio of sequences 213112 and 300203 (top red) and a 1:1.5 ratio. Expected exact masses: $[M_{Hybrid\text{-}O1}+H]^+$ = 3658.7; $[M_{Triplex}+H]^+$ = 5474.6. $M_{Triplex}$ refers to the mass of a multimeric specie comprised of three strands; one 213112 and two 300203 oligomers.

3.6. Conclusion

In conclusion, we have demonstrated the information-directed assembly of molecular ladders and grids bearing covalent bond-based rungs from encoded precursor strands using dual concurrent, orthogonal dynamic covalent interactions. We anticipate that the information density offered by this system to have broad utility in the development of complex and robust self-assembled nanostructures.

3.7 References

1. Crick, F. & Watson, J. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737 (1953).
2. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* 99, 237–247 (1982).
3. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* 440, 297–302 (2006).
4. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* 485, 623–626 (2012).
5. Santosh, M. & Maiti, P. K. Force induced DNA melting. *J. Phys. Condens. Matter* 21, (2009).
6. Bockelmann, U., Essevaz-Roulet, B. & Heslot, F. Molecular stick-slip motion revealed by opening DNA with piconewton forces. *Phys. Rev. Lett.* 79, 4489–4492 (1997).
7. Hartley, C. S., Elliott, E. L. & Moore, J. S. Covalent assembly of molecular ladders. *J. Am. Chem. Soc.* 129, 4512–4513 (2007).
8. Wei, T., Furgal, J. C., Jung, J. H. & Scott, T. F. Long, self-assembled molecular ladders by cooperative dynamic covalent reactions. *Polym. Chem.* 8, 520–527 (2017).
9. Reuther, J. F. *et al.* Dynamic covalent chemistry enables formation of antimicrobial peptide quaternary assemblies in a completely abiotic manner. *Nat. Chem.* 10, 45–50 (2018).
10. Seifert, H. M., Ramirez Trejo, K. & Anslyn, E. V. Four Simultaneously Dynamic Covalent Reactions. Experimental Proof of Orthogonality. *J. Am. Chem. Soc.* 138, 10916–10924 (2016).
11. Bracchi, M. E. & Fulton, D. A. Orthogonal breaking and forming of dynamic covalent imine and disulfide bonds in aqueous solution. *Chem. Commun.* 51, 11052–11055 (2015).
12. Rodriguez-Docampo, Z. & Otto, S. Orthogonal or simultaneous use of disulfide and hydrazone exchange in dynamic covalent chemistry in aqueous solution. *Chem. Commun.* 5301–5303 (2008) doi:10.1039/b808725c.
13. Hutin, M., Bernardinelli, G. & Nitschke, J. R. An iminoboronate construction set for subcomponent self-assembly. *Chem. Eur. J.* 14, 4585–4593 (2008).
14. Christinat, N., Scopelliti, R. & Severin, K. Multicomponent assembly of boron-based dendritic nanostructures. *J. Org. Chem.* 72, 2192–2200 (2007).
15. Christinat, N., Scopelliti, R. & Severin, K. Multicomponent assembly of boronic acid

based macrocycles and cages. *Angew. Chem. Int. Ed.* 47, 1848–1852 (2008).
16. Dunn, M. F., Wei, T., Zuckermann, R. N. & Scott, T. F. Aqueous dynamic covalent assembly of molecular ladders and grids bearing boronate ester rungs. *Polym. Chem.* 10, 2337–2343 (2019).
17. Leguizamon, S. C. & Scott, T. F. Sequence-selective dynamic covalent assembly of information-bearing oligomers. *Nat. Commun.* 11, 784 (2020).
18. Wei, T., Furgal, J. C. & Scott, T. F. In situ deprotection and dynamic covalent assembly using a dual role catalyst. *Chem. Commun.* 53, 3874–3877 (2017).
19. Yadav, J. S., Reddy, B. V. S. & Reddy, K. S. Yb(OTf)3·H2O: A novel reagent for the chemoselective hydrolysis of isopropylidene acetals. *Chem. Lett.* 430–431 (2001).
20. Kobayashi, S., Sugiura, M., Kitagawa, H. & Lam, W. W. L. Rare-earth metal triflates in organic synthesis. *Chem. Rev.* 102, 2227–2302 (2002).
21. Yan, J., Springsteen, G., Deeter, S. & Wang, B. The relationship among pK a, pH, and binding constants in the interactions between boronic acids and diols - It is not as simple as it appears. *Tetrahedron* 60, 11205–11209 (2004).
22. Edison, J. R. *et al.* Conformations of peptoids in nanosheets result from the interplay of backbone energetics and intermolecular interactions. *Proc. Natl. Acad. Sci. U. S. A.* 115, 5647–5651 (2018).

Chapter 4 Temperature-mediated Molecular Ladder Self-assembly Employing Diels-Alder Cycloaddition

4.1 Abstract

Dynamic covalent self-assembly processes often exhibit poor capacities for error-correction owing to the relatively low connectivity rearrangement rates of dynamic covalent interactions and the common use of reaction conditions where the equilibrium state remains fixed. Here, we report a dynamic covalent self-assembly technique employing temperature as an externally-applied stimulus to mediate the hybridization of peptoid oligomers bearing maleimide- and furan-based pendant groups to afford molecular ladders incorporating Diels-Alder adduct-based rungs. By raising or lowing the reaction temperature, this system enables the equilibrium state to be readily varied without altering reagent concentrations. Both triethylamine and the Lewis acidic scandium triflate were examined as candidate reaction catalysts; however, only scandium triflate increased the rate of single strand conversion. As the Diels-Alder cycloaddition reaction does not liberate a small molecule, a registry-dependent mass change was effected by employing a base-catalyzed thiol-Michael addition reaction between any un-reacted maleimide pendant groups and a low molecular weight thiol to enable the number of Diels-Alder adduct rungs to be readily determined by mass spectrometry. Finally, by employing a slow temperature ramp from high to low temperature, approximating the thermal cycle employed for nucleic acid hybridization, sequence-selective hybridization between model, tetra-functional precursor strands was demonstrated.

4.2 Introduction

Molecular self-assembly is commonly associated with relatively weak, intermolecular bonding as the transience of these interactions provides an inherent rearrangement mechanism to enable error correction. Nevertheless, dynamic covalent interactions have emerged as indispensable reactions to mediate nanoarchitecture fabrication, wherein the structural complexity attainable by self-assembly is combined with the strength of covalent bonds[1–3]. These dynamic, rearrangeable covalent interactions have been used extensively for the synthesis of structurally diverse, mechanical and chemical robust materials including covalent organic frameworks (COFs)[4–6], molecular cages[7–9], and molecular ladders[10–12]. In comparison with supramolecular interactions, dynamic covalent reactions suffer from relatively low dissociation rates which impede rearrangement amongst the assembled components and can result in the kinetic trapping of non-equilibrium species[13–16]. To overcome this limitation, dynamic covalent assembly processes are generally restricted to those employing constituent precursors bearing few reactive sites or the utilization of harsh reaction conditions[1]. Conversely, the inherently transient nature of supramolecular interactions permits the use of non-covalent building blocks with high degrees of functionality and mild reaction conditions to generate complex molecular architectures[17–19]. Particularly suitable for the formation of multi-dimensional structures, nucleic acids are often employed as sophisticated nanoconstruction media where encoded sequences self-assemble in a predetermined manner upon the gradual cooling of dissociated strands from raised temperatures[20–22]. In a recent effort to mimic the hybridization of complementary nucleic acid sequences with precursors bearing dynamic covalent reaction pairs, we reported a step-wise process involving the dissociation of encoded peptoid (i.e., oligo(*N*-substituted glycine)) strands bearing amine- and aldehyde-pendant species by treatment with large amounts of a Lewis acidic rare-earth metal triflate, $Sc(OTf)_3$, and subsequent extraction of this metal triflate to promote imine formation and error-correction of information-bearing molecular ladders[23]. Similarly, base-4-encoded peptoid strands bearing two sets of orthogonal dynamic covalent reaction pairs capable of forming either imine- or boronate ester-rungs, exhibited sequence-selective self-assembly using a mechanism in which acid-dissociated strands were introduced to alkaline conditions to promote rung formation between complementary sequences[24]. Whereas these methods moderate the kinetic trapping that prevails in dynamic covalent assembly systems to enable the self-assembly of complementary oligomeric sequences, both processes are limited by an abrupt equilibrium shift from dissociation to association and are dissimilar to the gradual temperature ramp employed for nucleic acid hybridization. The development of a covalent self-assembly system where the reaction equilibrium is progressively shifted is more readily realized with dynamic covalent reactions that are affected by an external stimulus (e.g., temperature, light intensity, etc.) than those influenced by a chemical stimulus (e.g., pH, reagent concentration, etc.). One such thermally-sensitive dynamic covalent interaction, the Diels-Alder cycloaddition between a maleimide and a furan, has been used extensively in self-healing polymeric materials[25,26]; however, it is rarely employed as a mechanism to mediate self-assembly processes. Here we demonstrate the temperature-mediated self-assembly of oligomers bearing maleimide- and furan-pendant groups to afford molecular ladders with Diels-Alder adduct-based rungs.

4.3 Experimental

4.3.1. General Experimental Procedure

^1H NMR spectra of the monomers were collected using a Bruker Avance 400 NMR spectrometer. Chemical shifts were measured in δ (ppm) relative to residual solvent (CD_3CN = 1.94, D_2O = 4.79, and $CDCl_3$ = 7.26). Electrospray ionization (ESI) mass spectra were recorded using an Agilent UPLC 1290 Q-TOF 6500 series spectrometer in positive ion mode. Reverse phase high performance liquid chromatography (RP-HPLC) was performed using both a preparative reversed phase Phenomenex Luna C18(2) column with a linear gradient of water and acetonitrile as the eluent at 30°C as well as an analytical scale column. Gel permeation chromatography (GPC) was performed using three Phenogel GPC/SEC columns (length 300 mm × diameter 7.8 mm, pore sizes of 500, 100, and 50 Å) in series, 92:6:2 (v/v/v) $CHCl_3$:MeOH:Et_3N as the eluent at room temperature. The RP-HPLC and GPC systems were equipped with a dual-parallel Shimadzu LC-20AD HPLC pump, a Shimadzu FRC-10A fraction collector, and monitored using a Shimadzu Prominence UV-Vis detector at 214, 254, or 313 nm. Unless otherwise noted, all reagents and materials were purchased from Sigma Aldrich, AK Scientific, Oakwood Products, and TCI America.

4.3.2. Monomer Synthesis

Synthesis of the *exo*-isomer of compound **1** was adapted from a published approach: A round bottom flask containing 20 g (204 mmol) of maleic anhydride was charged with 160 mL of toluene, stirred for 30 minutes and filtered. 16.64 g (244 mmol) of furan was added to the filtrate and stirred for 48 hrs at room temperature. The precipitate was collected, filtered, and washed with cold toluene to yield a white solid (22.03 g, 65% yield). ^{1}H NMR (400 MHz, CDCl$_3$) δ: 6.58 (2 H, s, -CHC*H*=C*H*CH-), 5.46 (2 H, s, -C*H*CH=CHC*H*-), 3.17 (2 H, s, O=CC*H*).

Synthesis of compound **2**: Compound 1 (10 g, 60 mmol), ethanol (30 mL), and trimethylamine (TEA, 10 mL) were placed in a 250 mL round bottom flask equipped with a magnetic stirrer. To this, a solution of *tert*-butyl-2-aminotheyl carbamate (10.12 g, 63.1 mmol) and 20 mL ethanol were added dropwise with continuous stirring, and mixed for one hour. The resulting mixture was refluxed at 85°C, using a reflux condenser, for 4 hrs. The solution was cooled to room temperature overnight and the white solid was filtered off and recrystallized from ethanol. The collected crystals were dried under vacuum to afford 2 (9.39 g, 51% yield). ^{1}H NMR (400 MHz, CDCl$_3$) δ: 6.51 (2 H, s, -CHC*H*=C*H*CH-), 5.26 (2 H, s, -C*H*CH=CHC*H*-), 3.63 (2 H, t, NC*H*$_2$CH$_2$NHC=O), 3.31 (2 H, t, NCH$_2$C*H*$_2$NHC=O), 2.85 (2 H, s, O=CC*H*), 1.41 (9 H, t, C(C*H*$_3$)$_3$).

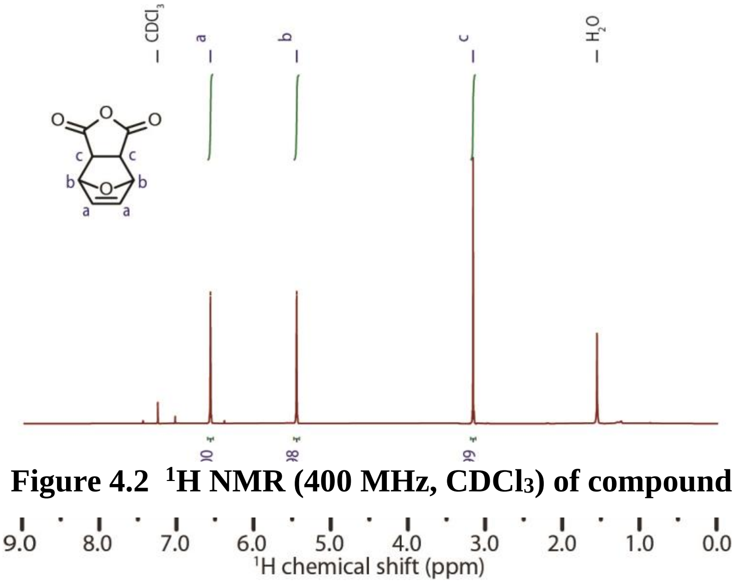
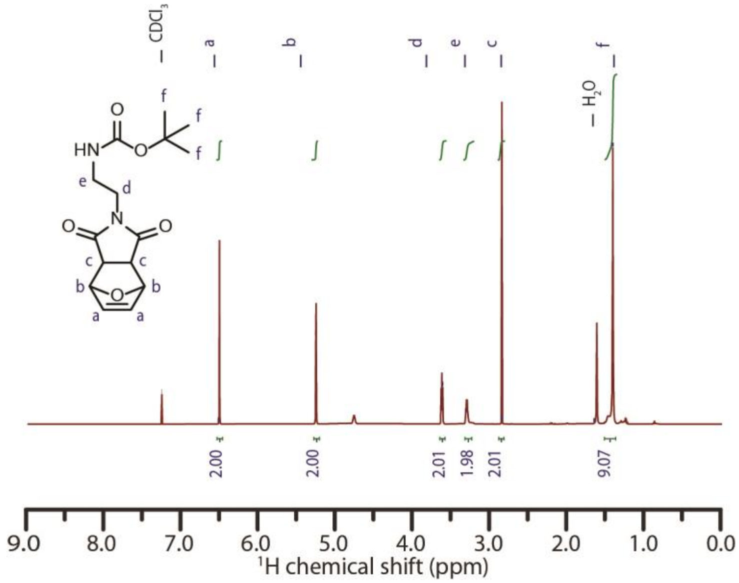


Figure 4.2 ^{1}H NMR (400 MHz, CDCl$_3$) of compound 2.

Figure 4.1 ^{1}H NMR (400 MHz, CDCl$_3$) of compound 1.



Synthesis of compound 3, Nfpm: In a 500 mL round bottom flask, compound 2 (9.4 g, 30.5 mmol) was dissolved in a 3:7 mixture of trifluoroacetic acid (TFA)/dichloromethane (DCM) and stirred at room temperature for one hour before blowing off with a gentle stream of nitrogen and subsequent solvent removal by evaporation. The resulting oil was added to diethyl ether to precipitate. After centrifugation, the supernatant was discarded and diethyl ether was again added to wash the product. This was repeated three times before collecting and drying the product under vacuum, quantitatively yielding compound 3 as a white solid. ^{1}H NMR (400 MHz, D_2O) δ: 6.65

(2 H, s, -CHC*H*=C*H*CH-), 5.37 (2 H, s, -C*H*CH=CHC*H*-), 3.86 (2 H, t, NC*H*$_2$CH$_2$NH$_2$), 3.31 (2 H, t, NCH$_2$C*H*$_2$NH$_2$), 3.21 (2 H, s, O=CC*H*).

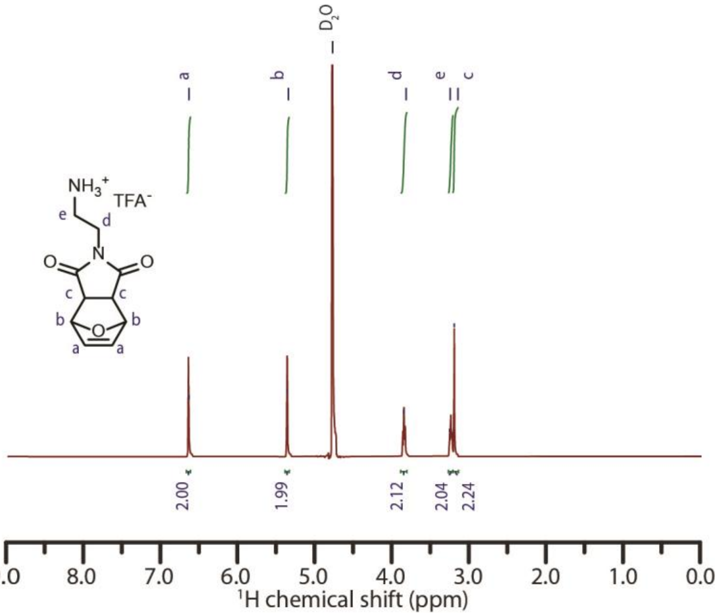


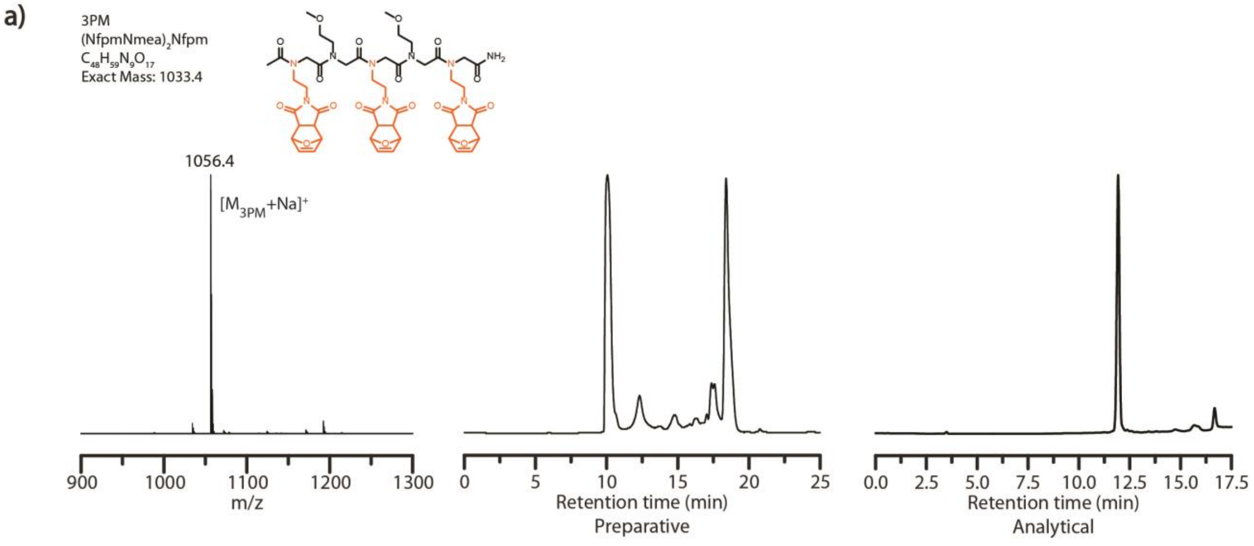
Figure 4.3 ^{1}H NMR (400 MHz, D$_2$O) of compound 3.

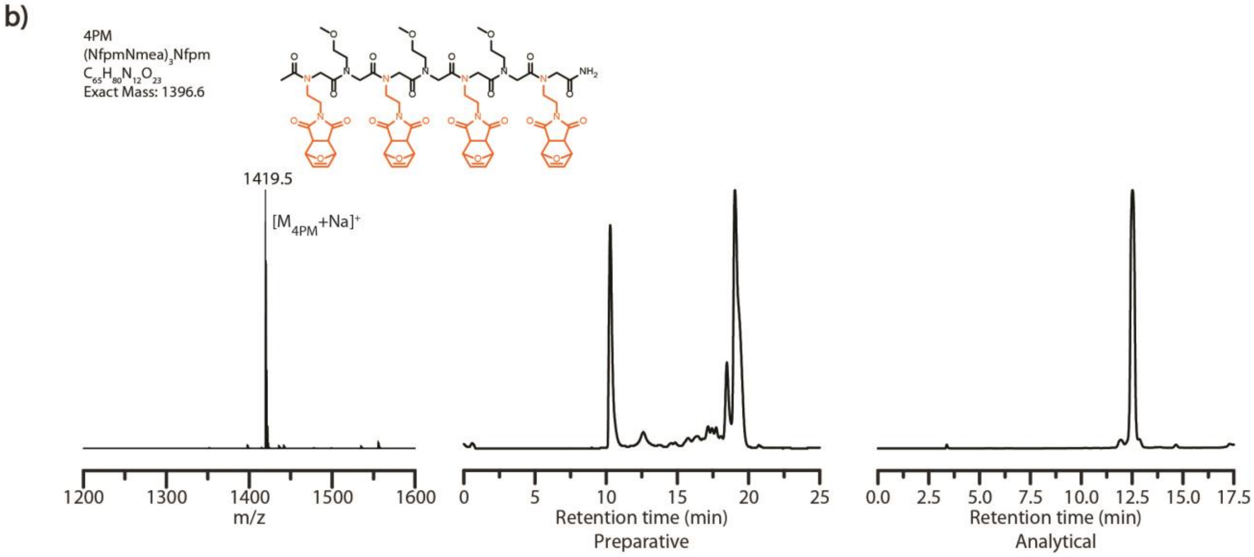
4.3.3. Oligomer Synthesis

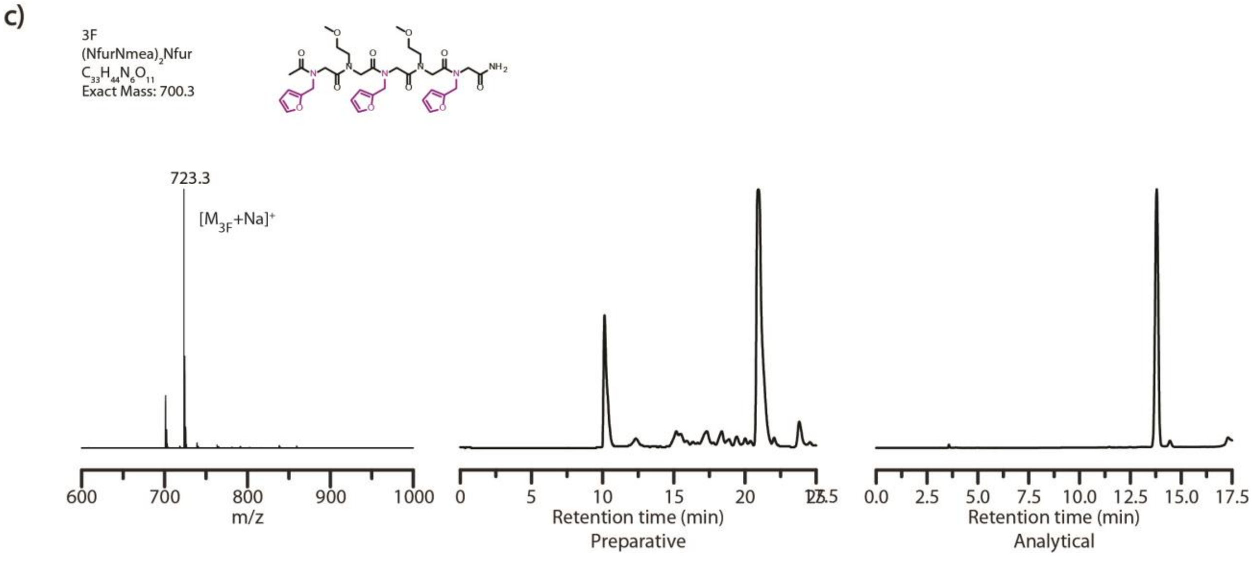
The peptoid-based oligomers were prepared using a microwave-assisted Liberty Blue automated peptide synthesizer (CEM Corporation) programmed with a peptoid synthetic method. Peptoids were synthesized on acid-labile Rink amide resin (ChemPep) bearing a fluorenylmethyloxycarbonyl (Fmoc)-protected amine 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 sequential addition reactions 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. Notably, addition of Nfpm, the furan-protected maleimide-bearing primary amine monomer, additionally requires 1.3 equivalents of *N*,*N*-diisopropylethylamine (DIPEA) to deprotonate the primary amine during the nucleophilic substitution. 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. The dynamic covalent functional group consisted of the prepared Nfpm and commercially available furfuryl amine, Nfur. The inert spacer monomers were the commercially available 2-methoxyethylamine (Nmea) and 2-ethoxyethoxyethylamine (Neee) prepared according to the established protocol of Wei et al.[2] All of the reagents were prepared in dimethylformamide (DMF).

4.3.4. Oligomer Cleavage and Purification

Peptoids were cleaved from acid-labile Rink amide resin by a 10 minute incubation with a cleavage cocktail containing 95% TFA and 5% water in a glass fritted reaction vessel. The resin was then rinsed with TFA to remove any residual peptoid on the resin. TFA and trace water were removed by blowing with a N_2 stream. Once only the peptoid residue remained, the residue was dissolved in a 50:50 solution of acetonitrile and water and purified by preparative scale reverse phase-HPLC using a linear gradient of acetonitrile and water. Major peaks were collected and characterized by ESI mass spectrometry to confirm the identity of each strand. Fractions containing the peptoids were lyophilized to a white powder.


a)
3PM
(NfpmNmea)₂Nfpm
C₄₈H₅₉N₉O₁₇
Exact Mass: 1033.4
1056.4
[M₃PM+Na]⁺
900
1000
1100
1200
1300
m/z
0
5
10
15
20
25
Retention time (min)
Preparative
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Retention time (min)
Analytical


b)
4PM
(NfpmNmea)₃Nfpm
C₆₅H₈₀N₁₂O₂₃
Exact Mass: 1396.6
1419.5
[M₄PM+Na]⁺
1200
1300
1400
1500
1600
m/z
0
5
10
15
20
25
Retention time (min)
Preparative
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Retention time (min)
Analytical


c)
3F
(NfurNmea)₂Nfur
C₃₃H₄₄N₆O₁₁
Exact Mass: 700.3
723.3
[M₃F+Na]⁺
600
700
800
900
1000
m/z
0
5
10
15
20
25
Retention time (min)
Preparative
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Retention time (min)
Analytical

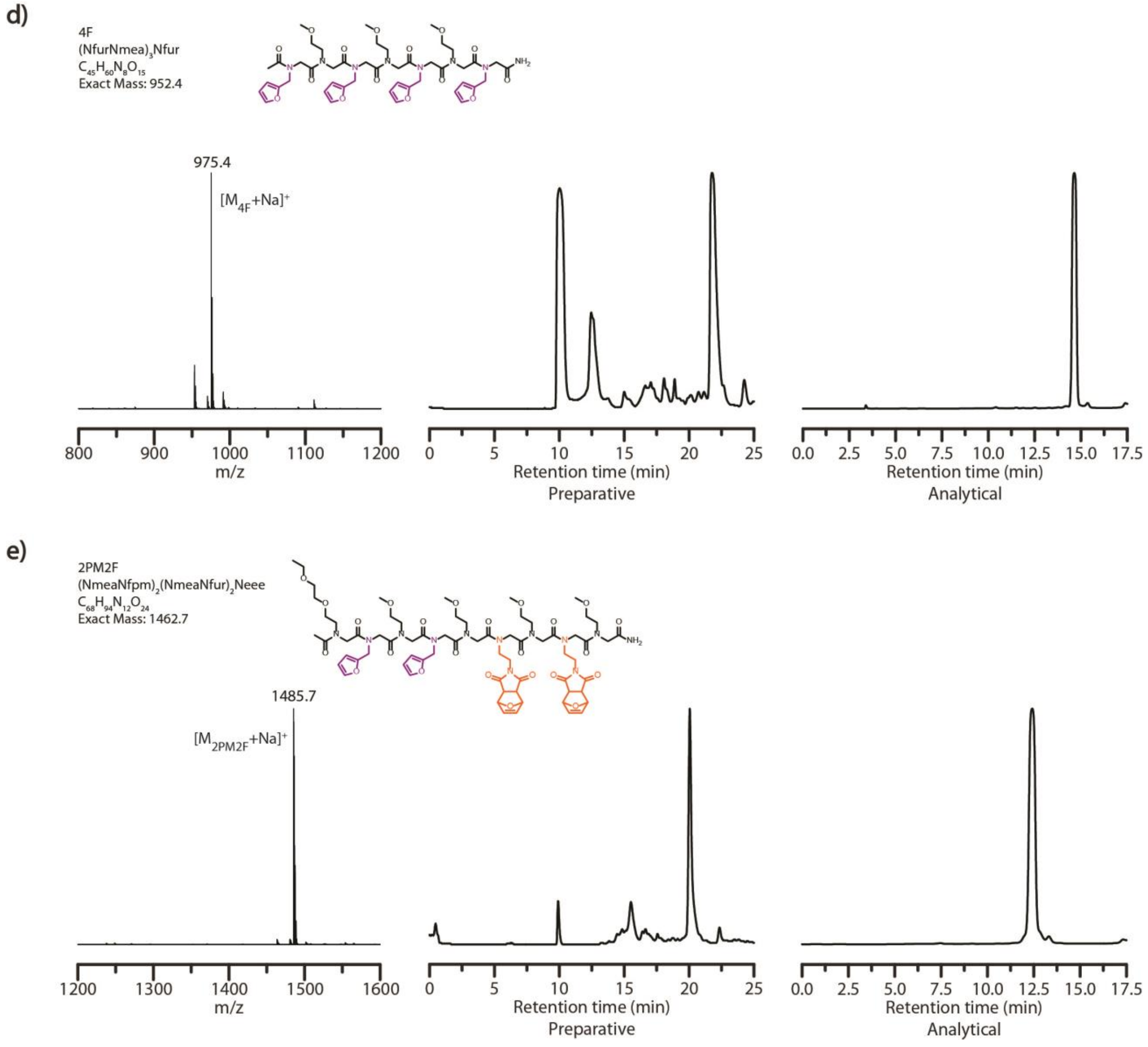


Figure 4.4 ESI mass spectra, preparative HPLC traces, and corresponding analytical HPLC traces of peptoids synthesized for this study.

4.3.5. Oligomer Deprotection

Peptoids were resuspended in acetonitrile to afford 10 mM stock solutions. 20 µL of furan-protected maleimide-bearing oligomers was added to 100 µL of anhydrous anisole and stirred at 140°C, or other specified temperature, in a vial without a cap for 30 minutes such that the adduct undergoes the retro-Diels-Alder reaction. For HPLC and NMR analysis of deprotected solutions, mixtures were subjected to vacuum evaporation under reduced pressure and resuspended in deuterated acetonitrile.

4.3.6. Molecular Ladder Formation

Dissociation and deprotection of peptoids was performed by adding 20 µL from a 10 mM stock solution of each desired oligo(peptoid) sequence to a 100 µL solution of anhydrous anisole which was stirred at 140°C for 30 minutes in a vial without a cap prior to cooling either 60°C or 100°C and addition of a further 200 µL of anhydrous anisole. TEA was often added to reaction solutions and stirred for 15 minutes before analysis ESI-MS to increase ionization efficiency of ladder species.

4.3.7. Molecular Ladder Degree of Alignment Analysis

To strand and hybridization mixtures, 5 µL of TEA and various equivalents of methyl 3-mercaptopropionate (S) per maleimide moiety were added and stirred for 30 minutes prior to analysis. For HPLC and NMR analysis of deprotected solutions, mixtures were evaporated under reduced pressure and resuspended in deuterated acetonitrile

4.4 Results and Discussion

4.4.1. Monomer Design

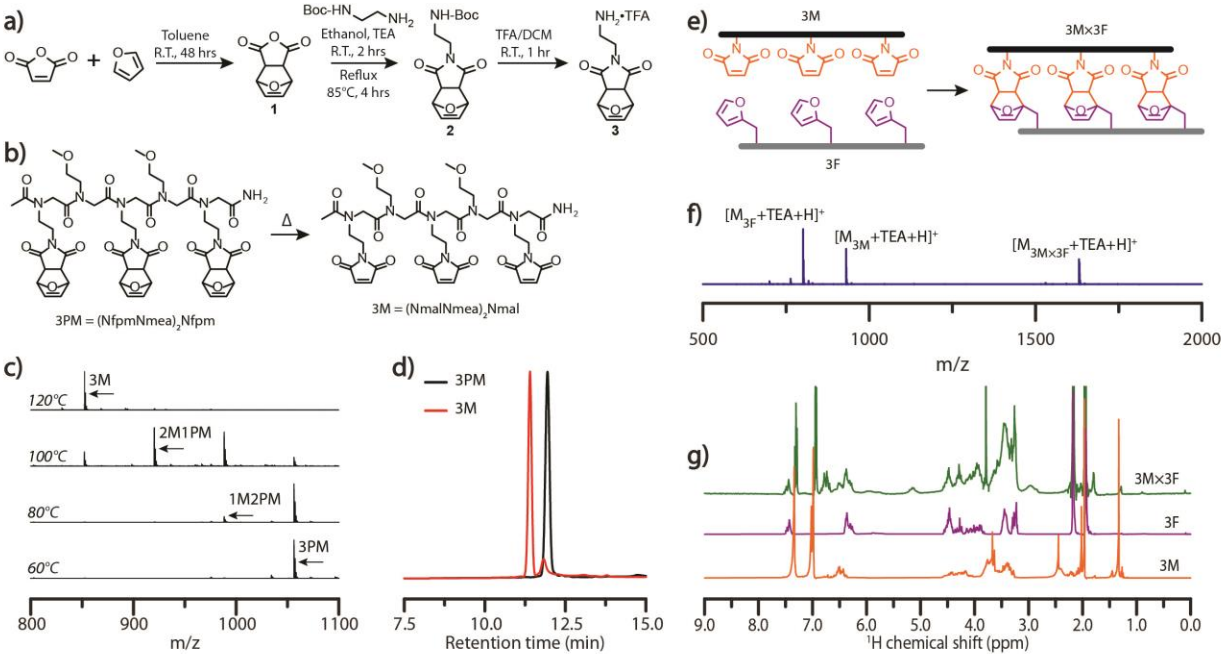


Figure 4.5 Dynamic covalent assembly of Diels-Alder-based molecular ladders. (a) Synthetic scheme for the furan-protected, maleimide-bearing peptoid monomer, 3 and (b) chemical structures depicting the deprotection of the Diels-Alder adduct-bearing peptoid, 3PM, to afford the maleimide-bearing oligomer 3M. (c) ESI mass spectra generated after exposing solutions of 3PM strands to 60°C, 80°C, 100°C, or 120°C for one hour. Here, 3M is the fully deprotected oligomer, while 2M1PM and 1M2PM represent partially deprotected oligomers bearing one and two furan-protected maleimide groups, respectively. (d) HPLC traces of a solution of 3PM single strands prior to (black) and after deprotection at 140°C for 30 minutes (red). (e) Idealized schematic diagram of the hybridization of complementary, trimeric maleimide- and furan-bearing oligomers (3M and 3F, respectively) to afford a peptoid-based molecular ladder bearing three Diels-Alder adduct rungs (3M×3F). (f) ESI mass spectrum of a 3PM and 3F hybridization reaction mixture after heating at 140°C for 30 minutes, then at 60°C for five days. Expected exact masses: $[M_{3F}+TEA+H]^+$ = 802.4; $[M_{3M}+TEA+H]^+$ = 931.4; $[M_{3M\times 3F}+TEA+H]^+$ = 1631.7. Exact masses found: $[M_{3F}+TEA+H]^+$ = 802.4; $[M_{3M}+TEA+H]^+$ = 931.5; $[M_{3M\times 3F}+TEA+H]^+$ = 1631.8. (g) ^{1}H-NMR spectra of the maleimide-bearing peptoid 3M (bottom, orange), the furan-bearing peptoid 3F (middle, purple), and a 3M×3F molecular ladder reaction mixture (top, green).

To examine the utility of the Diels-Alder cycloaddition reaction as a dynamic covalent interaction capable of mediating molecular self-assembly, molecular ladders were employed as target structures owing to their demonstrated capacity for sequence-selective assembly from information-bearing precursors with large numbers of reactive sites, and their facile characterization[23,27,28]. By employing the sub-monomer solid phase synthetic method, sequence-specific peptoid precursor strands bearing furan and maleimide pendant groups were synthesized from readily available primary amines[29].

To preclude potential hydrolysis or Michael addition side reactions, or premature Diels-Alder cycloaddition during the synthesis and purification of maleimide-bearing oligomers, a furan-protected maleimide (3) was synthesized as outlined in Figure 4.5. The Diels-Alder reaction commonly generates a mixture of *endo* and *exo* diastereomers, where the cyclo-reversion of the *endo* adduct proceeds at a lower temperature (~60°C) than the *exo* adduct (~110°C)[30,31]. As the peptoid synthesis is performed at 75°C, the *endo* form is ill-suited as a maleimide-protecting group here owing to its potential for undergoing premature deprotection. To preclude the deleterious influence of *endo* adducts, we employed an approach to selectively yield the *exo*-Diels-Alder cycloaddition adduct (**1**) from furan and maleic anhydride by toluene precipitation, described elsewhere[32]. *N*-Boc-ethylenediamine was added to compound **1** via a ring-opening mechanism and subsequently cyclized in ethanol under reflux to afford compound **2** which was then treated with trifluoroacetic acid (TFA) in dichloromethane to yield the final, furan-protected maleimide monomer (**3**). Furfuryl amine, 2-methoxyethylamine, and compound **3**, were employed as furan-bearing, inert spacer, and protected maleimide-bearing primary amine monomers, respectively, in the microwave-assisted synthesis of a small library of peptoids (Figure 4.4). The deprotonated form of compound **3** was generated *in situ* by the addition of *N*,*N*-diisopropylethylamine to the reaction mixture, ensuring reactivity of the monomer in the nucleophilic displacement of the terminal bromide.

4.4.2. Maleimide Deprotection

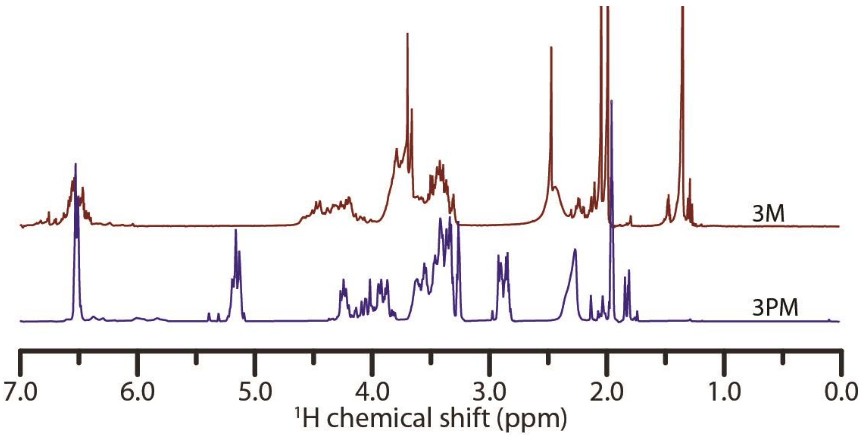


Figure 4.6 ^{1}H NMR (400 MHz, CD_3CN) of a trimeric peptoid bearing maleimide species before (bottom, blue) and after deprotection (top, red).

A peptoid oligomer bearing three furan-protected maleimide pendant groups, 3PM, was used to evaluate the extent of deprotection at elevated temperatures as incomplete deprotection would impede the subsequent self-assembly process (Figure 4.5b). Characterization of 3PM sequences deprotected for one hour at various elevated temperatures in anhydrous anisole, used as a high boiling-point solvent, was performed by electrospray ionization mass spectrometry (ESI-MS), configured with low cone voltage to eliminate in-source fragmentation. Here, the deprotection of a maleimide pendant group via the retro-Diels-Alder reaction yielded a mass change of -68 owing to the loss of furan. Mass spectra indicated an onset of the retro-Diels-Alder reaction at 80°C with complete deprotection at temperatures of 120°C (Figure 4.5c) and above, consistent with similar studies on *exo*-Diels-Alder isomers[31]. Notably, the maleimide pendant groups remained deprotected after cooling of the reaction mixture, attributable to the high volatility and consequent evaporation of furan at raised temperature. Reverse-phase HPLC of 3PM oligomers and the corresponding deprotected oligomer, 3M, confirmed the formation of a new species with a decreased retention time (Figure 4.5d.). The chromatogram of the post-thermal cycle 3M reaction mixture did show a small peak with a retention time similar to that of the 3PM sequence; however, this peak could not be identified by ESI-MS. Further characterization by ^{1}H-NMR demonstrated the quantitative deprotection of 3PM strands, as shown by the absence of peaks around 5.14 and 2.86 corresponding to the (-C*H*CH=CHC*H*-), and (O=CC*H*) protons, respectively, of the Diels-Alder adduct (Figure 4.6).

4.4.3. Initial Molecular Ladder Formation

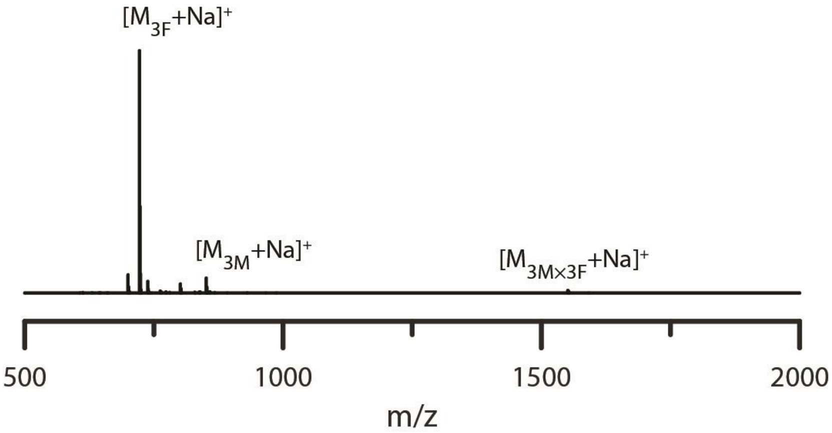


Figure 4.7 Mass spectra of a 3M×3F hybridization solution after five days at 60°C and absent any catalyst.

Initial ladder formation was established by heating reaction mixtures of complementary trimeric furan-protected maleimide- and furan-bearing oligomers, 3PM and 3F, respectively, to 140°C to simultaneously deprotect the maleimide-bearing peptoid and dissociating the two sequences, then subsequently annealing the reaction mixture at 60°C for five days (Figure 4.5e). Characterization of the resulting hybridization mixtures by ESI-MS confirmed the generation of the target ladder species (3M×3F). Although the 3M×3F molecular ladder and residual 3M oligomer both exhibited low intensity signals (Figure 4.7), addition of triethylamine (TEA), a known catalyst for Diels-Alder cycloaddition[33], to a reaction mixture which was subjected to the aforementioned thermal dissociation/annealing process increased the peak intensities (Figure 4.5f). ^{1}H-NMR and diffusion ordered spectroscopy (DOSY) NMR further verified the occurrence of a binding event with the appearance of adduct signals around 5.14 (C*H*CH=CH) and 2.95 (O=CC*H*) and a decrease in the diffusion coefficient upon complexation (Figure 4.5g and Figure 4.8).

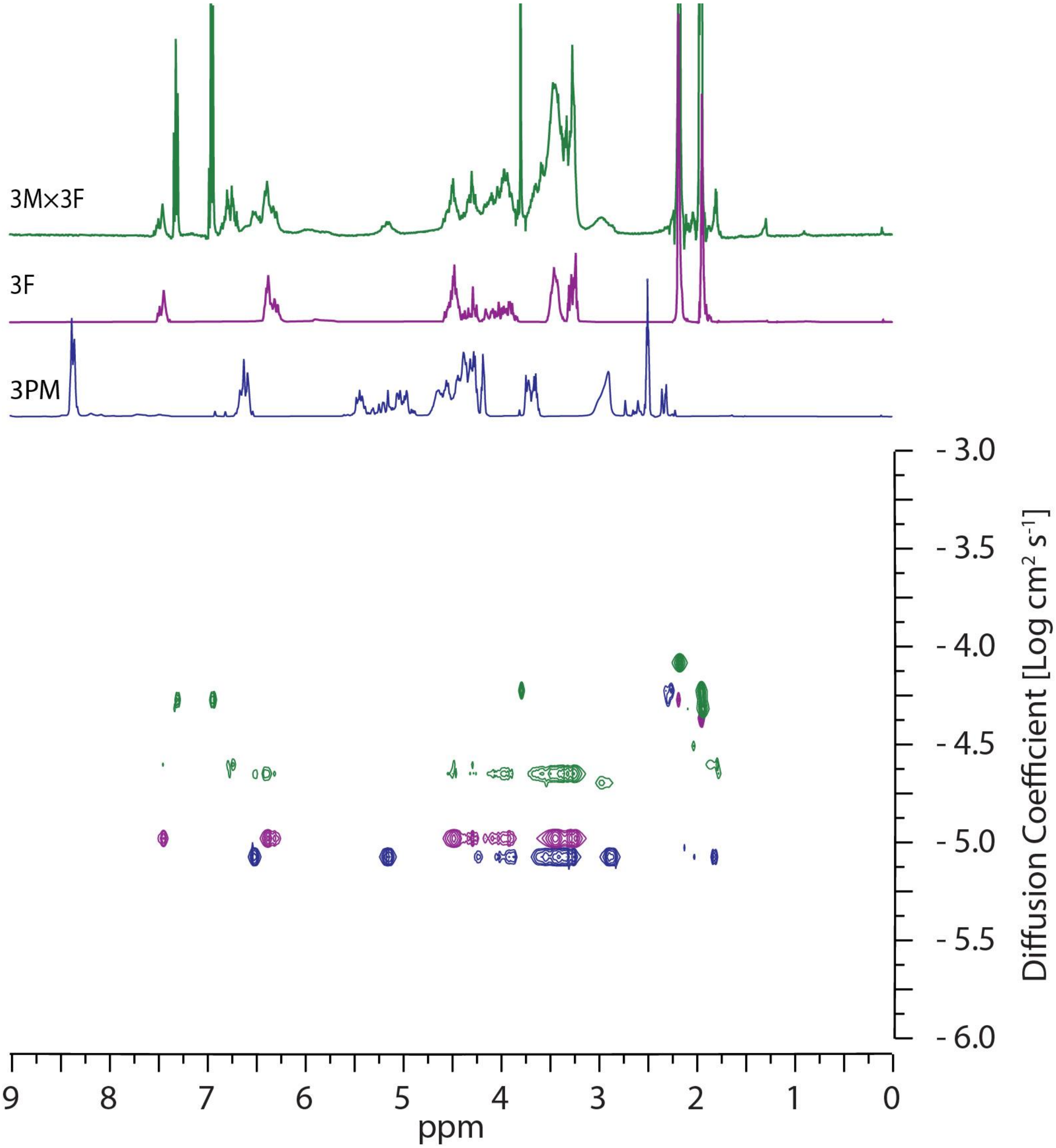


Figure 4.8 DOSY NMR spectra of 3PM (bottom, blue), 3F (middle, purple) and a 3M×3F molecular ladder mixture (top, green) after reaction for two days at 60°C and absent any catalyst. As hybridization reactions in deuterated anisole generated significant amounts of precipitate, CD_3CN was used as the solvent for DOSY NMR studies.

4.4.4. Gel Permeation Chromatography

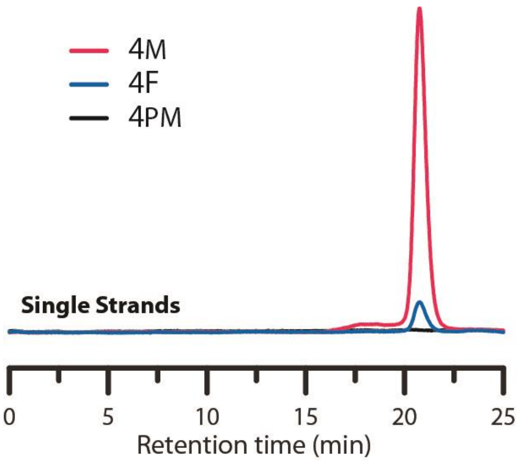


Figure 4.9 GPC traces of sequence-defined oligomers using a UV-Vis detector at 313 nm.

Signal intensity of ESI-MS analytes is dependent on a variety of factors including sample concentration, ease of analyte ionization, and matrix or solute interference, preventing the quantitative analysis of dissimilar species (e.g., single strand and dimeric ladder species)[34]. Consequently, gel permeation chromatography (GPC), was employed to examine the formation of dimeric and high-order ladder species. Interestingly, the protected maleimide-bearing peptoids, mimicking the Diels-Alder-rung molecular ladder species, were undetectable by GPC analysis using a UV-Vis detector at 254 or 313 nm, while the furan-bearing peptoid presented only a minor peak (Figure 4.9). Nevertheless, peptoids bearing deprotected-maleimide species were found to absorb strongly at 313 nm, enabling the unreacted maleimide concentration in single strand or hybrid species to be monitored over time by spiking reaction mixtures with inert, low dispersity polystyrene at a known concentration as an internal standard. Oligomer sequences with four reactive species per strand were employed for this study to augment the adsorption signal (Figure 4.10a). As the Diels-Alder reaction proceeds slowly at room temperature, the period each aliquot experienced in the GPC system was assumed to have negligible effects on the overall conversion.

This assumption was verified by analyzing aliquots of a hybridization reaction mixtures reacted at room temperature with and without a catalyst (Figure 4.11), where negligible formation of dimeric or multimeric ladder species was observed even after 72 hours.

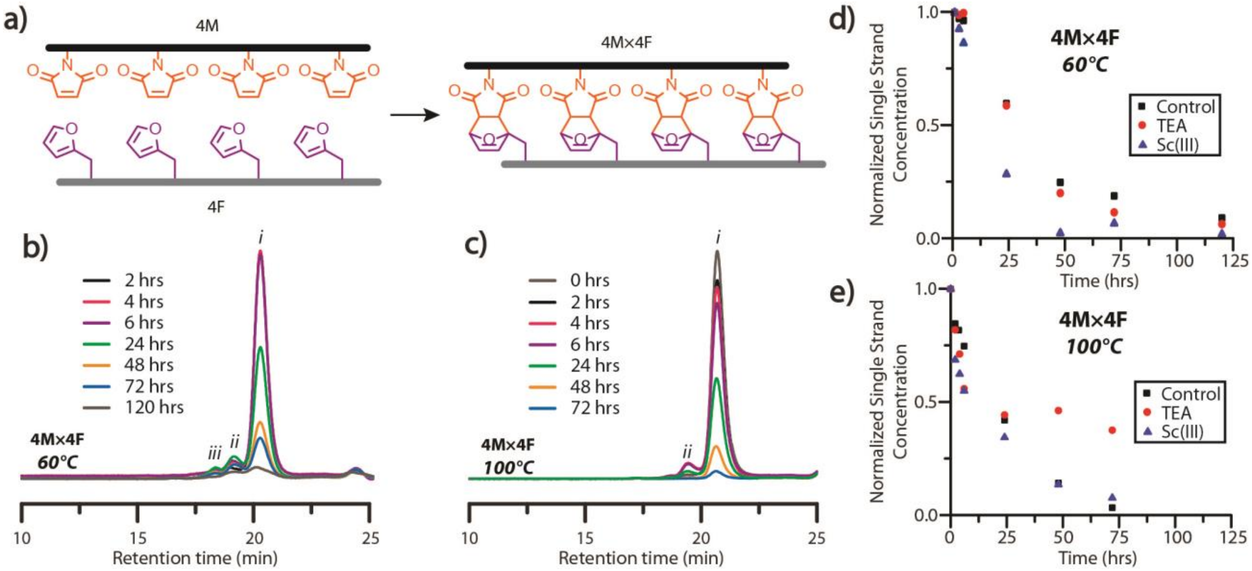


Figure 4.10 Time-resolved characterization of molecular ladder formation. (a) Idealized schematic diagram of the self-assembly of tetramaleimide (4M) and tetrafuran (4F) precursor strands to afford a molecular ladder bearing four Diels-Alder adduct-based rungs (4M×4F). GPC traces, normalized to a polystyrene internal standard, of aliquots periodically sampled from 4M×4F hybridization reaction mixtures reacted at (b) 60°C and (c) 100°C, where *i*, *ii*, and *iii* indicate single stranded, dimeric, and multimeric species, respectively, bearing residual maleimide groups. Normalized concentrations of single strand species in hybridization mixtures with and without catalysts and reacted at (d) 60°C and (e) 100°C, determined from deconvolution of overlapping GPC trace peaks by Gaussian curve fitting.

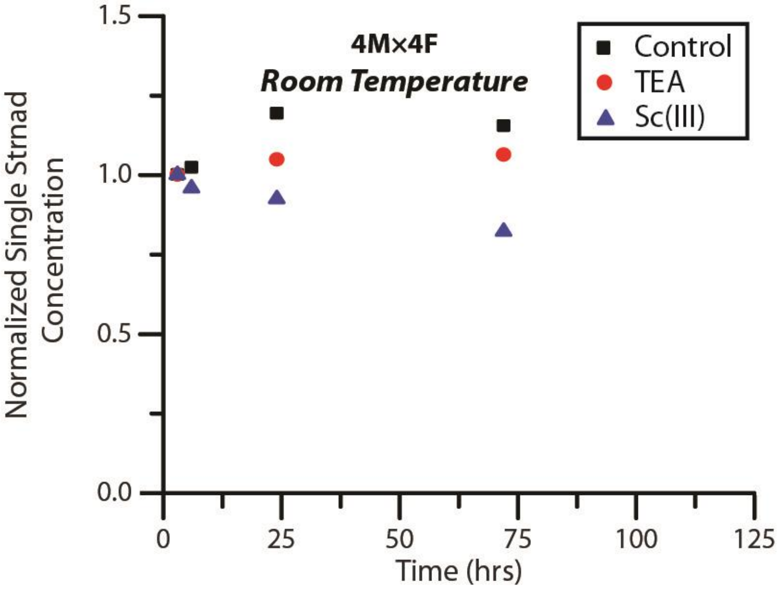


Figure 4.11 Normalized concentrations of single strand species in hybridization mixtures with and without added catalysts at room temperature found by deconvoluting GPC traces with fitted Gaussian functions and normalized to an internal polystyrene standard.

GPC analysis of aliquots from hybridization reaction mixtures at 60°C and 100°C showed the generation of dimeric ladder species within 2 hours, and high-order, multimeric species at 4 hours (Figure 4.10b and c). Multimeric molecular ladder species were consistently observed in aliquots of the 60°C reaction mixture (Figure 4.10b), indicating the persistence of out-of-registry ladders throughout the reaction. In contrast, negligible amounts of dimeric and multimeric species were observed in aliquots of the 100°C reaction mixture from 48 hours (Figure 4.10c), indicating the complete consumption of residual maleimide groups associated with out-of-registry structures and thus an increased capacity for strand rearrangement to form in-registry-ladders at this temperature.

To examine catalyst influence on the rate of molecular ladder formation, either TEA or scandium triflate[35] was added to 4M×4F reaction mixtures immediately following maleimide-deprotection at 140°C, and then cooled to either 60°C or 100°C. Aliquots of hybridization solutions spiked with polystyrene were collected and the single strand maleimide conversions were determined by GPC (Figure 4.10d and e). The addition of TEA did not significantly affect the rate of ladder formation at 60°C, suggesting that the signal enhancement in mass spectra of molecular ladder species in 3M×3F reaction mixtures observed above was a result of increased analyte ionization in the presence of TEA. This was confirmed by ESI-MS analysis of 3M×3F reactions immediately before and after addition of a catalytic amount of TEA to the solutions, yielding mass spectra with increased peak intensities attributable to molecular ladder species. Interestingly, evidenced by a discoloration of the reaction solution and a conversion plateau around 24 hours (Figure 4.10e), the 4M×4F reaction solutions at 100°C incorporating TEA as a catalyst experienced a deleterious side reaction, prompting the exclusion of TEA as a hybridization reaction catalyst. Nevertheless, TEA was added at room temperature prior to characterization by mass spectrometry to improve analyte ionization. At both 60°C and 100°C, the addition of scandium triflate increased the rate of single strand conversion, particularly within the first six hours.

4.4.5. Degree of Alignment

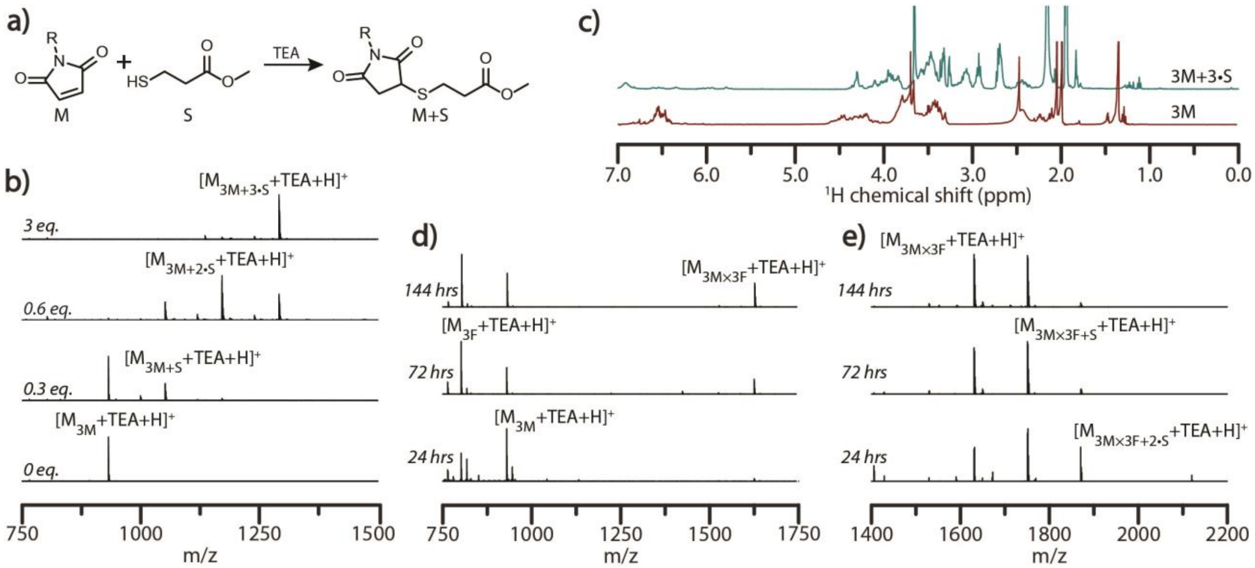


Figure 4.12 Molecular ladder registry. (a) Reaction scheme for the base catalyzed thiol-Michael addition between a maleimide and methyl 3-mercaptopropionate (S). (b) ESI mass spectra of thiol-Michael addition reaction mixtures between the peptoid 3M, bearing three maleimide pendant groups, and reacted for 30 minutes with various equivalents of S to maleimide groups. (c) ^{1}H-NMR spectra of a solution of 3M strands before (bottom, red) and

after (top, blue) reaction with 3 equivalents of S to maleimide groups. ESI mass spectra of 3M×3F hybridization mixtures (d) after reaction at 60°C for various periods as shown, and (e) after reaction at 60°C for various periods as shown, then subjected to a thiol-Michael addition with a threefold excess of S. Expected exact masses: $[M_{3F}+TEA+H]^+$ = 802.4; $[M_{3M}+TEA+H]^+$ = 931.4; $[M_{3M\times3F}+TEA+H]^+$ = 1631.7; $[M_{3M+S}+TEA+H]^+$ = 1051.4; $[M_{3M+2\bullet S}+TEA+H]^+$ = 1171.3; $[M_{3M+3\bullet S}+TEA+H]^+$ = 1291.3; $[M_{3M\times3F+S}+TEA+H]^+$ = 1751.7; $[M_{3M\times3F+2\bullet S}+TEA+H]^+$ = 1871.7;. Exact masses found: $[M_{3F}+TEA+H]^+$ = 802.4; $[M_{3M}+TEA+H]^+$ = 931.4; $[M_{3M\times3F}+TEA+H]^+$ = 1631.7; $[M_{3M+S}+TEA+H]^+$ = 1051.5; $[M_{3M+2\bullet S}+TEA+H]^+$ = 1171.5; $[M_{3M+3\bullet S}+TEA+H]^+$ = 1291.5; $[M_{3M\times3F+S}+TEA+H]^+$ = 1751.8; $[M_{3M\times3F+2\bullet S}+TEA+H]^+$ = 1871.8.

In imine- and boronate ester-based, self-assembled molecular ladder systems examined previously, the number of rungs (i.e., inter-strand bonds) formed between complementary oligomeric sequences was readily determined by mass spectrometry, where each imine or boronate ester bond formed yielded a mass change of -18 or -36, respectively, owing to the loss of water generated upon amine/aldehyde or boronic acid/diol condensation[27,36]. As the Diels-Alder cycloaddition reaction does not liberate a small molecule, in- and out-of-registry constructs are indistinguishable by conventional mass spectrometry. To afford a registry-dependent mass change, we employed a base-catalyzed thiol-Michael addition reaction between any un-reacted maleimide pendant groups and methyl 3-mercaptopropionate (S) as a reactive, low-molecular weight thiol (Figure 4.12a). Yielding a mass change of +120 (i.e., the exact mass of S) for each available maleimide pendant group that arises from out-of-registry ladder species, this approach enables the number of Diels-Alder adduct rungs to be readily determined by mass spectrometry. The efficiency of the thiol addition to strands bearing multiple reactive sites was initially examined by adding various amounts of S and a catalytic amount of TEA to solutions of 3M in anhydrous anisole. Mass spectra of the solutions demonstrated that, whereas incompletely reacted oligomers could be detected by ESI MS when the thiol was the limiting reagent, the efficient and near-complete maleimide consumption was observed when three equivalents of S to maleimide groups was employed to yield the fully-substituted strand 3M+3•S (Figure 4.12b). This fully-substituted strand was subjected to evaporation under reduced pressure to remove the solvent, excess S, and TEA, then redissolved in deuterated acetonitrile and analyzed by ^{1}H-NMR (Figure 4.12c) and HPLC (Figure 4.13). These revealed a 93% maleimide conversion, evident by a decrease in the ^{1}H-NMR maleimide peak around 6.5 ppm and increase in peaks attributable to the conjugated thiol between 2.6 and 3.2 ppm. Whereas, the deprotected peptoid 3M afforded a peak with a retention time of 11.4 minutes in reverse phase HPLC, the progressive addition of thiol diminished the prevalence of the 3M peak and yielded an emergent peak at an increased retention time of 14.4 minutes, attributable to the generated 3M+3•S.

.

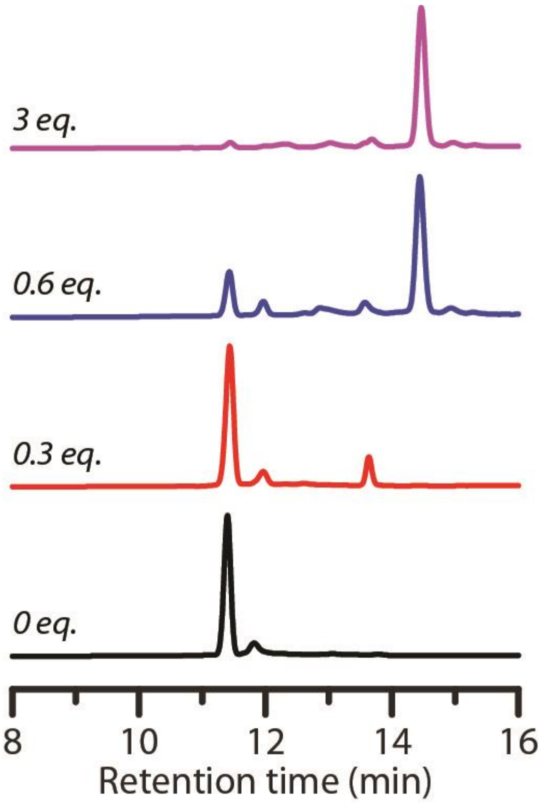


Figure 4.13 HPLC traces generated from thiol-Michael addition reaction mixtures between the peptoid, 3M, and various equivalents of S to maleimide groups.

Hybridization solutions containing 3M and 3F sequences at 60°C and 100°C were quenched at various times by rapid cooling to room temperature and subsequently reacted with three equivalents of the thiol S to maleimide groups in the presence of TEA to ensure quantitative conjugation with residual maleimides. Aliquots of the mixture before and after addition of S were characterized by ESI-MS to assess the ladder formation and determine the extent of strand alignment (Figure 4.12d and 4.12e and Figure 4.14). Mass spectrum peaks attributable to dimeric molecular ladder species progressively emerged with increased reaction times at both temperatures examined, signifying further conversion over time and supporting the previous GPC analysis. Upon addition of S, ESI-MS revealed that hybridization of 3M and 3F at both 60°C and 100°C does not yield complete registry of molecular ladders within 24 hours, but rather a combination of ladder species bearing 1, 2, or 3 rungs (Figure 4.12e and Figure 4.14). At extended reaction times, the peak attributable to the in-registry, 3-rung ladder does increases relative to the other ladder species for both reaction conditions, indicating rearrangement of the constituent precursor strands towards the target, in-registry ladder product, albeit at an enhanced rate at 100°C..

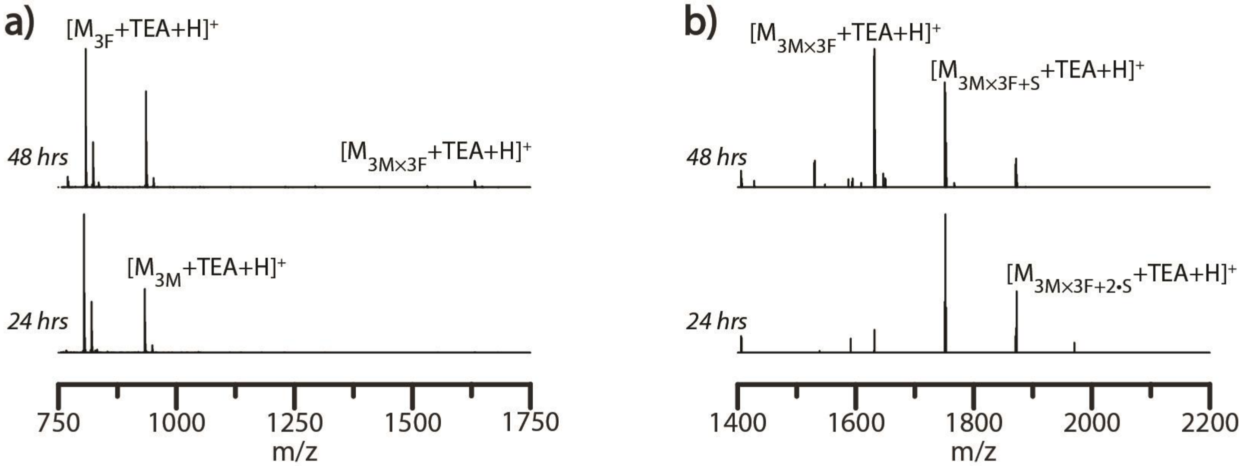


Figure 4.14 ESI mass spectra of 3M×3F hybridization mixtures (a) after reaction at 100°C for 24 and 48 hours, and (b) after reaction at 100°C for 24 and 48 hours, then subjected to a thiol-Michael addition with a threefold excess of S. Expected exact masses: $[M_{3F}+TEA+H]^+ = 802.4$; $[M_{3M}+TEA+H]^+ = 931.4$; $[M_{3M\times 3F}+TEA+H]^+ = 1631.7$; $[M_{3M+S}+TEA+H]^+ = 1051.4$; $[M_{3M+2\cdot S}+TEA+H]^+ = 1171.3$; $[M_{3M+3\cdot S}+TEA+H]^+ = 1291.3$; $[M_{3M\times 3F+S}+TEA+H]^+ = 1751.7$; $[M_{3M\times 3F+2\cdot S}+TEA+H]^+ = 1871.7$;. Exact masses found: $[M_{3F}+TEA+H]^+ = 802.4$; $[M_{3M}+TEA+H]^+ = 931.4$; $[M_{3M\times 3F}+TEA+H]^+ = 1631.7$; $[M_{3M+S}+TEA+H]^+ = 1051.5$;

$[M_{3M+2\cdot S}+TEA+H]^+$ = 1171.5; $[M_{3M+3\cdot S}+TEA+H]^+$ = 1291.5; $[M_{3M\times 3F+S}+TEA+H]^+$ = 1751.8; $[M_{3M\times 3F+2\cdot S}+TEA+H]^+$ = 1871.8.

4.4.6. Sequence-selective Ladder Assembly

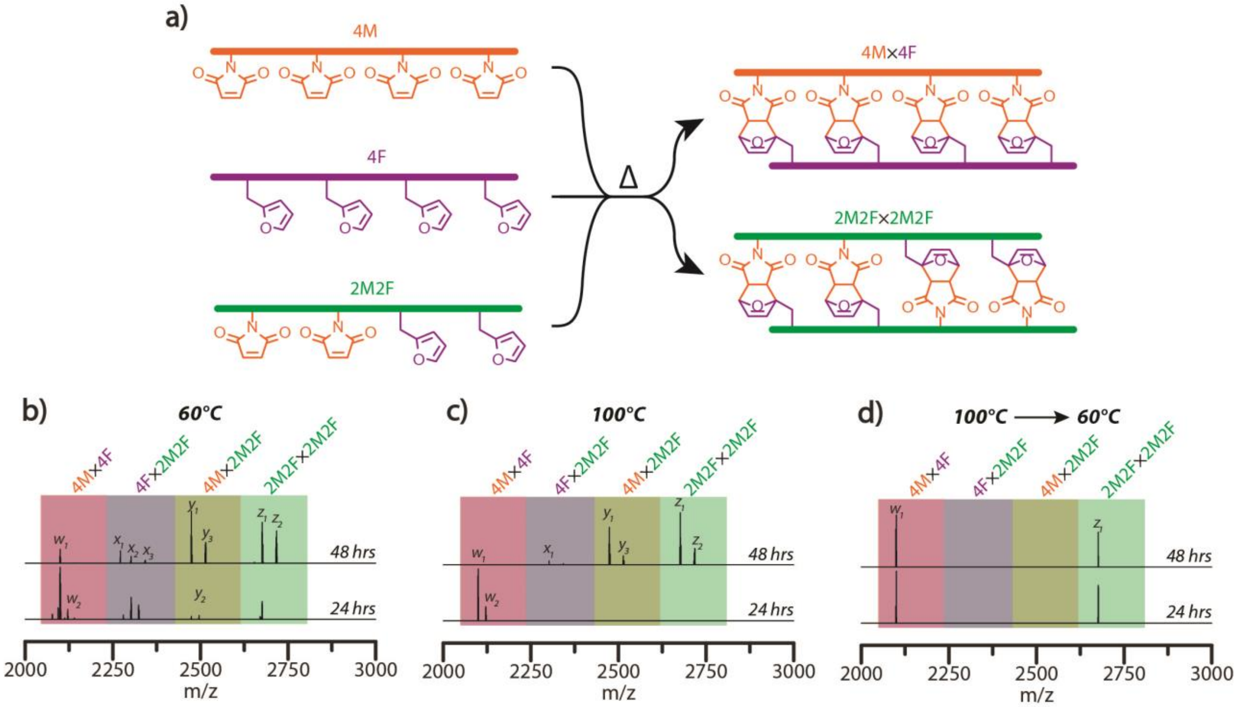


Figure 4.15 Dynamic covalent assembly of sequence-defined molecular ladders (a) Idealized schematic diagram of the concurrent self-assembly of 4PM, 4F, and 2PM2F precursor strands to afford the four-rung molecular ladders 4M×4F and 2M2F×2M2F. ESI mass spectra of molecular ladder reaction mixture aliquots removed from hybridization solutions heated at 140°C for 30 minutes then cooled and maintained at (b) 60°C, (c) 100°C, or (d) 100°C for 24 hours, allowed to slowly cool, then maintained at 60°C. Expected exact masses: w_1 = $[M_{4M\times 4F}+Na]^+$ = 2099.8; w_2 = $[M_{4M\times 4F}+2Na\text{-}H]^+$ = 2121.8; x_1 = $[M_{4F\times 2M2F}+H]^+$ = 2280.0; x_2 = $[M_{4F\times 2M2F}+Na]^+$ = 2303.0; x_3 = $[M_{4F\times 2M2F}+2Na\text{-}H]^+$ = 2324.0; y_1 = $[M_{4M\times 2M2F}+Na]^+$ = 2474.0; y_2 = $[M_{4M\times 2M2F}+2Na\text{-}H]^+$ = 2496.0; y_3 = $[M_{4M\times 2M2F}+MeCN+Na]^+$ = 2515.1; z_1 = $[M_{2M2F\times 2M2F}+Na]^+$ = 2676.2; z_2 = $[M_{2M2F\times 2M2F}+MeCN+Na]^+$ = 2717.2. Exact masses found: w_1 = $[M_{4M\times 4F}+Na]^+$ = 2099.8; w_2 = $[M_{4M\times 4F}+2Na\text{-}H]^+$ = 2121.8; x_1 = $[M_{4F\times 2M2F}+H]^+$ = 2280.0; x_2 = $[M_{4F\times 2M2F}+Na]^+$ = 2303.0; x_3 = $[M_{4F\times 2M2F}+2Na\text{-}H]^+$ = 2324.0; y_1 = $[M_{4M\times 2M2F}+Na]^+$ = 2474.0; y_2 = $[M_{4M\times 2M2F}+2Na\text{-}H]^+$ = 2496.0; y_3 = $[M_{4M\times 2M2F}+MeCN+Na]^+$ = 2515.0; z_1 = $[M_{2M2F\times 2M2F}+Na]^+$ = 2676.2; z_2 = $[M_{2M2F\times 2M2F}+MeCN+Na]^+$ = 2717.1.

Having demonstrated a capacity for effective error correction and rearrangement of the inter-strand Diels-Alder adduct-based rungs at 100°C, we examined a temperature-mediated approach, analogous to the thermal cycling employed to effect sequence-selective nucleic acid hybridization, to achieve the hybridization selectivity of maleimide- and furan-bearing precursor strands to yield multiple target molecular ladders in a single pot. Here, several unique, sequence-defined oligomers in a single pot reaction mixture were deprotected and dissociated at high temperature, then allowed to react and anneal under lower temperature conditions. One of the sequences employed was a peptoid bearing two furan-protected maleimide and two furan pendant groups (2PM2F), mass-labeled with additional 2-ethoxyethoxyethylamine and 2-methoxyethylamine residues at its N- and C-terminal ends, respectively; upon in situ deprotection, this oligomeric sequence should undergo self-hybridization to afford the molecular ladder 2M2F×2M2F. Concurrent hybridization selectivity for complementary peptoid pairs was examined by allowing the tetrafunctional, mass-labeled sequences 2PM2F, 4PM, and 4F, in a 2:1:1 molar ratio, to react simultaneously in a single pot reaction mixture to yield two distinct target molecular ladder species (Figure 4.15a). Initially, reaction mixtures were heated at 140°C for 30 minutes, then cooled and maintained at either 60°C or 100°C while aliquots were collected, thermally quenched, and characterized by ESI-MS (Figure 4.15b and 4.15c, respectively). At a reaction temperature of 60°C, all possible dimeric combinations between and complementary and non-complementary sequences (i.e., 4M×4F, 4F×2M2F, 4M×2M2F, and 2M2F×2M2F) were observed (Figure 4.15b), even after prolonged reaction times (Figure 4.16), indicating that oligomer hybridization was unselective for sequence under these reaction conditions. At a reaction temperature of 100°C, ESI-MS showed the generation of only the 4M×4F species after 24 hours; however, all other possible combinations were present after 48 hours (Figure 4.15c), again showing unselective oligomer hybridization. Importantly, as these reaction conditions involve near-step temperature changes that provide insufficient time for connectivity rearrangement, misconfigured ladder species between non-complementary precursor strands remain present in the reaction mixtures. To better approximate the slow cooling step of a thermal cycle employed for nucleic acid hybridization, a modified temperature ramp profile was employed where the reaction mixture temperature was initially reduced from 140°C to 100°C and held at that temperature for 24 hours, then slowly reduced further to 60°C and held there for up to 48 hours. As shown in Figure 4d, ESI-MS revealed that this temperature profile exclusively yielded the two target molecular ladder assemblies (i.e., 4M×4F and 2M2F×2M2F) after 24 hours at 60°C, and that this distribution of reaction products was maintained after a further 24 hours at 60°C, confirming the capacity of this Diels-Alder adduct-based system for temperature-mediated, sequence-selective hybridization.

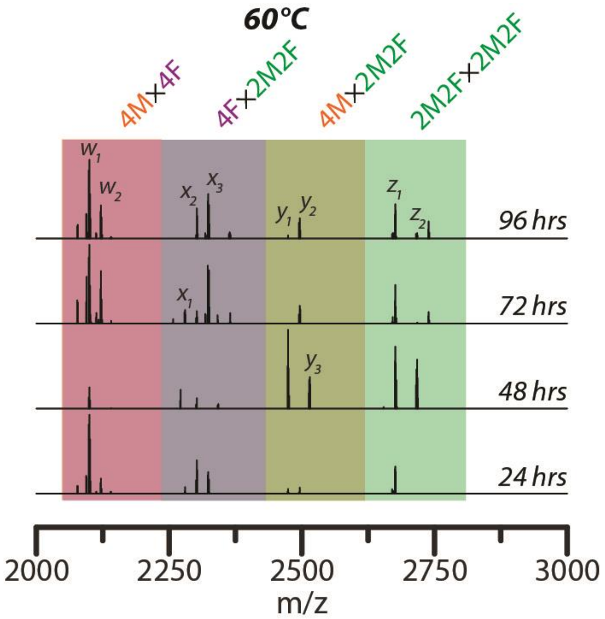


Figure 4.16 Dynamic covalent assembly of sequence-defined molecular ladders. ESI mass spectra of molecular ladder reaction mixture aliquots removed from hybridization solutions maintained at 60°C. Each time point shown indicates the period at 60°C. Expected exact masses: $w_1 = [M_{4M\times 4F}+Na]^+ = 2099.8$; $w_2 = [M_{4M\times 4F}+2Na\text{-}H]^+ = 2121.8$; $x_1 = [M_{4F\times 2M2F}+H]^+ = 2280.0$; $x_2 = [M_{4F\times 2M2F}+Na]^+ = 2303.0$; $x_3 = [M_{4F\times 2M2F}+2Na\text{-}H]^+ = 2324.0$; $y_1 = [M_{4M\times 2M2F}+Na]^+ = 2474.0$; $y_2 = [M_{4M\times 2M2F}+2Na\text{-}H]^+ = 2496.0$; $y_3 = [M_{4M\times 2M2F}+MeCN+Na]^+ = 2515.1$; $z_1 = [M_{2M2F\times 2M2F}+Na]^+ = 2676.2$; $z_2 = [M_{2M2F\times 2M2F}+MeCN+Na]^+ = 2717.2$. Exact masses found: $w_1 = [M_{4M\times 4F}+Na]^+ = 2099.8$; $w_2 = [M_{4M\times 4F}+2Na\text{-}H]^+ = 2121.8$; $x_1 = [M_{4F\times 2M2F}+H]^+ = 2280.0$; $x_2 = [M_{4F\times 2M2F}+Na]^+ = 2303.0$; $x_3 = [M_{4F\times 2M2F}+2Na\text{-}H]^+ = 2324.0$; $y_1 = [M_{4M\times 2M2F}+Na]^+ = 2474.0$; $y_2 = [M_{4M\times 2M2F}+2Na\text{-}H]^+ = 2496.0$; $y_3 = [M_{4M\times 2M2F}+MeCN+Na]^+ = 2515.0$; $z_1 = [M_{2M2F\times 2M2F}+Na]^+ = 2676.2$; $z_2 = [M_{2M2F\times 2M2F}+MeCN+Na]^+ = 2717.1$.

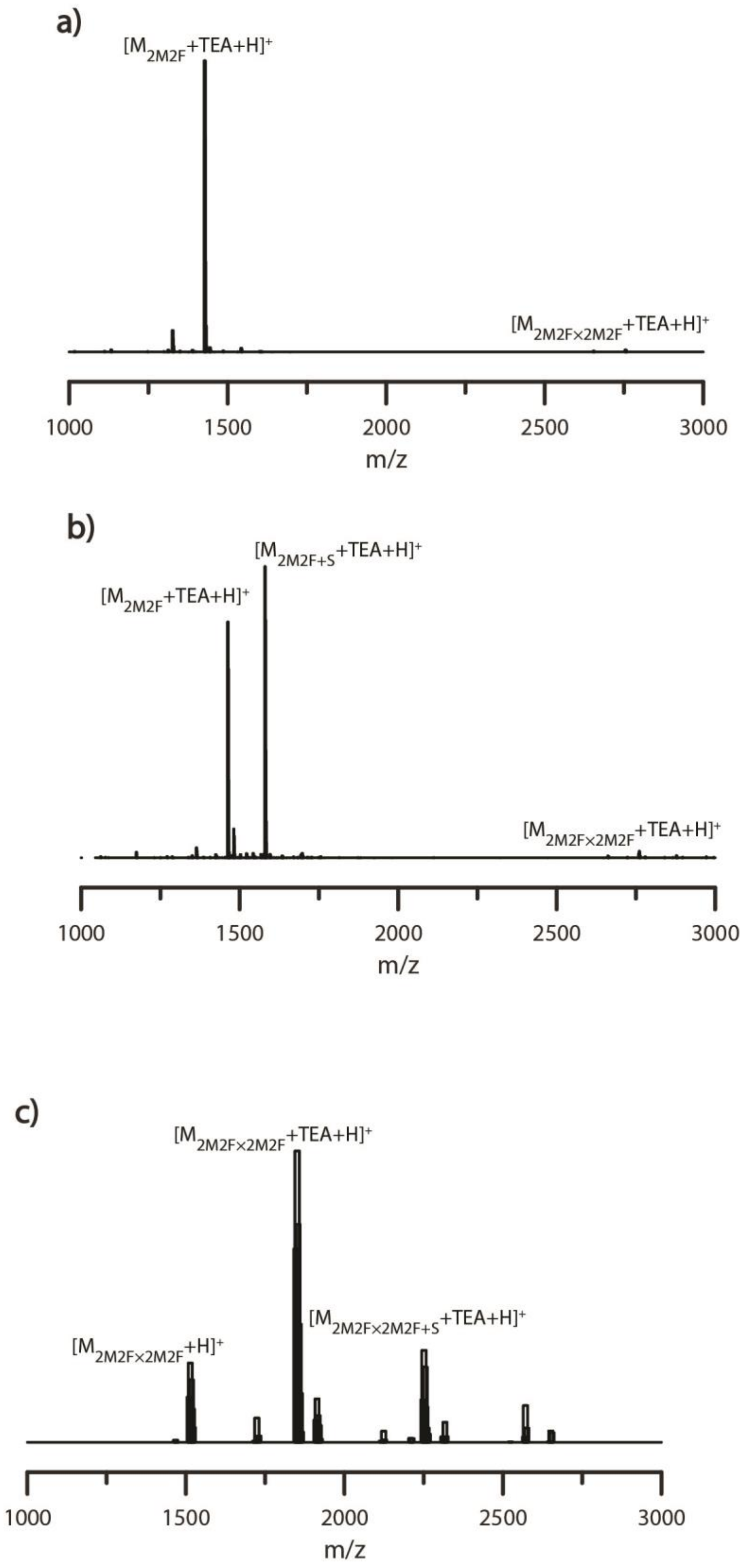


Figure 4.17 (a) ESI mass spectra were generated of 2M2F×2M2F hybridization reaction mixtures at 60°C. The 3M×3F hybridization reaction mixtures were subjected to a thiol-Michael addition with S and subsequently characterized by additional ESI mass spectra (b)

and (c). Expected exact masses: $[M_{2M2F}+TEA+H]^+ = 1428.7$; $[M_{2M2F+S}+TEA+H]^+ = 1548.8$; $[M_{2M2F\times 2M2F}+H]^+ = 2654.2$; $[M_{2M2F\times 2M2F}+TEA+H]^+ = 2755.3$; $[M_{2M2F\times 2M2F+S}+TEA+H]^+ = 2875.3$. Exact masses found: $[M_{2M2F}+TEA+H]^+ = 1428.7$; $[M_{2M2F+S}+TEA+H]^+ = 1548.7$; $[M_{2M2F\times 2M2F}+H]^+ = 2654.2$; $[M_{2M2F\times 2M2F}+TEA+H]^+ = 2755.3$; $[M_{2M2F\times 2M2F+S}+TEA+H]^+ = 2875.3$.

4.5 Conclusion

In summary, we have demonstrated the self-assembly of well-characterized ladder species employing the Diels-Alder dynamic covalent interaction. By utilizing a furan-protected maleimide residue, sequence-defined peptoids can be synthesized simultaneously bearing both furan and maleimide species while precluding premature hybridization. While fundamental, the sequence-selective formation of two distinct ladder species using temperature-mediated reversible covalent interactions establishes a self-assembly mechanism requiring no internal stimuli but capable of alleviating or eliminating kinetic trapping. Because of this work, we anticipate further interest and development of the Diels-Alder cycloaddition reaction for the self-assembly of robust, covalent molecular structures.

4.6 References

1. Jin, Y., Yu, C., Denman, R. J. & Zhang, W. Recent advances in dynamic covalent chemistry. *Chem. Soc. Rev.* **42,** 6634–6654 (2013).
2. Furgal, J. C., Dunn, M., Wei, T. & Scott, T. F. in *Dynamic Covalent Chemistry: Principles, Reactions, and Applications* 389–434 (2017).
3. Rowan, S. J., Cantrill, S. J., Cousins, G. R. L., Sanders, J. K. M. & Stoddart, J. F. *Dynamic covalent chemistry. Angew. Chem. Int. Ed.* **41,** (2002).
4. Ockwig, N. W., Co, A. P., Keeffe, M. O., Matzger, A. J. & Yaghi, O. M. Porous , Crystalline , Covalent Organic Frameworks. **310,** 1166–1171 (2005).
5. Feng, X., Ding, X. & Jiang, D. Covalent organic frameworks. *Chem. Soc. Rev.* **41,** 6010–6022 (2012).
6. Ma, X. & Scott, T. F. Approaches and challenges in the synthesis of three-dimensional covalent-organic frameworks. *Commun. Chem.* **1,** (2018).
7. Zhang, G. & Mastalerz, M. Organic cage compounds-from shape-persistency to function. *Chem. Soc. Rev.* **43,** 1934–1947 (2014).
8. Jin, Y., Wang, Q., Taynton, P. & Zhang, W. Dynamic covalent chemistry approaches toward macrocycles, molecular cages, and polymers. *Acc. Chem. Res.* **47,** 1575–1586 (2014).

9. Granzhan, A., Schouwey, C., Riis-Johannessen, T., Scopelliti, R. & Severin, K. Connection of metallamacrocycles via dynamic covalent chemistry: A versatile method for the synthesis of molecular cages. *J. Am. Chem. Soc.* **133,** 7106–7115 (2011).
10. Hartley, C. S., Elliott, E. L. & Moore, J. S. Covalent assembly of molecular ladders. *J. Am. Chem. Soc.* **129,** 4512–4513 (2007).
11. Wei, T., Jung, J. H. & Scott, T. F. Dynamic covalent assembly of peptoid-based ladder oligomers by vernier templating. *J. Am. Chem. Soc.* **137,** 16196–16202 (2015).
12. Iida, H. *et al.* Double-Stranded Helical Oligomers Covalently Bridged by Rotary Cyclic Boronate Esters. *Chem. Asian J.* **12,** 927–935 (2017).
13. Elliott, E. L., Hartley, C. S. & Moore, J. S. Covalent ladder formation becomes kinetically trapped beyond four rungs. *Chem. Commun.* **47,** 5028–5030 (2011).
14. Zhu, G. *et al.* Formation Mechanisms and Defect Engineering of Imine-Based Porous Organic Cages. *Chem. Mater*. **30,** 262–272 (2018).
15. Baram, J., Weissman, H. & Rybtchinski, B. Supramolecular polymer transformation: A kinetic study. *J. Phys. Chem. B* **118,** 12068–12073 (2014).
16. Wang, Q. *et al.* Dynamic covalent synthesis of aryleneethynylene cages through alkyne metathesis: Dimer, tetramer, or interlocked complex? *Chem. Sci.* **7,** 3370–3376 (2016).
17. Gong, B. Molecular duplexes with encoded sequences and stabilities. *Acc. Chem. Res.* **45,** 2077–2087 (2012).
18. Chakrabarty, R., Mukherjee, P. S. & Stang, P. J. Supramolecular coordination: Self-assembly of finite two- and three-dimensional ensembles. *Chem. Rev.* **111,** 6810–6918 (2011).
19. Stupp, S. I. *et al.* Supramolecular Materials: Self-Organized Nanostructures. *Science* **276,** 384–389 (1997).
20. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99,** 237–247 (1982).
21. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).
22. Tørring, T., Voigt, N. V., Nangreave, J., Yan, H. & Gothelf, K. V. DNA origami: A quantum leap for self-assembly of complex structures. *Chem. Soc. Rev.* **40,** 5636–5646 (2011).
23. Leguizamon, S. C. & Scott, T. F. Sequence-selective dynamic covalent assembly of information-bearing oligomers. *Nat. Commun.* **11,** 784 (2020).
24. Leguizamon, S. C., Dunn, M. & Scott, T. F. Sequence-directed Dynamic Covalent Assembly of Base-4-encoded Oligomers. *Chem. Commun.* **56,** 7817–7820 (2020).
25. Adzima, B. J., Aguirre, H. A., Kloxin, C. J., Scott, T. F. & Bowman, C. N. Rheological and chemical analysis of reverse gelation in a covalently cross-linked diels-alder polymer network. *Macromolecules* **41,** 9112–9117 (2008).
26. Liu, Y. L. & Chuo, T. W. Self-healing polymers based on thermally reversible Diels-Alder chemistry. *Polym. Chem.* **4,** 2194–2205 (2013).
27. Wei, T., Furgal, J. C., Jung, J. H. & Scott, T. F. Long, self-assembled molecular ladders by cooperative dynamic covalent reactions. *Polym. Chem.* **8,** 520–527 (2017).
28. Reuther, J. F. *et al.* Dynamic covalent chemistry enables formation of antimicrobial peptide quaternary assemblies in a completely abiotic manner. *Nat. Chem.* **10,** 45–50 (2018).
29. Zuckermann, R. N., Kerr, J. M., Kent, S. B. H. & Moos, W. H. Efficient Method for the Preparation of Peptoids [Oligo(N-substituted glycines)] by submonomer solid-phase synthesis. *J. Am. Chem. Soc.* **114,** 10646–10647 (1992).
30. Froidevaux, V. *et al.* Study of the diels-alder and retro-diels-alder reaction between furan derivatives and maleimide for the creation of new materials. *RSC Adv.* **5,** 37742–37754 (2015).

31. Discekici, E. H. *et al.* Endo and Exo Diels-Alder Adducts: Temperature-Tunable Building Blocks for Selective Chemical Functionalization. *J. Am. Chem. Soc.* **140,** 5009–5013 (2018).
32. Mane, S. R. *et al.* A unique polymeric nano-carrier for anti-tuberculosis therapy. *J. Mater. Chem.* **22,** 19639–19642 (2012).
33. Okarnura, H., Iwagawa, T. & Nakatani, M. A Base Catalyzed Diels-Alder Reaction of 3-Hydroxy-2-Pyrone. *Tetrahedron Lett.* **36,** 163 (1995).
34. Ho, C. *et al.* Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin Biochem Rev.* **24,** 3–12 (2003).
35. Fukuzumi, S., Yuasa, J., Miyagawa, T. & Suenobu, T. Mechanism of scandium ion catalyzed diels-alder reaction of anthracenes with Methyl Vinyl Ketone. *J. Phys. Chem. A* **109,** 3174–3181 (2005).
36. Dunn, M. F., Wei, T., Scott, T. F. & Zuckermann, R. N. Aqueous dynamic covalent assembly of molecular ladders and grids bearing boronate ester rungs. *Polym. Chem.* **10,** 2337–2343 (2019).

Chapter 5 Rapid Gel-card Agglutination Assays for Serological Analysis Following SARS-CoV-2 Infection in Humans

5.1 Original Publication Information

Alves. D.[*], Curvello, R.[*], Henderson, E.[*], Kesarwani, V.[*], Walker, J.A.[*], **Leguizamon, S.C.[*],** McLiesh, H.[*], Raghuwanshi, V.[*] , Samadian, H.[*] ,Wood, E.M., McQuilten, Z.K., Graham, M., Wieringa, M., Korman, T.M., Scott, T.F., Banaszak-Holl, M.M., Garnier, G., Corrie, S.R. Rapid gel-card agglutination assays for serological analysis following SARS-CoV-2 infection in humans. **Article ASAP** DOI: 10.1021/acssensors.0c01050

**Authors contributed equally*

Modifications have been made to the original document in order to adapt the content to the proper format.

Given the rapid nature necessitated for the development of the SARS-Cov-2 assay, a large team was employed. Although my primary contribution to this work was in the recognition of specific epitopes, synthesis of peptide sequences, and development of bio-conjugation chemistries and techniques, I also assisted in the conceptualization and analysis of the entire project.

5.2 Abstract

The detection of IgG/IgM response specific to SARS-Cov-2 in human blood samples by high-throughput and rapid serology assays is urgently required to improve our understanding of the effects of COVID-19 across the world. Short-term applications include rapid case identification and contact tracing to limit viral spread, while population screening to determine the extent of viral

infection across populations is a longer-term need. Assays developed to address these needs should match the ASSURED criteria. Agglutination tests based on commonly employed blood typing methods are identified as a viable option. These blood typing tests are employed in hospitals worldwide, are high-throughput, fast (10-30 minutes) and automated in most cases. Here, the application of agglutination assays to SARS-CoV-2 serology testing is described by combining column agglutination testing with peptide-antibody bioconjugates which facilitate red cell crosslinking only in the presence of plasma containing antibodies against SARS-CoV-2. This simple, rapid, and easily scalable approach has immediate application in SARS-CoV-2 serological testing, and is a useful platform for assay development beyond the COVID-19 pandemic.

5.3 Introduction

Coronavirus Disease (COVID)-19 due to SARS-CoV-2 (Severe Acute Respiratory Syndrome - coronavirus – 2) has caused a worldwide viral pandemic, with >642,000 deaths and >15,900,000 cases reported internationally,[1] with Australia reporting 13,950 cases and 145 deaths[2], as of 26th July 2020. Large-scale efforts are underway to develop vaccines and anti-viral therapy, and epidemiological methods of social distancing and quarantine are being used to reduce the spread of infection, and the rapid development and deployment of diagnostic tests is of key importance.[3, 4] PCR (polymerase chain reaction) tests are already widely available to confirm SARS-CoV-2 infection from respiratory samples, [3] and direct antigen assays are also emerging to detect current infection.[5, 6] However, there is currently a lack of high-throughput, lab-based blood screening tests that detect the antibody response to viral infection. These serology tests are required for population screening, case identification, contact tracing, and potentially to confirm vaccine efficacy during clinical trials and vaccine distribution.[7-9] While the full picture of the immune response to SARS-CoV-2 is still emerging, recent reports suggests IgG and IgM antibodies are produced either sequentially or simultaneously, with titres reaching a plateau 6 days after seroconversion ,[10] and that SARS-CoV-2 antigens elicit highly specific antibody responses not present in naïve individuals, including those previously infected with other coronaviruses.[11, 12] A key unanswered question is whether or not SARS-CoV-2 infection yields long-lived antibody responses for long-term protective immunity; mass serology testing is required to comprehensively address this question.

Several approaches for serology testing are already being distributed around the world. Point-of-care paper-based tests for antibodies are under evaluation and available in some countries, but they cannot be used for high-throughput screening (15-30 minutes/sample), and specificity/sensitivity is not expected to meet the standard of laboratory-based tests. The current gold standard for serology methods is laboratory-based indirect Enzyme-Linked Immunosorbent Assay (ELISA), in which antibodies from patient serum are captured onto a protein-coated microwell plate followed by enzymatic detection using an anti-Ig secondary antibody.[3, 4] These assays can be performed manually or using automated systems,[13] however they are still multi-step processes requiring multiple antibodies and reagents.

Alternative approaches should meet the ASSURED (Affordable, Sensitive, Specific, User-friendly, Rapid/Robust, Equipment-free, Deliverable to end-users) criteria,[14] and be rapidly scalable and customisable. Blood-typing and antibody screening is performed in hospital laboratories all over the world, using the robust column agglutination test (CAT) technology. Detection of antibodies in patient plasma or serum involves pipetting a mixture of reagent red blood cells (RRBCs) and antibody-containing serum/plasma onto a gel card containing separation media, incubating the card for 5-15 minutes, and using a centrifuge to separate agglutinated cells from free cells (Figure 1A), resulting in strong red lines on top of the gel column in the case of a "positive" test. A wide range of RRBCs expressing different surface antigens are available for assay development, along with corresponding antibodies of varying affinity and avidity. During the early years of the HIV (human immunodeficiency virus) epidemic, Kemp et al developed a variant of the blood typing assays to detect HIV-induced antibodies present in the blood of previously infected patients.[15] Anti-glycophorin antibodies or Fabs[16] were bioconjugated to viral peptide gp41, so that in the presence of a human blood sample, the antibodies would bind any RBCs, while the peptide would bind to anti-HIV-IgG, producing agglutination reactions only in the presence of blood plasma collected from HIV-positive people. The degree of agglutination was then read on a microscope slide by a trained reader. However, to automate the process and broadly deploy this approach, an alternative to microscopic examination of agglutination reactions is required.

In this study, we developed a serology test to detect SARS-CoV-2 antibodies from human plasma using gel card agglutination tests. The CAT technology was selected to maximize rapid testing and comprehensive serology mapping, for two reasons. First, CAT is currently available in the blood/analytical laboratory of all major hospitals throughout the world, with equipment and trained personnel already in place. The CAT technology is already automated and high throughput (>100 tests/hour). The second is that many companies are currently manufacturing the gel cards widely used for blood typing analysis. Production of SARS-CoV-2 gel card diagnostics only requires the substitution of the current RRBCs with bioconjugated cells, using the current processing and technology. We have found that by producing bioconjugates of anti-D-IgG and peptides from SARS-CoV-2 spike protein, and immobilising these to RRBCs, we observe selective agglutination assays in gel-cards in the presence of plasma collected from patients recently infected with SARS-CoV-2 in comparison to healthy plasma and negative controls.

5.4 Experimental

5.4.1. General Experimental Procedure

Chemicals were purchased from Sigma-Aldrich unless otherwise identified. Seeblue plus2 protein standard (ThermoFisher Scientific), GelCode blue stain reagent (ThermoFisher Scientific), goat anti-human IgG (H+L) cross-adsorbed secondary antibody; HRP-labelled, (ThermoFisher Scientific), gel apparatus (Life Technology), PBS (Gibco, Invitrogen), syringe filters (0.22 μm–0.45 μm, Pall Inc.), HiTrap Protein A column (GE Healthcare), anti-D-IgG FFMU ("for further manufacturing use"; CSL), NHS-(PEG)2-maleimide (Quanta Biodesign), zebapsin columns (ThermoFisher Scientific), Celpresol (CSL).

5.4.2. Peptide Synthesis

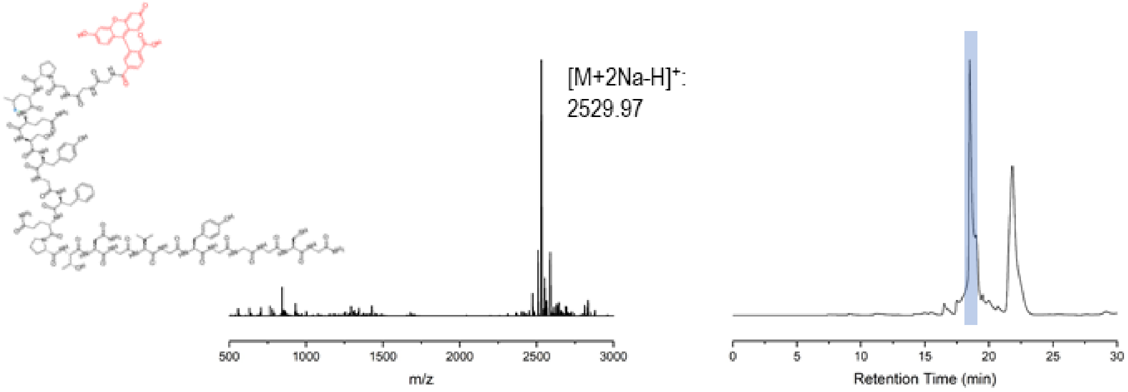
Peptides were synthesized via microwave-assisted solid phase synthesis using Rink amide resin (0.05 mmol scale, 100-200 mesh, 1% DVB, ChemPep Inc.) as the resin. Syntheses were performed in an automated microwave synthesizer (Liberty Blue, CEM Corporation, North Carolina). Resin was swelled at room temperature for 5min with N,N-dimethylformamide (DMF) before deprotection with 20% 4-methylpiperidine in DMF (v/v) for 30 s at 75 °C and 90 s at 90 °C. Subsequently, Fmoc-amino acids (0.25 mmol, ChemPep Inc.) were coupled with a 1:1:1:2 ratio of amino acid/1-hydroxybenzotriazole hydrate (HOBT, AK Scientific, Inc.)/HCTU (AK Scientific, Inc.)/N,N-diisopropylethylamine (DIPEA) in 4 mL DMF at 70 °C for 5 min prior to deprotection with 20% 4-methylpiperidine in DMF (v/v) for 5 min at 75 °C. Fluorescein was included in the N-terminal by coupling of 5,6-carboxyfluoroscein (0.25 mmol) to the final amino acid in the sequence, after Fmoc-deprotection, with a 1:1:1:2 ratio of amino acid/HOBT/HCTU/DIPEA in 4 mL DMF at 70 °C. The fluorescein coupling step was performed twice to ensure complete labeling of the synthesized sequences.

After solid-phase synthesis, the resultant dried Rink-amide resin was transferred to a 25 mL solid-phase peptide synthesis vessel (CG-1866, Chemglass) and treated with 10 mL of trifluoroacetic acid (TFA)/phenol/water/triisopropylsilane (88/5/5/2) cleavage cocktail for 2 hours while bubbling with nitrogen at room temperature. The TFA cleavage solution was collected by filtering through the fritted glass into a 25 mL round bottom flask. The remaining resin was further rinsed twice with 5 mL of fresh TFA cleavage cocktail to collect any residual peptoid. The cleavage solution was combined and precipitated into cold diethyl ether. Upon centrifugation, the precipitate was collected and reconstituted in HPLC grade water/acetonitrile (MeCN) with 0.1% TFA. Peptides were purified by preparative HPLC using a linear gradient of MeCN and water: (1) 10% MeCN, 0.1–2.1 min; (2) 10–95% MeCN, 2.1–23.1.1 min; (3) 95% MeCN, 23.1–26.1 min. Purified peptides were lyophilized to yield off-white powder and characterized by ESI-MS.

Peptide 1

FAM-GGG-PLQSYGFQPTNGVGY-GGGCG

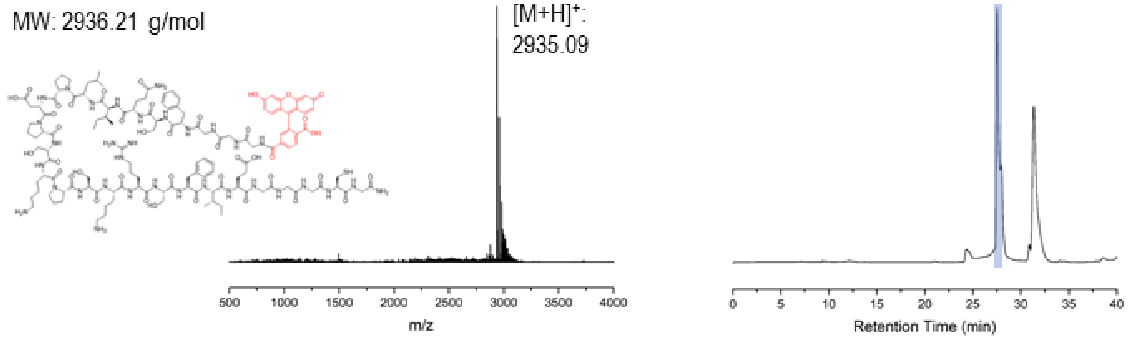
MW: 2487.60 g/mol



Peptide 2

FAM-GGG-FSQILPDPSKPSKRSFIE-GGGCG

MW: 2936.21 g/mol



Peptide 5

FAM-GGG-FGAGAALQIPFAMQMAYRFNGI-GGGCG

MW: 3234.64 g/mol

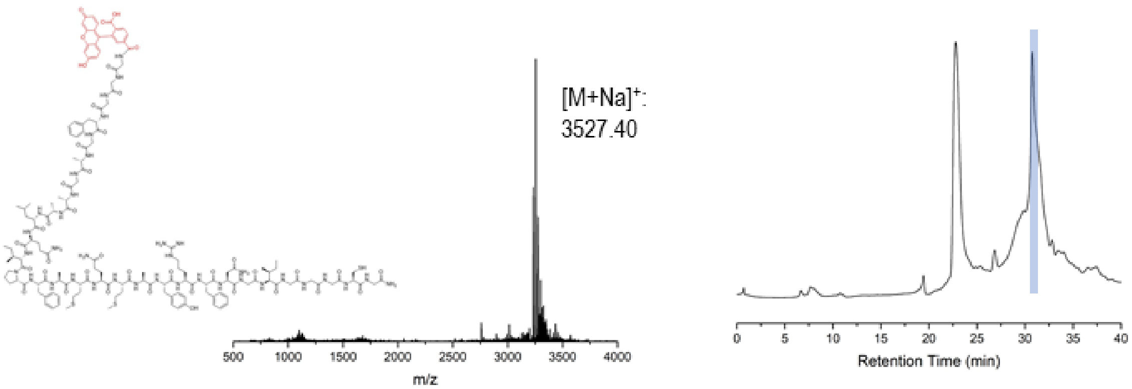


Figure 5.1 Peptide sequences, expected structures and LC/MS results.

5.4.3. Anti-D-IgG Purification and Bioconjugation Reaction

Anti-D-IgG was purified from FFMU product using Protein A affinity chromatography and buffer exchange. 100 mL of the FFMU was filtered through a 0.22 µm filter before being loaded onto purification column. A HiTrap Protein A column (GE Healthcare) was connected to AKTA Chromatography System (AKTA Start, GE Healthcare), equilibrated with 1X PBS buffer (pH 7.4). Filtered supernatant was loaded onto the column followed by washing with 1X PBS (3 column volume). Proteins were eluted using glycine elution buffer (1X PBS, 0.1M glycine, pH 2.8), neutralised by adding 150 µL 1M Tris pH 9 per 1 mL eluate, and then exchanged to 1x PBS buffer (pH 7.4) using a HiPrep 26/10 desalting column. The resulting purified anti-D-IgG was then concentrated and stored at 4°C for several weeks, over which time we observed no precipitation or reduction in absorbance at 280 nm as measured on a Nanodrop spectrometer.

5.4.4. Bioconjugate Reaction

Bioconjugation reactions were performed with ~3 mg/mL anti-D-IgG concentrations at 100 µL scale. Firstly, a 50-fold molar excess of NHS-$(PEG)_2$-maleimide was added to the antibody and incubated for 2 hours at 4°C. This mixture was then desalted using Zebaspin columns to remove the free PEG, before reaction with peptide. The thiolated peptide was added to the antibody solution in 15-20-fold molar excess, incubated at room temperature for 1 hr at room temperature, before another desalting step to remove free peptide. The resulting bioconjugates were loaded into NuPage 4-12%, Bis-Tris gels for SDS-PAGE analysis in accordance with the manufacturer's instructions. Gels were analysed in a UVP Biospectrum gel imager using Cy2 excitation/emission filters to confirm antibody labelling with the fluorescent peptide, prior to Coomassie staining (GelCode) and imaging under white light. Bioconjugates were stored as 1:1 glycerol:1XPBS stocks at -20°C and were stable for at least several weeks. Concentrations of bioconjugates were monitored using Nanodrop spectrometer.

5.4.5. Flow Cytometry

Flow assays were performed to confirm bioconjugate binding to RRBCs and to determine the concentration of biconjugate required to saturate the cells. Data was collected on a Beckman Coulter Cytoflex. A single gate identifying the RRBCs on a forward/side scatter plot was applied to the fluorescence histogram displaying FAM-positive cells. Data was exported to Flowjo LLC for off-line analysis, and plotted for display in GraphPad Prism.

5.4.6. Column Agglutination Tests

Stocks of bioconjugate-labelled RRBCs (R2R2 cells) were prepared fresh daily. 0.8% RRBCs in naïve human plasma solution were prepared and bioconjugate was added directly to achieve the desired ratio of bioconjugate:D-antigen on the cells. Following 20 minutes incubation at room temperature, the cells were pelleted and washed four times in Celpresol solution. Neutral or Coombs' reagent gel cards (*i.e.* containing anti-human-IgG) produced by Haemokinesis were employed as required for CAT assays. 50 µL of 0.8% bioconjugate-coated RRBCs (unless otherwise stated, 2:1 bioconugate: D-antigen ratio was used) were added to the gel-cards along with 25 µL plasma. The cards were incubated at 25°C for 10 minutes, then centrifuged for 11 minutes (Haemokinesis gel-card centrifuge), and results recorded digitally using the Haemokinesis gel-card reader. De-identified human plasma or serum samples were provided by Monash Pathology and the Australian Red Cross Lifeblood, obtained with written informed consent in accordance with the recommendations of Blood Service Human Research Ethics Committee (BSHREC) and the Monash University Human Research Ethics Committee (MUHREC).

5.4.7. Indirect IgG ELISA

A mixture of spike S1, spike S1+S2, spike-RBD and nucleocapsid proteins (mass ratio 1:1:1:0.5; Jomar Life Research) was coated in 96-well plates (Nunc flat bottom, Maxisorp; 0.22 μg/mL total protein concentration) and incubated overnight at 4 °C. Plates were blocked with 200 μL of blocking buffer (5% skim milk diluent in 1X PBS) for 2 h at room temperature. Plasma samples were then diluted into dilution buffer (5% skim milk diluent in 1X PBST) and titrated down the plate at 50 μL/well. After 1 hour incubation at 37°C, the wells were vigorously washed six times with PBST (1× PBS, 0.05% Tween 20). anti-IgG-HRP antibody was diluted into PBST (1:15000), and 50 μL was added to each well. After a 60 minute incubation at 37 °C, the anti-IgG-HRP was decanted and washed six times with PBST. 50 μL of 3,3′,5,5′-tetramethylbenzidine (TMB, Thermo Fisher Scientific) was then added to each well, followed by 1 M hydrochloric acid as a stopping solution. The colorimetric reaction was analyzed by recording absorbance at 450 nm in Tecan Infinite M Nano plate reader and resulted plotted for display in GraphPad Prism. The limit of quantification was defined here as 3 standard deviations above the mean signal from wells tested without plasma (0.065 Absorbance units).

5.5 Results and Discussion

5.5.1. Peptide Selection and Bioconjugation

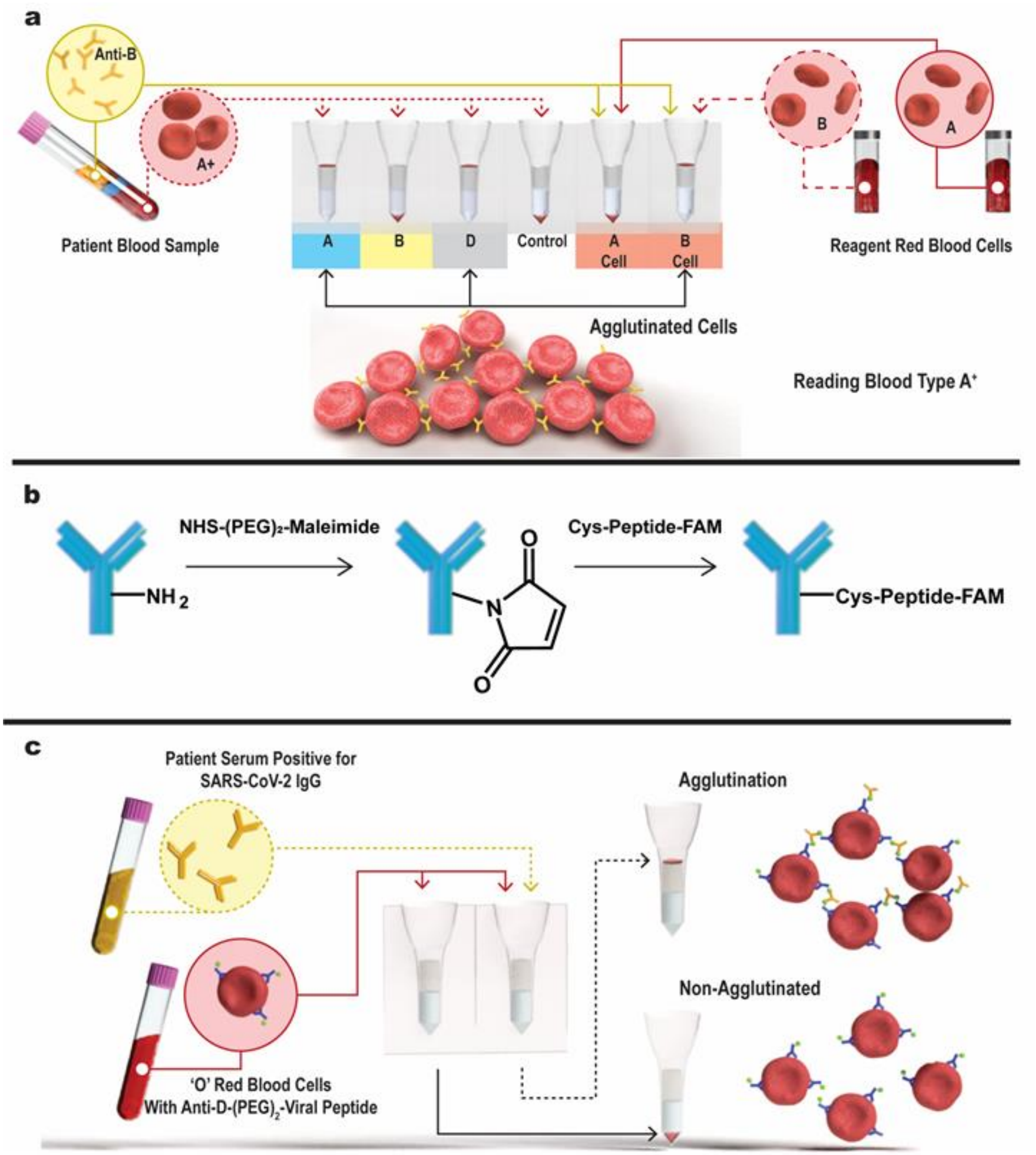

a
Anti-B
A+
Patient Blood Sample
A
B
D
Control
A Cell
B Cell
B
A
Reagent Red Blood Cells
Agglutinated Cells
Reading Blood Type A⁺
b
NH 2
NHS-(PEG)$_2$-Maleimide
O
N
O
Cys-Peptide-FAM
Cys-Peptide-FAM
c
Patient Serum Positive for SARS-CoV-2 IgG
Agglutination
Non-Agglutinated
'O' Red Blood Cells With Anti-D-(PEG)$_2$-Viral Peptide

Figure 5.2 Schematic of blood-typing CAT assay and the introduction of antibody-peptide bioconjugates to produce SARS-CoV-2 serology assay. (a) In a typical blood typing assay, RRBCs are incubated with patient samples on a gel-card prior to centrifugation to generate a pattern of agglutination results to determine a blood type. (b) Reaction scheme employed to produce the antibody-peptide bioconjugate in a two-step process.

We designed a SARS-CoV-2 specific serology assay using an agglutination approach, based on the widespread availability of simple CAT technology commonly employed for blood typing (Figure 5.2a). In a standard blood typing assay, patient RBCs and plasma are separately reacted with reagent antibodies or RBCs, respectively, in gel cards to identify a specific blood type. When RBCs agglutinate, they cannot pass through the gel card, and hence a visible red line is observed following card centrifugation. In Figure 5.2a, agglutination of (i) patient antibodies with B+ reagent cells, and (ii) patient RBCs with reagent anti-A-IgM and anti-D-IgM, confirms the A+ blood type. In contrast, we introduced an antibody-peptide bioconjugate which would aggregate RRBCs only in the presence of antibodies against SARS-CoV-2 (Figure 5.2b). We selected anti-D-IgG as the antibody scaffold, based on its strong affinity for the Rh D-antigen on R2R2 cells. Peptides from the SARS-CoV-2 spike protein were selected based on emerging computational and experimental epitope mapping studies,[17] and synthesised via automated solid-phase peptide synthesis (SPSS) with C-terminal cysteine tags and N-terminal FAM labels (Figure 5.1.). The bioconjugation process consisted of reacting NHS-PEG(2)-maleimide with the lysine side-chains of non-reduced anti-D-IgG, then coupling the cysteine-terminated peptide to the maleimide-tagged antibodies via thioether bonds (Figure 5.2c).

Table 5.1 List of peptides synthesized, Sars-Cov-2 protein sequence, and the start and end amino acid position on the protein sequence.

Peptide ID	Protein	Name	Seq start	Seq finish
0	ACE2 Binding			
1	**SARS-CoV-2 Spike**	**SARS-CoV-2-S491-505**	**491**	**505**
2	**SARS-CoV-2 Spike**	**SARS-CoV-2-S802-819**	**802**	**819**
3	SARS-CoV-2 M	SARS-CoV-2-M1-24	1	24
4	SARS-CoV-2 M	SARS-CoV-2-M132-151	132	151
5	**SARS-CoV-2 Spike**	**SARS-CoV-2-S888-909**	**888**	**909**
6	SARS-CoV-2 Spike	SARS-CoV-2-S675-687	675	687

Table 5.2 **List of peptides synthesized with modifiers.**

Peptide ID	Sequence
0	FAM-GGG-IEEQAKTFLDKFNHEAEDLFYQS-GGG-CG
1	**FAM-GGG-PLQSYGFQPTNGVGY-GGGCG**
2	**FAM-GGG-FSQILPDPSKPSKRSFIE-GGGCG**
3	FAM-GGG-MADSNGTITVEELKKLLEQWNLVI-GGGCG
4	FAM-GGG-PLLESELVIGAVILRGHLRI-GGGCG
5	**FAM-GGG-FGAGAALQIPFAMQMAYRFNGI-GGGCG**
6	FAM-GGG-QTQTNSPRRARSV-GGGCG

Antibody-peptide bioconjugates were designed to bind SARS-CoV-2-specific antibodies raised in response to viral infection. Peptides were selected based on early predictions from database mining projects, where P1, P2 and P5 are all elements of the SARS-CoV-2 spike protein (Figure 5.1, Table 5.2 and Table 5.3). Following crude purification of anti-D-IgG (Figure 5.3.) and the simple bioconjugation procedure outlined in Figure 5.2c, the antibody/peptide bioconjugates were analysed by gel electrophoresis to confirm peptide labelling. Fluorescence analysis revealed clear bands ~150 kDa only for those samples labelled with FAM, with no evidence of fluorescence emission for the anti-D-IgG control (Figure 5.4a). After staining the same gel with Coomassie Blue, all four bioconjugates were clearly visible, confirming successful peptide labelling (Figure

5.4b). Labelling efficiency was estimated at 5-10 % based on Nanodrop analysis, from ~10-20-fold molar excess of peptide, above which we observed evidence of protein precipitation, in agreement with previous studies.[15] While the labelling efficiency was lower than expected, the gels showed some evidence of impurities in the purified anti-D-IgG sample, and the FAM incorporation into peptides may not be 100% as the dye incorporation is the final step in the peptide synthesis process. As expected, the bioconjugates bound red cells efficiently, showing strong and titratable FAM intensity in flow cytometry (Figure 5.4c). Saturation occurred at a bioconjugate:cell ratio of 1-3:1, and it was later found that maintaining this ratio in gel-card assays was critical to forming clear agglutination signal.

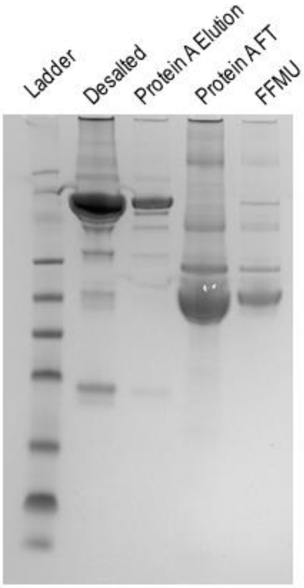


Figure 5.3 SDS-PAGE gel showing results of Protein A-purification of anti-D-IgG FFMU supernatant to purified and desalted anti-D-IgG. “Protein A FT” = flow through from Protein A column; “Protein A Elution” = eluant from Protein A column, “Desalted” = concentrated product after buffer exchange into PBS.

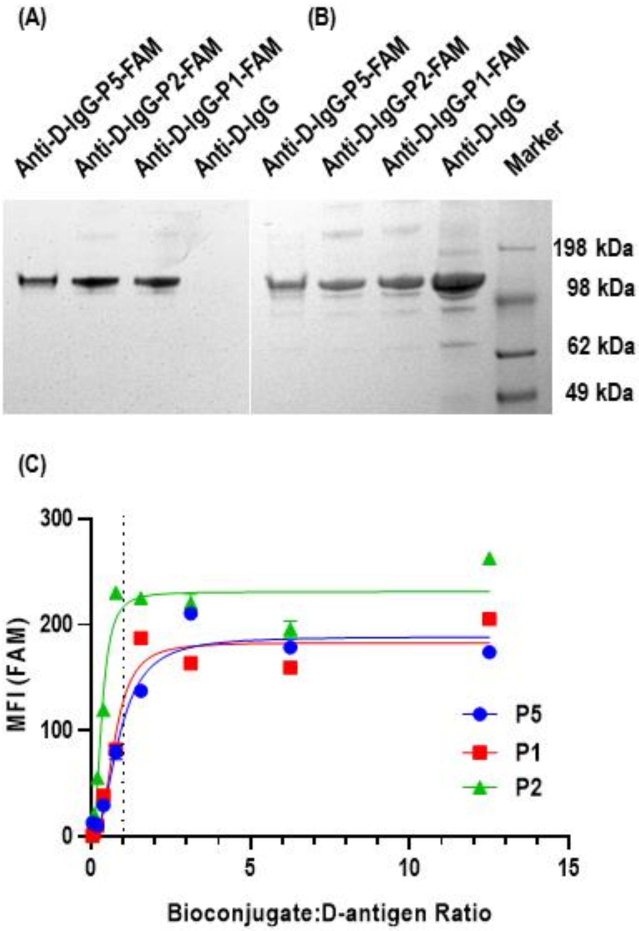


Figure 5.4 Anti-D-IgG/peptide bioconjugate characterisation. (a) Fluorescence scan of protein gel under FAM filter. (b) Brightfield image of the same gel following Coomassie staining. (c) Graph showing the bioconjugate binding to D+RRBCs using flow cytometry, and the effect of bioconjugate titration. Dotted line indicates equimolar bioconjugate and D-antigen in the incubation reaction.

5.5.2. Red Blood Cell Agglutination

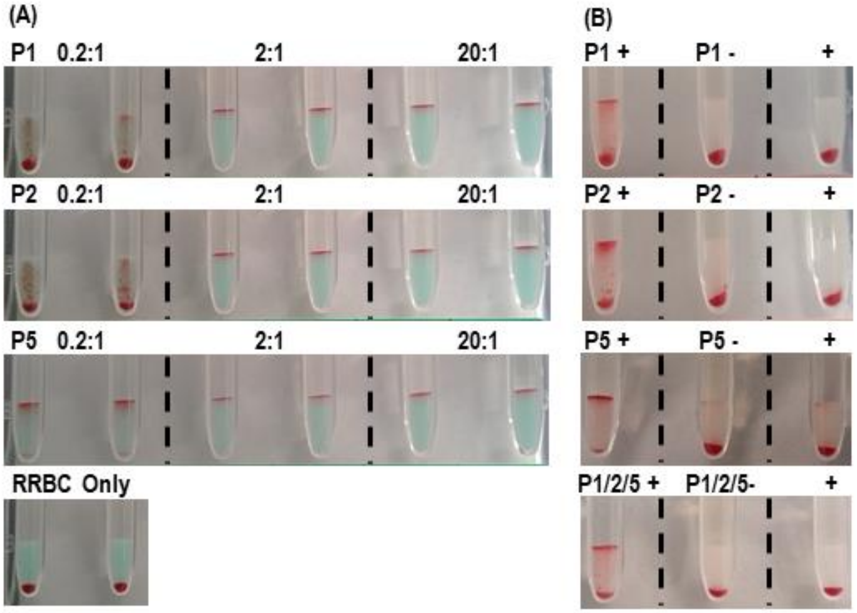
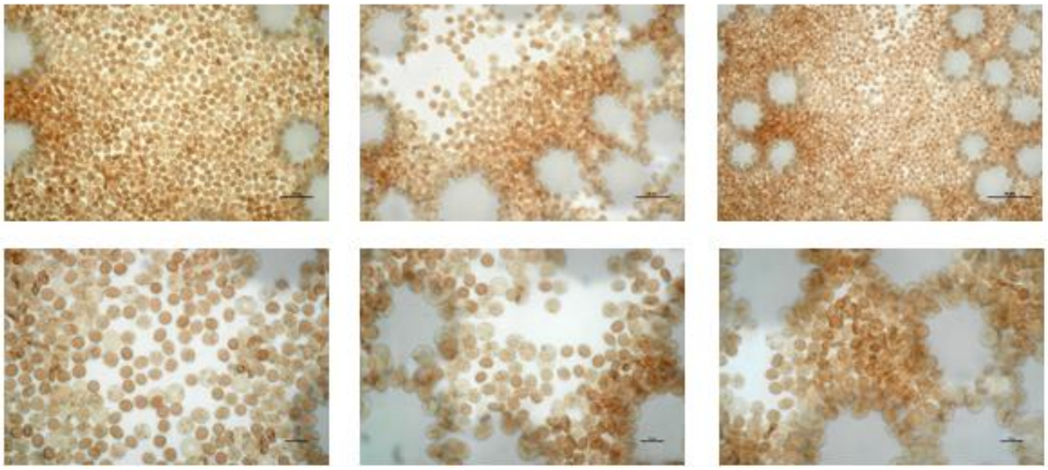


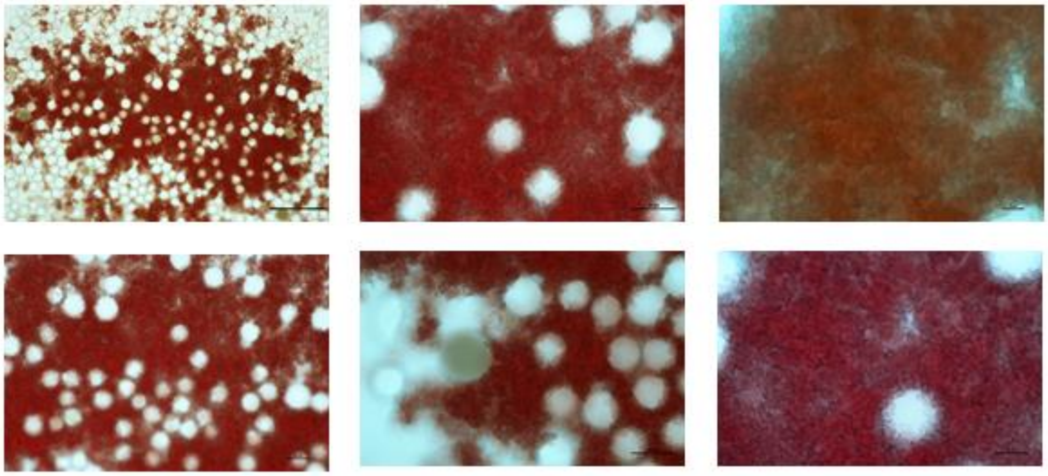
Figure 5.5 Optimisation of gel-card assays for SARS-CoV-2 serology. (a) Testing the ability of bioconjugates to crosslink RRBCs independent of attached peptide, using anti-IgG in solution in PBS. "P*n*" indicates the peptide used in the reactions (*n* = 1, 2 or 5), and ratios indicate the bioconjugate to D-antigen (on cells). (b) Demonstrating the selective agglutination of SARS-CoV-2 antibodies present in a clinical sample in comparison to negative controls, using bioconjugate-saturated RRBCs. Reactions involving SARS-CoV-2-positive samples indicated by "+" and SARS-CoV-2-negative samples indicated by "-". Reactions labelled "P1/2/5" indicate that bioconjugates were pre-mixed to provide the same total peptide concentration as used for other reactions.

We next investigated the agglutination potential of bioconjugate-coated RRBCs. To ensure that the bioconjugate did not inhibit the potential for RRBCs to agglutinate, we first incubated bioconjugate-coated RRBCs with anti-IgG antibodies prior to centrifugation (Figure 5.5a).[18] While RRBCs alone spun through the gel card as a negative control, if the RRBCs were saturated with bioconjugate (*i.e.* bioconjugate:D-antigen ratio >1:1), we observed strong agglutination with no evidence of free cells travelling through the gel card. If the RRBCs were not saturated (*e.g.* 0.2:1 ratio), we observed only partial agglutination in, or on top of, the gel columns, presumably because there was not sufficient bioconjugate present to crosslink cells with the anti-IgG ("Coombs' reagent”). This suggested that operating the assays under conditions of bioconjugate saturation is required for successful agglutination and retention of cells above the gel column.

We next progressed to optimize gel card agglutination assays to distinguish between SARS-CoV-2-positive and SARS-CoV-2-negative patient samples. Preliminary optimization experiments suggested that common protocols used in blood-typing assays were also appropriate for SARS-CoV-2 serology assays; i.e. 5-10-minute incubation of gel cards at 25°C, followed by 11 minutes centrifugation. Consistent with our findings related to Figure 5.5a, a significant factor affecting the degree of agglutination was the bioconjugate:cell ratio. If this ratio was not high enough to ensure cell saturation, a higher proportion of RRBCs were observed to pellet upon centrifugation. When using saturated bioconjugate-cell reagents, we were able to detect positive agglutination results with each of the three bioconjugates tested. Importantly, negative control reactions involving either SARS-CoV-2-negative samples, or RRBCs and SARS-CoV-2-positive samples without bioconjugates, all revealed no agglutination behavior. Only rarely did we observe “complete” agglutination (with no cells pelleted following centrifugation), which was expected because not all cell-bound antibodies were tagged with peptide. Reactions involving mixed bioconjugates (mixtures prepared after cell incubation) were also tested, showing similar results as for single-bioconjugate reactions, which is important, as we expect that multiple immunodominant peptides will be required to minimize false positives in large cohorts. Importantly, during the optimization phase of assay development, we occasionally observed false positive results, *i.e.* a red line appearing above the gel column after incubation of bioconjugated-coated RRBCs and SARS-CoV-2-negative plasma. However, upon microscopy it was revealed that these cells were not agglutinated (Figure 5.6); this was mainly related to the presence of glycerol carried over from the bioconjugate stock solution. While 1:1 glycerol/PBS mixtures are commonly used for long-term -20°C storage of bioconjugates, the glycerol carryover into RRBCs for agglutination reaction should be minimized or removed.



(A) False positives



(B) True Positives

Figure 5.6 Light microscopy images of RRBCs collected from the top layer of gel cards. (a) False positive gel cards revealed cells that were not agglutinated, and showed some apparent interactions with the gel beads. While retained on top of the gel card, the lack of agglutination confirms the false positive status. This was most often associated with the presence of glycerol in the RRBC stocks, carried over from the bioconjugate solutions. (b) True positive gel cards revealed clear regions of strong agglutination, a 3D network of cells aggregated together in a continuous pattern.

5.5.3. Testing with Clinical Samples

Table 5.3 List of clinical samples used in the study

Sample	Time between PCR and sample collection (Date of collection)	Sample Matrix
COVID19-positive (PCR-confirmed)		
1	30	Serum
2	34	Serum
3	8	Serum
4	27	Serum
5	28	Serum
COVID19-negative (collected prior to pandemic)		
6	07/19/2017	Serum
7	01/20/2014	Plasma
8	07/28/2014	Plasma
9	05/21/2012	Serum
10	12/02/2015	Serum

Following optimization of the gel card assays to distinguish between SARS-CoV-2-positive samples and negative controls, we tested 10 clinical samples in both gel cards and indirect IgG ELISA. The ELISA was designed to capture and detect IgG antibodies from plasma which bound to SARS-CoV-2 proteins coated onto the plates. This assay cannot detect IgM antibodies which are also likely to be present in many samples,[10] however we expect that IgM levels are likely to recede over time in at least some individuals, whereas IgG levels are likely to remain high over time and hence are appropriate to confirm immune response to infection. We observed strong IgG signals for all 5 PCR-confirmed SARS-CoV-2-positive samples, while all negative samples tested at or below the LOQ for the assay. Given that samples used for serology testing contain a unique and complex polyclonal mixture of IgG and IgM antibodies, it is not meaningful to calibrate or quantify relative to a standard, hence the data is regularly reported based on dilutions or titres.[10-12] Notably, one of the positive samples was collected only 8 days post-PCR, suggesting that the assay may be capable of detecting IgG levels well before they are expected to peak (~19 days[10]). Importantly, the SARS-CoV-2-negative samples tested were collected prior to the pandemic (Table 5.3), as it is possible that samples collected from apparently healthy people during the pandemic could indeed be from asymptomatic carriers. We then tested the same 10 samples in gel card agglutination tests using a mixture of all three bioconjugates, and included negative controls for which RRBCs were not coated with bioconjugates. Here we added an additional 5-minute incubation step at 4°C prior to centrifugation as this is one way of enhancing agglutination reactions to achieve strongly visible bands across a range of samples. The same effect could likely be achieved by using agglutination enhancing solutions, including low ionic-strength saline (LISS). All 5 PCR-confirmed SARS-CoV-2-positive samples yielded positive agglutination results, while the negative samples showed no agglutination above the gel. The lack of positive results in ELISA or agglutination tests for these negative samples is particularly encouraging, because in a recent study four out of five SARS-CoV-2-negative blood samples showed high levels of antibodies against seasonal coronaviruses but no cross-reactive antibodies that bind SARS-CoV-2.[19] This is consistent with earlier estimates that nearly 100% of adolescents have antibodies against seasonal coronaviruses.[20] This analysis shows that the gel card agglutination tests can provide serological results for SARS-CoV-2 infection within 30 minutes, using an approach consistent with blood-typing assays used routinely in hospital labs around the world.

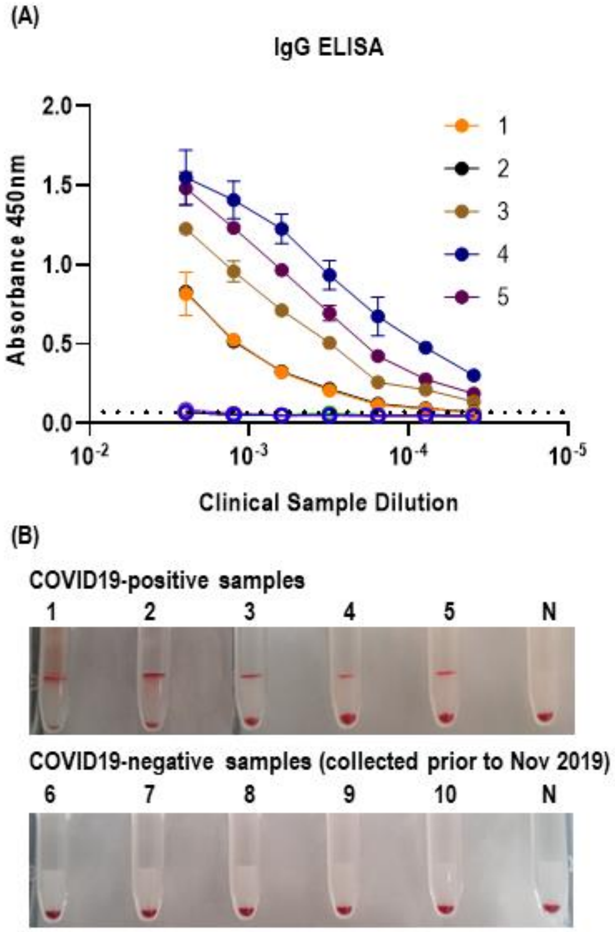


Figure 5.7 Clinical sample analysis comparing indirect IgG ELISA against agglutination approach using RRBCs coated with P1, P2 and P5 bioconjugates prior to mixing. (A) Indirect IgG ELISA results comparing 5 PCR-confirmed SARS-CoV-2-positive samples (filled circles) against 5 samples collected from healthy individuals prior to SARS-CoV-2 pandemic (empty circles). The dotted line indicates the limit of quantification (LOQ) for the assay, determined to be three standard deviations above wells containing PBST instead of clinical sample. (B) Digital images of gel card assays comparing 5 PCR-confirmed SARS-CoV-2-positive samples against 5 samples collected from healthy individuals prior to SARS-CoV-2 pandemic. Negative controls ("N") samples were performed using RRBCs and clinical sample (sample 5 for positives; sample 10 for negatives) without bioconjugates.

5.6 Conclusion

In this study, we have taken widely used blood-typing tests and converted them into SARS-CoV-2 serology tests. Given the rapid turnaround time, high throughput, and level of clinical acceptance, we suggest that with further testing in large sample cohorts to accurately characterize false positive/negative rates, CAT assays could provide an alternative to ELISAs. At the time of writing this manuscript, there is still much to learn about SARS-CoV-2 virology, the nature and spectrum of immune responses, and the utility of serology assays as many countries begin to emerge from strict quarantine. Key limitations at present include the lack of knowledge on which SARS-CoV-2 peptides are immunodominant, whether or not predicted B-cell epitopes will efficiently bind antibodies (IgG, IgM, *etc*) in patient blood, and what is the level of cross-reactivity between antibodies raised against previous coronavirus infections in large sample cohorts. It is also unclear if distinguishing between IgM and IgG would be clinically relevant, given the lack of consistent class switching trends reported to date, however it would likely be relevant in terms of understanding the immune response in more detail. One recent study used a peptide microarray to determining IgG/IgM binding epitopes from the blood of 40 infected individuals, and identified that 80% of the samples contained both IgG and IgM antibodies against just four epitope sequences (which includes our current P2 peptide). [21] This confirms that pursuing an approach involving multiple bioconjugates is likely required in order to minimize false-negative results in large sample cohorts. An alternative approach would be to create bioconjugates using whole proteins instead of peptide epitopes; the advantage would be that protein sequences are often much faster to determine in comparison to identifying a minimal list of immunodominant peptides, which is important in terms of pandemic response. However, bioconjugation reactions involving whole (and potentially novel) proteins are likely to be less efficient due to size/charge comparisons, requiring individual tailoring of reaction conditions whereas many peptide bioconjugates can be prepared simultaneously.

In this study we have transformed blood-typing tests into SARS-CoV-2 serology tests using robust gel card agglutination reactions in combination with easily prepared antibody-peptide bioconjugates. We found concordance between results for gel card assays and an indirect IgG

ELISA across 10 clinical samples, 5 of which were PCR-confirmed SARS-CoV-2 positive samples. During assay development, we found that it was critical to ensure that RRBCs were saturated with bioconjugates, otherwise agglutination did not occur or was extremely inefficient. We also found that very small amounts of glycerol in the RRBC cell stocks caused false positive results, hence care must be taken when choosing buffer and storage solutions to investigate the effects of all additives. This assay was designed and operated based on the infrastructure routinely available in blood-typing laboratories worldwide, and this approach is now ready to be evaluated for clinical application with a large sample cohort, leading to the next phase of industry partnership for scale-up and distribution.

5.6 References

1. Coronavirus Update (Live), Worldometer. https://www.worldometers.info/coronavirus/
2. Australian Government Department of Health. https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/coronavirus-covid-19-current-situation-and-case-numbers.
3. Carter, L. J.; Garner, L. V.; Smoot, J. W.; Li, Y.; Zhou, Q.; Saveson, C. J.; Sasso, J. M.; Gregg, A. C.; Soares, D. J.; Beskid, T. R.; Jervey, S. R.; Liu, C., Assay Techniques and Test Development for COVID-19 Diagnosis. *ACS Central Science* **2020,** *6* (5), 591-605.
4. Udugama, B.; Kadhiresan, P.; Kozlowski, H. N.; Malekjahani, A.; Osborne, M.; Li, V. Y. C.; Chen, H.; Mubareka, S.; Gubbay, J. B.; Chan, W. C. W., Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano* **2020,** *14* (4), 3822-3835.
5. Seo, G.; Lee, G.; Kim, M. J.; Baek, S.-H.; Choi, M.; Ku, K. B.; Lee, C.-S.; Jun, S.; Park, D.; Kim, H. G.; Kim, S.-J.; Lee, J.-O.; Kim, B. T.; Park, E. C.; Kim, S. I., Rapid Detection of COVID-19 Causative Virus (SARS-CoV-2) in Human Nasopharyngeal Swab Specimens Using Field-Effect Transistor-Based Biosensor. *ACS Sensors* **2020,** *14* (4), 5135-5142.
6. Qiu, G.; Gai, Z.; Tao, Y.; Schmitt, J.; Kullak-Ublick, G. A.; Wang, J., Dual-Functional Plasmonic Photothermal Biosensors for Highly Accurate Severe Acute Respiratory Syndrome Coronavirus 2 Detection. *ACS Sensors* **2020**, https://doi.org/10.1021/acsnano.0c02439.
7. Lipsitch, M.; Kahn, R.; Mina, M. J., Antibody testing will enhance the power and accuracy of COVID-19-prevention trials. *Nature Medicine* **2020**, https://doi.org/10.1038/s41591-020-0887-3.
8. Lee, C. Y.-P.; Lin, R. T. P.; Renia, L.; Ng, L. F. P., Serological Approaches for COVID-19: Epidemiologic Perspective on Surveillance and Control | Immunology. *Frontiers in Immunology* **2020**, https://doi.org/10.3389/fimmu.2020.00879.

9. Amy K. Winter, S. T. H., The important role of serology for COVID-19 control - The Lancet Infectious Diseases. *The Lance Infectious Diseases* **2020**, https://doi.org/10.1016/S1473-3099(20)30322-4.

10. Long, Q.-X.; Liu, B.-Z.; Deng, H.-J.; Wu, G.-C.; Deng, K.; Chen, Y.-K.; Liao, P.; Qiu, J.-F.; Lin, Y.; Cai, X.-F.; Wang, D.-Q.; Hu, Y.; Ren, J.-H.; Tang, N.; Xu, Y.-Y.; Yu, L.-H.; Mo, Z.; Gong, F.; Zhang, X.-L.; Tian, W.-G.; Hu, L.; Zhang, X.-X.; Xiang, J.-L.; Du, H.-X.; Liu, H.-W.; Lang, C.-H.; Luo, X.-H.; Wu, S.-B.; Cui, X.-P.; Zhou, Z.; Zhu, M.-M.; Wang, J.; Xue, C.-J.; Li, X.-F.; Wang, L.; Li, Z.-J.; Wang, K.; Niu, C.-C.; Yang, Q.-J.; Tang, X.-J.; Zhang, Y.; Liu, X.-M.; Li, J.-J.; Zhang, D.-C.; Zhang, F.; Liu, P.; Yuan, J.; Li, Q.; Hu, J.-L.; Chen, J.; Huang, A.-L., Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nature Medicine* **2020**, https://doi.org/10.1038/s41591-020-0897-1.

11. Assis, R. R. d.; Jain, A.; Nakajima, R.; Jasinskas, A.; Felgner, J.; Obiero, J. M.; Adenaiye, O.; Tai, S.; Hong, F.; Norris, P. J.; Stone, M.; Simmons, G.; Bagri, A.; Schreiber, M.; Buser, A.; Holbro, A.; Battegay, M.; Hosimer, P.; Noesen, C.; Milton, D. K.; Group, P. S.; Davies, D. H.; Contestable, P.; Corash, L. M.; Busch, M. P.; Felgner, P. L.; Khan, S., Analysis of SARS-CoV-2 Antibodies in COVID-19 Convalescent Blood using a Coronavirus Antigen Microarray. *bioRxiv* **2020**, https://doi.org/10.1101/2020.04.15.043364.

12. To, K. K.-W.; Tsang, O. T.-Y.; Leung, W.-S.; Tam, A. R.; Wu, T.-C.; Lung, D. C.; Yip, C. C.-Y.; Cai, J.-P.; Chan, J. M.-C.; Chik, T. S.-H.; Lau, D. P.-L.; Choi, C. Y.-C.; Chen, L.-L.; Chan, W.-M.; Chan, K.-H.; Ip, J. D.; Ng, A. C.-K.; Poon, R. W.-S.; Luo, C.-T.; Cheng, V. C.-C.; Chan, J. F.-W.; Hung, I. F.-N.; Chen, Z.; Chen, H.; Yuen, K.-Y., Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study - The Lancet Infectious Diseases. *The Lance Infectious Diseases* **2020,** *20* (5), 565-574.

13. Bryan, A.; Pepper, G.; Wener, M. H.; Fink, S. L.; Morishima, C.; Chaudhary, A.; Jerome, K. R.; Mathias, P. C.; Greninger, A. L., Performance Characteristics of the Abbott Architect SARS-CoV-2 IgG Assay and Seroprevalence in Boise, Idaho. *Journal of Clinical Microbiology* **2020**, https://doi.org/10.1128/JCM.00941-20.

14. Mabey, D.; Peeling, R. W.; Ustianowski, A.; Perkins, M. D., Diagnostics for the developing world. *Nature Reviews Microbiology* **2004,** *2* (3), 231-240.

15. Kemp, B.; Rylatt, D.; Bundesen, P.; Doherty, R.; McPhee, D.; Stapleton, D.; Cottis, L.; Wilson, K.; John, M.; Khan, J.; et, a., Autologous red cell agglutination assay for HIV-1 antibodies: simplified test with whole blood. **1988,** *241* (4871), 1352-1354.

16. Wilson, K. M.; Gerometta, M.; Rylatt, D. B.; Bundesen, P. G.; McPhee, D. A.; Hillyard, C. J.; Kemp, B. E., Rapid whole-blood assay for HIV-1 seropositivity using an Fab-peptide conjugate. *Journal of Immunological Methods* **1991,** *138* (1), 111-119.

17. Grifoni, A.; Sidney, J.; Zhang, Y.; Scheuermann, R. H.; Peters, B.; Sette, A., A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host & Microbe* **2020,** *27* (4), 671-680.

18. N, Y.; H, M.; L, G.; W, S.; G, G., Paper-based Assay for Red Blood Cell Antigen Typing by the Indirect Antiglobulin Test. *Analytical and bioanalytical chemistry* **2016,** *408* (19), 5231-5238.

19. Khan, S.; Nakajima, R.; Jain, A.; de Assis, R. R.; Jasinskas, A.; Obiero, J. M.; Adenaiye, O; Tai, S.; Hong, F.; Milton, D. K.; Davies, H.; Felgner, P. L., Analysis of Serologic Cross-Reactivity Between Common Human Coronaviruses and SARS-CoV-2 Using Coronavirus Antigen Microarray. *bioRxiv* **2020**, https://doi.org/10.1101/2020.03.24.006544.

20. Zhou, W.; Wang, W.; Wang, H.; Lu, R.; Tan, W., First infection by all four non-severe acute respiratory syndrome human coronaviruses takes place during childhood. *BMC Infectious Diseases* **2013**, 13, 433.

21. Wang, H.; Hou, X.; Wu, X.; Liang, T.; Zhang, X.; Wang, D.; Teng, F.; Dai, J.; Duan, H.; Guo, S.; Li, Y.; Yu, X., SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. *bioRxiv* **2020,** https://doi.org/10.1101/2020.03.26.994756 .

Chapter 6 Summary and Future Directions

6.1 Summary of Research

This dissertation explores the concept of information-directed self-assembly by encoded polymeric strands, particularly with the use of dynamic covalent interactions. This understanding is key to expanding the complexity and stability of self-assembled structures.

In the second chapter, we discussed the design of a self-assembly process by which a multi-role Lewis-acid catalyst, scandium triflate, was first used in high concentrations to dissociate peptoid strands encoded with amine- and aldehyde-pendant groups (‘1’s and’0’s respectively). Subsequent extraction of a majority of the scandium triflate, shifted the equilibrium towards imine formation, and immediately produced in- and out-of-registry duplex formations. Overtime these could be annealed at either room temperature or 70°C to form sequence-selective molecular ladders. These information-bearing ladders were demonstrated to be formed under both a competitive environment, with the addition of a non-complementary sequence to the reaction mixture, and in a one-pot, simultaneous hybridization with other encoded oligomer sequences to form distinct, in-registry ladder species. Finally, this process was used as a non-destructive data recognition medium where a strand encoded with a ‘secret message’ could be identified and decrypted through MS, by reading the intense MS signal produced from the strand’s hybridization with a known library of sequences containing the ‘secret message’s’ complementary sequence.

The third chapter builds upon the sequence-selective hybridization in Chapter 2 by introducing boronate ester-forming moieties into the strands as a second, orthogonal dynamic covalent interaction, ultimately directing the assembly of base-4 molecular ladders and grids. We first explored analogous pH-mediated reactions for each of the dynamic covalent pairs which would allow for sequence-selective self-assembly in aqueous environment, realized by purifying the constituent strands under acidic conditions prior to their self-assembly in water at a pH of roughly 9. This chapter then utilized NMR to conform the orthogonality between the two dynamic covalent interactions before fabricating an orthogonal molecular grid from a core strand comprised of catechol and aldehyde functional groups and two complementary side strands. The design of our molecular grid enabled the sequential reaction of the two dynamic pairs as well a simultaneous approach. These conditions led to the successful formation of a molecular ladder and grid comprised of the two chemistries. This effectively doubled the information density of our system, which in turn highlights the potential of our system in forming complex highly controlled sequence-specific assemblies.

In the fourth chapter, a furan-protected maleimide primary amine monomer was developed to enable the encoding of peptoid oligomers bearing both furan and maleimide groups precluding premature self-hybridization, or ‘hair-pinning.’ The monomer was designed to exclusively produce the *exo*-isomer of the Diels-Alder adduct as to enable the microwave-assisted synthesis of peptoids, dramatically decreasing coupling times and thus overall speed of synthesis. At temperatures above 120°C, the furan protecting the maleimide monomer was removed, and its low boiling point ensured it escaped the uncapped reaction mixture, preventing interference with future hybridizations. Heating of the reaction mixture to elevated temperatures not only deprotected the maleimide, but dissociated strands which were subsequentially hybridized at reduced temperatures. Self-assembly and conversion of single strand species was characterized at 60°C and 100°C. The thiol-Michael addition of thiols to dienes was employed to determine the degree of alignment between species, with an increase in in-registry ladders being observed at 60°C for extended periods. Finally, a method in which reaction mixtures at 140°C were cooled to 100°C for 24 hrs before further cooling and annealing at 60°C gave evidence to suggest the sequence-specific assembly of ladder species utilizing a Diels-Alder cycloaddition.

The fifth chapter describes the development of a diagnostic device detecting antibodies produced by humans in response to SARS-CoV-2. As development was a pressing matter, the team chose to use rapid gel card agglutination assays commonly employed for blood typing. These blood typing tests are employed in hospitals worldwide, are high-throughput, fast (10-30 minutes) and automated in most cases. Peptide-antibody bioconjugates, which facilitate red cell crosslinking only in the presence of plasma containing antibodies against SARS-CoV-2, were developed in-house utilizing only the materials on hand. Thus, the thiol-Michael addition was again employed to selectively react cysteine-bearing peptides to a maleimide linker attached to red-cell binding antibodies. Peptides were chosen from current literature as epitopes with computational evidence of high specificity towards antibodies against SARS-CoV-2. Ultimately, detection was possible using this device within 30 minutes.

6.2 Dynamic Covalent Assembly Future Directions

Molecular ladders are one-dimensional assembled structures, and while two-dimensional grids were realized in Chapter 3, they were limited to three strands. Greater dimensions are possible by dynamic covalent interactions, as seen in work on covalent organic frameworks and molecular cages, but again formed by functionally-limited constituents. The following sections will outline the future direction of information-directed assembly to achieve two- and even three-dimensional structures.

6.2.1. 2-D Grid Structures

Given the success of base-4 molecular grids, initial attempts at larger, complex structures entailed the formation of grids formed from any number of strands (Figure 6.1). Here, it is proposed to begin with a single amine-bearing strand, with pendant groups either on a single side or double sided, depending on the desire for uni- or bi-functional growth. This strand would be reacted with a complementary strand bearing aldehydes on one side and alloc-protected amines on the other. Upon formation, the imines assembled would be reduced by a reducing agent, and the alloc-protecting groups removed with palladium as described in Chapter 2. The assembly could further be reacted with new aldehyde/alloc-protected amine strands and the process repeated any number of times. As each strand is added iteratively, the strands could be functionalized with orthogonal click-chemistries enabling the formation of a customized 'nano-breadboard.'

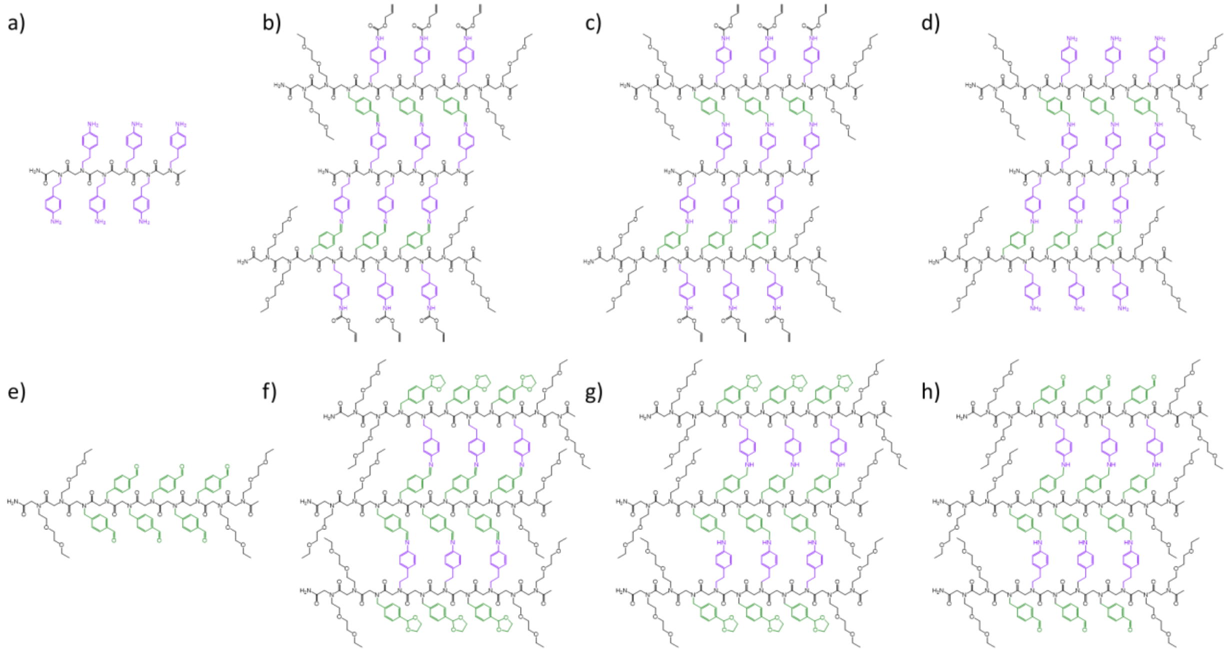


Figure 6.1 Iterative grid assembly. (a) A bi-directional amine bearing core-strand is (b) reacted with two equivalents of a strand bearing free-aldehydes on one face and alloc-protected amines on another. (c) The assembly is reduced and (d) the alloc-protecting groups are removed for further reactions. (e) Similarly, a bi-directional aldehyde bearing core-strand is (b) reacted with two equivalents of a strand bearing free-amines on one face

and acetal-protected aldehydess on another. (c) The assembly is reduced and (d) the acetal-protecting groups are removed for further reactions.

Unfortunately, initial attempts at grid formation were thwarted by the inability to remove palladium from the assembly after alloc-deprotection. The palladium complex was unable to be removed by extraction given its similar solubility to peptoids, and the few solvents it does prefer caused precipitation of the peptoid assemblies. Purification by filtration, as palladium complexes stick to silica, also removed peptoid assemblies from the reaction mixture. Thus, a second attempt consisted of an aldehyde core reacted iteratively with peptoids bearing free-amines and protected-aldehydes. The initial three stranded assembly formed efficiently overnight, and MALDI-MS indicated reduction of the complex with sodium cyanoborohydride by a shift in the mass spectra owing to hydrogen addition upon reduction of the imine bond. Assemblies were deprotected using TFA, and then neutralized by basic water before extraction in an organic layer. This process was repeated twice and resulted in Figure 6.2. The correct assembly is identifiable in the mass spectra, but ultimately the solutions contain a variety of products as the only purification is extraction in the process which doesn’t remove incomplete assemblies. One common issue that arose from this process was the inability to confirm imine reduction, as MALDI-MS is un-precise, commonly resulting in spectra of ±10 m/z units and showing false-positive results of complete imine reduction. To help circumvent these issues, future attempts will attempt the assembly on-resin to readily enable purification of by-products.

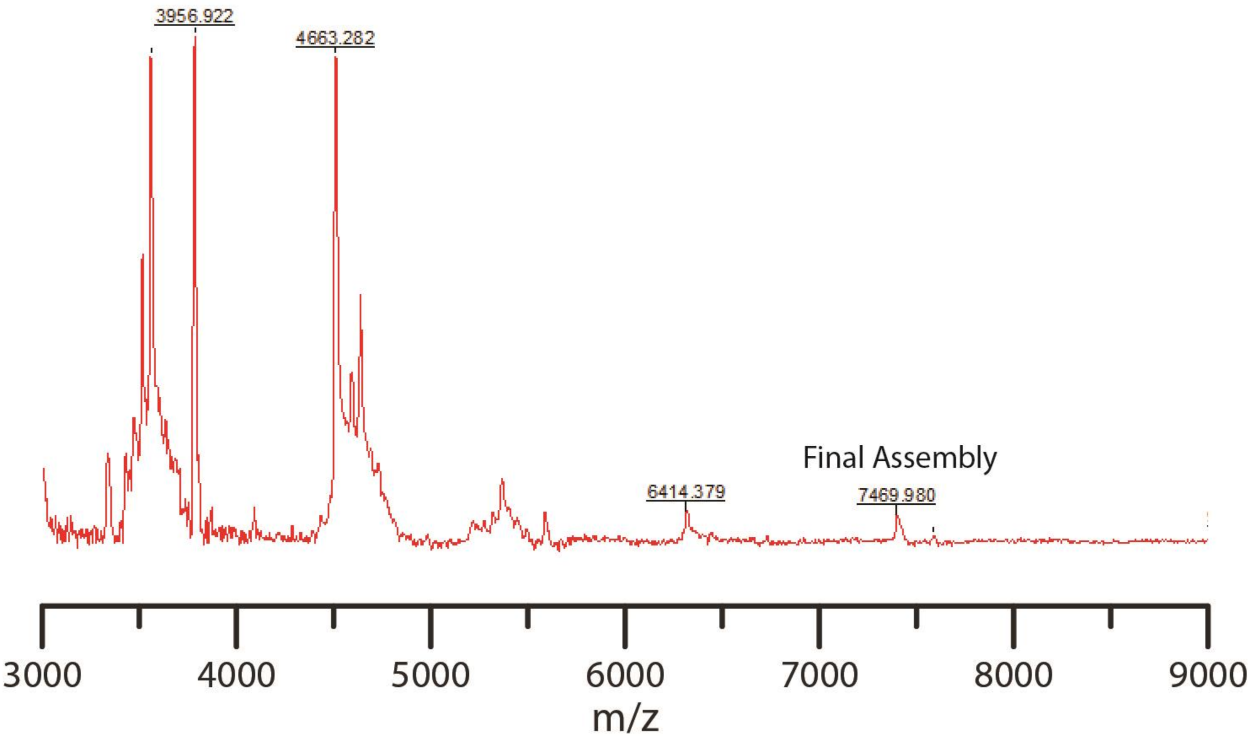


Figure 6.2 MALDI- mass spectra of a grid assembly from five strands.

Ultimately, the self-assembly process envisaged is one not requiring an iterative method, but rather the simultaneous self-assembly of information-encoded species. Thus, we attempted the hybridization of a molecular grid using an imine-forming system of information-bearing sequences, however the complexity of the system triggered the precipitation of strands upon extraction as a large cross-linked network instantly formed upon addition of an aqueous layer (Figure 6.3). Here, the novel Diels-Alder cycloaddition self-assembly process described in Chapter 4 could potentially enable the formation of encoded grids via a thermal cycling and will be attempted in the future.

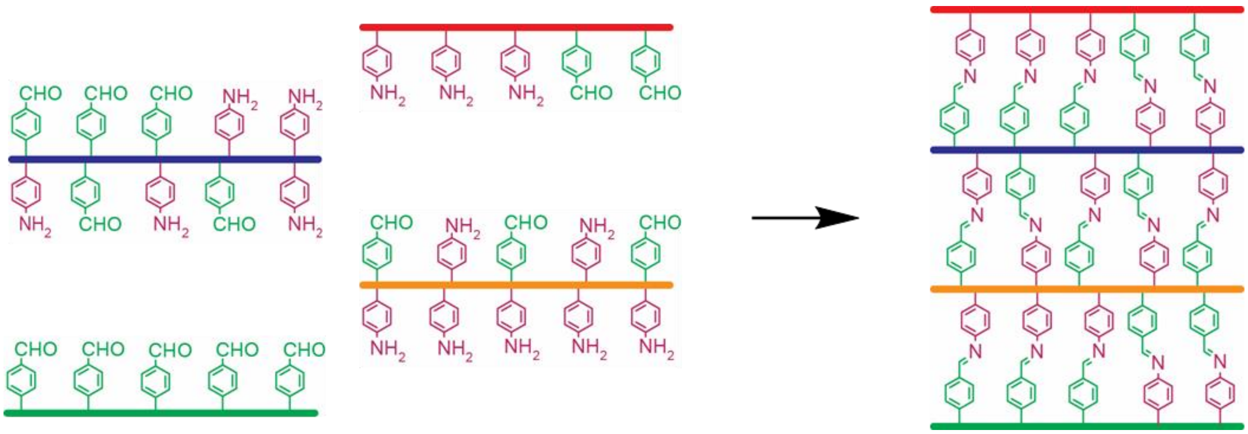


Figure 6.3 Self-assembly of a molecular grid from information-bearing peptoid strands.

6.2.2. Non-linear Branched Structures

Junctions were the first multidimensional information-directed assemblies formed from DNA strands[1]. In an attempt to emulate this work by Nadrian Seeman, three sequences were designed and synthesized that would theoretically form a trimeric junction. However, the mass spectra obtained after using the dissociation/extraction/annealing approach in Chapter 2, indicated unselective formation of the junctions and, interestingly, the assemblies formed were out-of-registry. This phenomena could either be because the assemblies were kinetically trapped or potentially due to the low information-density provided by base-2 encoding. In his profound initial work, Seeman suggests that the repetition of bases within the junction core could cause migratory issues, and thus preclude proper formation. To test this theory, we similarly designed base-4 sequences and subsequently reacted them to form junctions. However, here again we saw incorrect species forming and mixed registries of the assemblies, thus signaling an issue with the self-assembly process rather than information-density. With the new development of thermal cycling utilizing Diels-Alder cycloadditions, we are looking to generate dynamic covalent junctions analogous to the initial DNA junctions shown by Seeman.

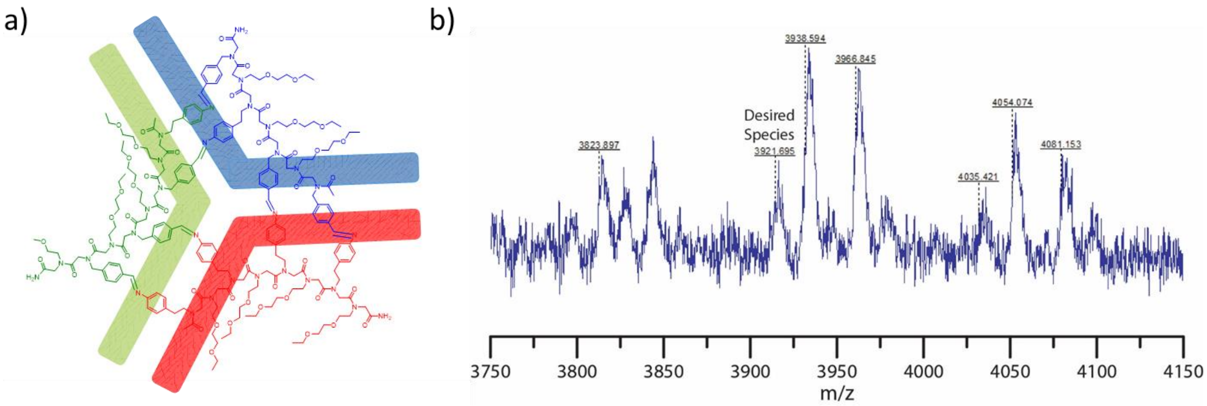
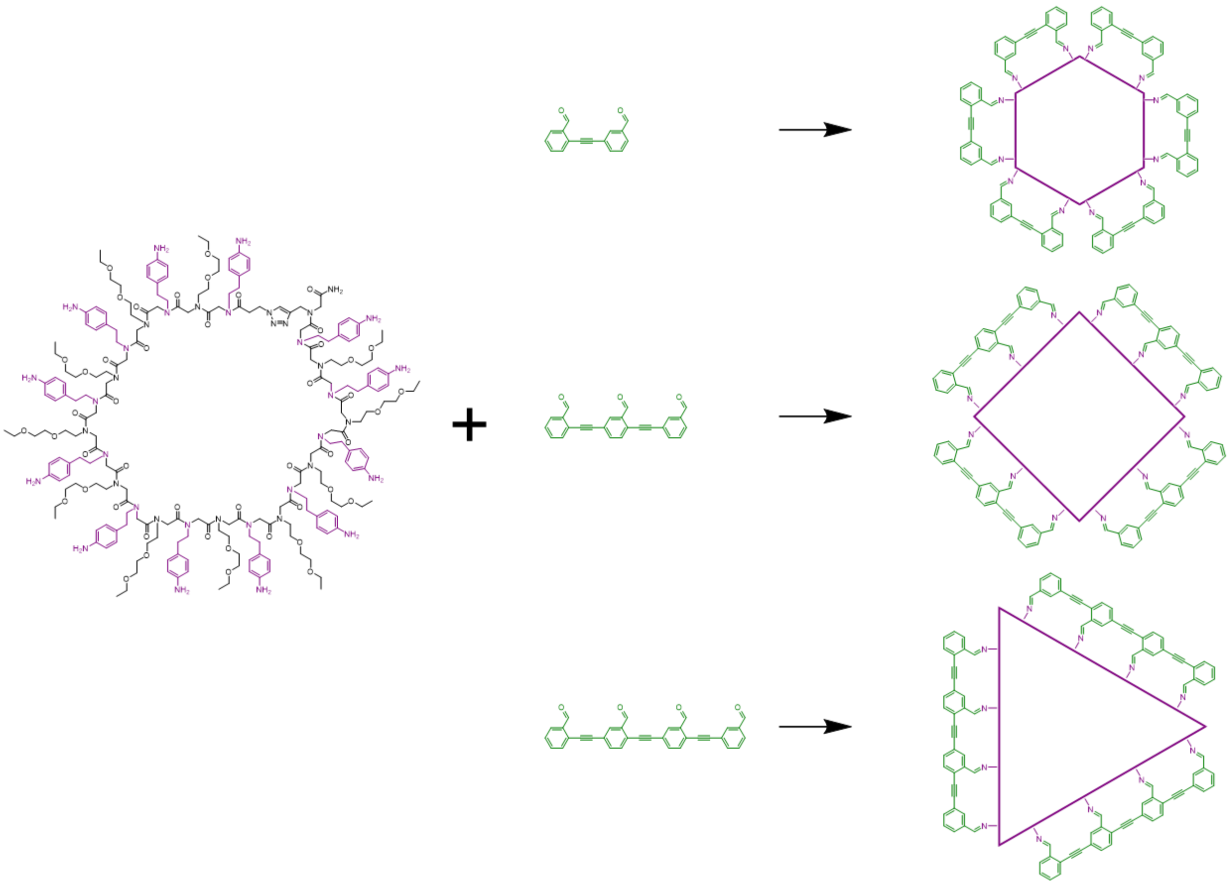


Figure 6.4 **(a) Schematic of a molecular junction from the self-assembly of three peptoids and (b) the MALDI mass spectrum of a reaction mixture attempting the assembly.**

6.2.3. Dynamic Covalent Scaffolding

As the information-directed assemblies of strands encoded by residue sequence has found limited success, another process for shape formation was considered inspired by Vernier templating and DNA origami scaffolding. Vernier templating uses the length of strands to form a higher ordered templated species, much like DNA origami uses small scaffolds to fill a larger strand and fix it into place. Here, we propose using oligomers of varying length to impose or scaffold a shape onto a larger macrocyclic strand. The large strand can be a peptoid to afford flexibility, while a stiffer shorter strand is used to provide a rigid shape. For example, a large peptoid bearing 24 reactive sites could form a tight oval or line with two strands with 12 functional sites, a square with four hexamers, triangle with three octamers, hexagon with eight trimers, or a circular shape with 12 dimers.

Large macrocycles are fairly difficult to build, as precursors typically co-react rather than cyclizing. Several chemistries were tested to determine the reaction with the greatest chance of macrocyclization. Copper azide-alkyne click chemistries were too expensive and difficult to perform on the large scale required for self-assembly experiments, while thiol-ene chemistries were largely incompatible with synthesis on a single oligomer. Fortunately, incorporation of glutamic or aspartic acid as the first residue on photolabile resin, and leaving the terminal end as a primary amine enables the on-resin formation of a macrocycle under dilute amide coupling conditions upon cleavage of the glutamic/aspartic acid protecting groups. Furthermore, an alloc-amine peptoid monomer could be used on the same strand and deprotected post cyclization to afford a large, functionalized macrocycle. The corresponding aldehyde scaffold oligomer is still under consideration as a relatively stiff backbone is required.



Scheme 6.1 Macrocycle scaffolding to assemble shapes dependent on scaffold length.

6.3 Peptide-antibody Bioconjugates Future Directions

At the moment, peptide bioconjugation to the red blood cell-binding antibodies requires first the addition of am NHS-ester-maleimide linker to the antibody, followed by the thiol-Michael addition of the cysteine-containing peptide to the maleimide. This process was found to be inefficient, requiring a substantial excess of valuable and somewhat-expensive peptide and resulting in non-ideal yields to be commercially viable. As such, we are currently developing a peptide sequence that incorporates the maleimide monomer from Chapter 4 enabling direct coupling of the peptide to the cysteines available from the reduction of the antibody.

6.4 Importance

The ability to direct the self-assembly of oligomeric strands based on their residue sequence and mediated by dynamic covalent interactions as demonstrated here is a crucial step towards the fabrication of complex, unimolecular constructs while mitigating the kinetic trapping of non-equilibrium species that often prevails in dynamic covalent assembly systems. Although this dissertation involved molecular ladders bearing covalent rungs, this multi-step approach to dynamic covalent assembly process may also be useful for other applications in which the alleviation or elimination of kinetic trapping is critical. Specifically, this process will provide significantly improved synthetic access to robust, complex covalent nanostructures such as molecular cages and crystalline, porous polymer networks.

6.5 References

1. Seeman, N. C., Nucleic Acid Junctions and Lattices. *J. Theor. Biol.* **1982,** *99* (2), 237-247.