Click Chemistry on DNA and Targeting RNA structure with Peptide Boronic Acids

Jason Brian Crumpton

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Chemistry

Webster L. Santos, Chair

Felicia A. Etzkorn

Paul R. Carlier

Daniel G. S. Capelluto

(04/20/2012)

Blacksburg, VA

Keywords: MALDI-MS, peptide sequencing, matrix, DNA, click chemistry

Click Chemistry on DNA and Targeting RNA structure with Peptide Boronic Acids

Jason Brian Crumpton

ABSTRACT

The utilization of click chemistry to perform inter- and intramolecular ligation on DNA has become ubiquitous in the literature. Advances in copper (I) stabilizing ligands that prevent DNA degradation via redox pathways have provided nucleic acid researchers access to the efficiency and quantitative nature of the click reaction. The majority of ligation procedures in the literature are performed in solution after DNA assembly and modification with alkyne reporter groups. However, without specialty alkyne reagents that can be sequentially and selectively deprotected, the solution phase method requires that the click reaction be performed on all DNA-attached alkynes simultaneously. Therefore, the variability of the azide reagent is limited to a singular R group. However, performing the click reaction on DNA during synthetic elongation (immediately after each alkyne installation) allows for the possibility of performing multiple click reactions with variable azide reagents. Unfortunately, most solid phase click procedures require long reaction times or the utilization of microwave irradiation to accelerate the reaction. The development of methods for the ligation of azides to alkynes without the use of microwave irradiation on solid phase is potentially very useful. Herein, we report a simple, efficient, and robust solid phase synthetic method for the ligation of azido-diamondoids to the alkyne-modified phosphate backbone of DNA with click chemistry using $[Cu(CH_3CN)_4]PF_6$ without stabilizing ligand. Interestingly, it was found that as the size of diamondoid increased, a corresponding increase in melting temperature of hybridized duplexes was observed. The developed method has the potential to

complement existing DNA ligation procedures for applications in biotechnology and diagnostics.

Interest in peptides incorporating boronic acid moieties is increasing due to their potential as therapeutics/diagnostics for a variety of diseases such as cancer. The utility of peptide boronic acids may be expanded with access to vast libraries that can be deconvoluted rapidly and economically. Unfortunately, current detection protocols using mass spectrometry are laborious and confounded by boronic acid trimerization, which requires time consuming analysis of dehydration products. These issues are exacerbated when the peptide sequence is unknown, as with *de novo* sequencing, and especially when multiple boronic acid moieties are present. Thus, a rapid, reliable and simple method for peptide identification is of utmost importance. Herein, we report the identification and sequencing of linear and branched peptide boronic acids containing up to five boronic acid groups by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). Protocols for preparation of pinacol boronic esters were adapted for efficient MALDI analysis of peptides. Additionally, a novel peptide boronic acid detection strategy was developed in which 2,5-dihydroxybenzoic acid (DHB) served as both matrix and derivatizing agent in a convenient, *in situ*, on-plate esterification. Finally, we demonstrate that DHB-modified peptide boronic acids from a single bead can be analyzed by MALDI-MSMS analysis, validating our approach for the identification and sequencing of branched peptide boronic acid libraries.

It is well known that RNA ligands incorporating basic and intercalating moieties display high RNA affinity. Unfortunately, these ligands are also often plagued by promiscuous binding to off-target substrates. Due to the potential utility of RNA ligands

in biology and medicine, it is imperative to elucidate RNA binders which display high specificity as well as affinity. Boronic acid peptides promise unique RNA binding motifs through the interaction between the empty p-orbital of boron and the 2'-hydroxyl group of RNA. Herein, we describe the incorporation of lysine and phenylalanine boronic acid analogues into a branched peptide combinatorial library in an effort to impart increased selectivity towards the HIV-1 Rev Response Element (RRE). We were able to easily select and deconvolute 6 resulting "hit" peptides from 65,536 unique library members by high throughput screening and *de novo* sequencing. Although we were unable to evaluate peptide selectivity towards RRE due to general insolubility in aqueous media, we demonstrated the efficient deconvolution of a branched peptide library that incorporates boronic acids.

ACKNOWLEDGEMENTS

I would like to start by thanking my committee members for guiding me throughout my time here at Tech. Thank you for taking the time to read and revise the inarticulateness of the documents I've submitted to you, for testing my knowledge, and for giving honest feedback on my strengths as well as my shortcomings.

I would also like to thank my advisor Dr. Webster Santos for his guidance over the years. Dr. Santos is the very antithesis of the old adage that nice guys finish last. I don't think I will ever have a fairer, more understanding, and patient yet demanding boss. I cannot count the times that I was leaving the lab in the evening, only to find Webster hard at work in his office. Thank you for always setting a good example for the rest of your group through your work-ethic, your love of science, your knowledge, and, at the end of the day, your ability to just relax and go have a beer.

My thanks go out to my talented friends and colleagues in the Santos lab, Jing Sun, Michael Perfetti, Wenyu Zhang, Xi Guo, Amanda Nelson, Jessica Wynn, Molly Congdon, Joe Calderone, and Emily Morris for being such wonderful, hard-working, and interesting people. I hope that you'll always be proud of your decision to push beyond the comfort and safety of the known and venture out into that wild blue yonder. In the spirit of acknowledging excellence, I would like to thank Dr. Mithun Raje and Dr. Brandon Thorpe for always setting the bar particularly high in our lab. The breadth of your knowledge, skill, confidence, and rigorousness that you applied to your work and towards science in general were always inspirational to me.

I would also like to single out Dr. David Bryson and Ken Knott of the Santos lab:

- David, I will always be indebted to you for allowing me to talk freely with you about anything and for making me feel that my opinion was always of value. You are one of the most giving, gifted, and patient people that I have ever had the pleasure of knowing, and you constantly inspire me to be a better person and scientist.

- Ken, I want to thank you for always being there for me when the chips were down. You always had a knack for knowing when I needed an ear to bend, and you never hesitated to try to help me pick myself back up after I stumbled and fell along my path. Your humor, honesty, and zest for life (and coffee) have left me with a good number of striking memories.

I want to thank Dr. Samantha Higgins for her friendship, advice, and encouragement along the way; your words helped much more than you realize.

I am extremely grateful to Jon Bernard for his instrumental help in the construction of software that has proven essential for the preservation of the sanity of everyone associated with the HIV project in the Santos Lab. Jon, you have been a great friend, climbing buddy, and collaborator during my time here at Tech; please do not lose touch.

My heartfelt thanks goes out to Dr. Keith Ray for being such an influential figure in my scientific career here at Tech. Without your help and patience, my life probably would have been steered in a much different direction.

Thank you J.P. Gannon, Amy Snyder, Jenn Chang, Adam Walker, Joe Rollin, Elin Arnaudin, Mike Swedo, James Wilson, and Chris Helak for your friendship and for all of

the memories. My apologies go out to all of the other delightful people I have had the pleasure of getting to know during my time in Blacksburg whom I did not mention in this short list.

My utmost love and thanks go out to my amazing parents, Jamey and Billy, as well as to the rest of my family. I would never be where I am today if not for your love, support, and encouragement in everything that I have ever decided to attempt. I surely would have failed long ago had it not been for you all.

- Mom, I want to thank you for your unfailing love and for always setting an example for what it means to be a good person. You are one of the strongest people that I know, and you have helped me understand the true meaning of inner strength and resilience throughout the trials and tribulations of life; I love you.

- Dad, I want to thank you for making sure that I had a strong role-model while I was growing up. You taught me that happiness is something that you have to actively cultivate and like everything else worthwhile in life, takes effort and comes with a price; I love you.

- To my Dot-Dot and Oopop, I want to thank you for being such amazing people. You both always selflessly gave much more than you took, and you never asked of anything other than for your children and grandchildren to be happy. You set an example for how to live life fully and gracefully that I can only aspire to emulate.

- To my grandmother Clare, I want to thank you for your love and for the memories of our time together at the beach when I was younger. Your smile and laugh

are exactly the same now, and just as beautiful, as the way I remember them when I was younger.

Finally, I extend my love and thanks to two very special women that I've had the pleasure of having as a part of my life:

- To Laura Hamm, you were and continue to be one of my oldest and closest friends. I want to thank you for reminding me, through your words and example, to try not to take myself too seriously and to always have a little fun along the way. You taught me so much about always trying to see the best in other people as well as in myself. You have always been there for me, whether it be at the heights of my joy or the depths of my despair. I will always be indebted to you for making me feel like I was worth your love, your forgiveness, and your friendship. Thank you for the memories; I'll cherish them always.

- To Stephanie McGinness, I want to thank you for coming into my life so unexpectedly. Even though our time together was short, you helped me remember what it was like to truly and undeniably love someone. It was without a doubt one of the happiest periods in my life. You taught me just how important it is to never take my blessings for granted. Most importantly, you helped me understand that if I'm ever lucky enough to know exactly what I want in life (which for me has always been a challenge), to never settle for less than that.

DEDICATION

This manuscript is dedicated to the notion that happiness does indeed come from within.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Abbreviations

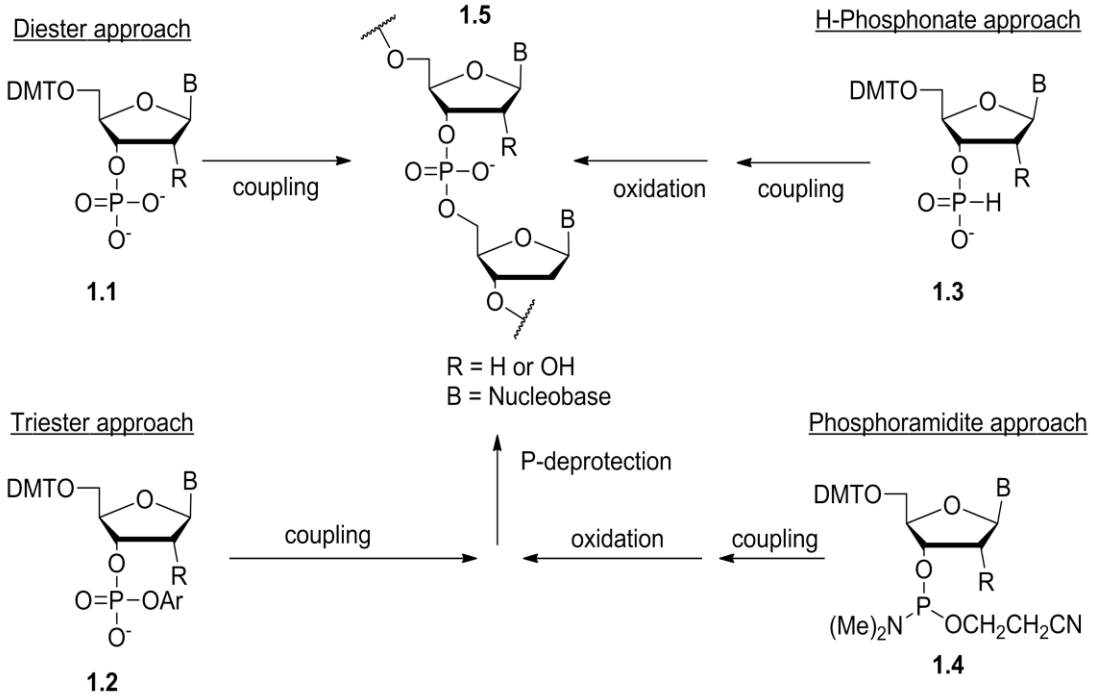
μCP = microcontact printing
AC = ammonium citrate dibasic
AChE = acetylcholinesterase
ACN = acetonitrile
ANP = 3-amino-3-(2-nitrophenyl) propionic acid
ARM = arginine rich motif
ART = anti-retroviral therapy
AZT = 3'-azido-2',3'-dideoxythymidine
BB = branched boronic acid
BCP = backbone cyclic peptide
BIV = bovine immunodeficiency virus
CD = circular dichroism
CHCA = α-cyano-4-hydroxycinnamic acid
CHIQ = 1-chloro-4-hydroxyisoquinoline
CI = chemical ionization
CMBT = 5-chloro-2-mercaptobenzothiazole
COPAS = complex object parametric analyzer and sorter
CPG = controlled pore glass
CRI = co-receptor inhibitor
CuAAC = Cu (I) catalyzed [3+2] azide-alkyne cycloaddition
DABP = 3,4-diaminobenzophenone
DBU = 1,8-diazabicyclo [5.4.0] undec-7-ene
DHAP = 2',6'-dihydroxyacetophenone
DHB = 2,5-dihydroxybenzoic acid
DIEA = *N*,*N*-diisopropylethylamine
DIP/EI = direct insertion probe electron ionization
DMT = 4,4'-dimethoxytrityl
ECCI = electron capture chemical ionization
EDTA = ethylenediaminetetraacetic acid
EMSA = electrophoretic mobility shift assay
ER = endoplasmic reticulum
ESI = electrospray ionization
ETT = 5-ethylthio-1H-tetrazole
FAB = fast atom bombardment
FAM = 6-carboxyfluorescein
F_{BPA} = 4-borono-L-phenylalanine
FI = fusion inhibitor
FIGS = fusion-induced gene stimulation
FITC = fluorescein isothiocyanate
HAART = highly active antiretroviral therapy
HBV = hepatitis B virus
HCTU = 2-(6-chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HPA = 3-hydroxypicolinic acid

HPLC = high performance liquid chromatography
HPTA = *tris*-(hydroxypropyltriazolylmethyl)amine
IN = integrase
IR = integrase inhibitor
K_{BBA} = N-ε-(4-boronobenzoyl)-L-lysine
LSI = liquid secondary ionization
LTR = long terminal repeat
MALDI = matrix assisted laser desorption ionization
MBO = multivalent binding oligomers
MQ = Milli-Q
MW = microwave
NB = no binding
NNRTI = non-nucleoside reverse transcriptase inhibitor
NRTI = nucleoside/nucleotide reverse transcriptase inhibitor
OBOC = one bead one compound
ODN = oligodeoxynucleotide
ON = oligonucleotide
PAGE = polyacrylamide gel electrophoresis
PBL = peptide boronolectin
PCR = polymerase chain reaction
PI = protease inhibitor
Piv-Cl = pivaloyl chloride
PS-ODN = phosphorothioate oligodeoxynucleotide
PTSA = p-toluenesulfonic acid
RRE = rev response element
RT = reverse transcriptase
SPPS = solid phase peptide synthesis
SU = surface
TAR = trans-activation response element
TBDMS = *tert*-butyldimethylsilyl
TBTA = *tris*-(benzyltriazolylmethyl)amine
TCA = trichloroacetic acid
TEA = triethylamine
TEAA = triethylammonium acetate
THAP = 2',4',6'-trihydroxyacetophenone monohydrate
^{TL}DNA = triazole linked DNA
TM = transmembrane

Chapter 1 DNA synthesis and post-synthetic modification

1.1 P(III) and P(V) chemistry

In order to synthesize oligonucleotides (ONs) with natural or unnatural features, there are several methodologies that the synthetic chemist can take advantage of. The synthesis of ON polymers is fundamentally tied to the formation of the phosphate backbone. Therefore, the scientific community has put forth a significant amount of time and effort towards the optimization of the requisite activation and coupling of nucleotide building blocks. Since the earliest examples of ON synthesis, this process has been dictated by either P(III) or P(V) chemistries (Scheme 1.1). The earliest methodologies for artificial DNA synthesis were pioneered and dominated by P(V) chemistries such as the diester and triester approaches.[1] The use of P(V) precursors such as **1.1** and **1.2** were



Scheme 1.1 Approaches for forming phosphate backbones

preferred due to the fact that they were less prone to oxidative degradation, and therefore easier to handle than their P(III) counterparts. However, by the early 1970's, it started to become more apparent that methodologies based on P(III) derivatives such as the H-phosphonate (**1.3**) and phosphoramidite (**1.4**) approaches also had the potential for great synthetic utility. Indeed, the phosphoramidite approach is currently the most common method for ON synthesis on solid support. [2]

Phosphorus compounds utilized can be subdivided into tetracoordinate P(V) (**1.6**), tricoordinate P(III) (**1.7**), and tetracoordinate P(III) compounds (**1.8**). In terms of utility, each has its own specific advantages and disadvantages.

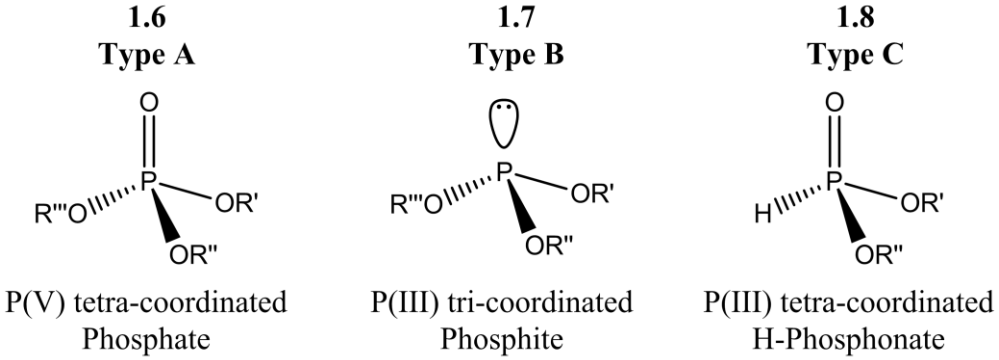
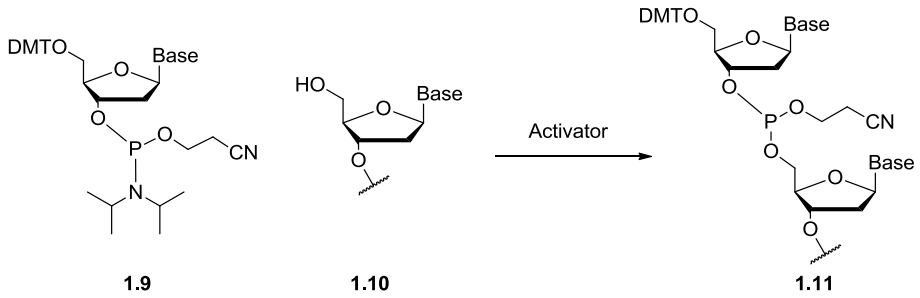


Figure 1.1 Phosphorus building blocks for ON synthesis

Type **A** phosphorus compounds are characterized by a tetrahedral geometry with a very stable phosphoryl group. Due to phosphorus's high oxidation state, it is exceptionally stable during storage and convenient to handle during organic transformations. Unfortunately, the utility of type **A** phosphoryl groups is hindered by lower synthetic transformation efficiency than type **B** or **C** building blocks, even after activation with condensing agents.

Type **B** phosphite compounds are characterized by a trigonal pyramidal geometry with a lone electron pair on the phosphorus to yield a slightly basic soft nucleophile.

However, when the phosphorus is attached to appropriate activating groups, these compounds can also react with various nucleophiles, such as the 5'-hydroxyl of nucleoside **1.10** to yield P(III) ON **1.11** (Scheme 1.2). Unfortunately, type **B** phosphites can be difficult to handle and, due to their low oxidation state, are susceptible to spontaneous oxidation and hydrolysis upon storage or handling.[3]



1.9 **1.10** **1.11**

Scheme 1.2 ON dimer (**1.11**) synthesis via phosphoramidite (**1.9**) activation/coupling

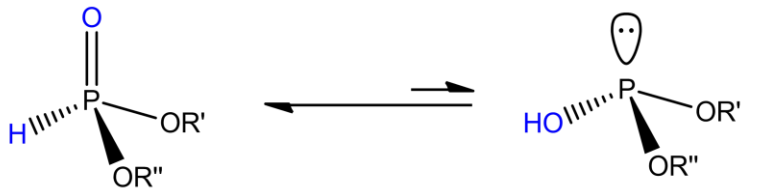
Type **C** phosphorus compounds possess a tetrahedral geometry with a phosphoryl group. Unlike type **B**, they lack a lone pair of electrons on phosphorus. One distinguishing characteristic of this class of compounds is the P-H bond that makes type **C** structurally quite similar to type **A** compounds (tetrahedral with no lone pair of electrons). Therefore, they have been found to be more stable and less prone to spontaneous oxidation under storage than their type **B** counterparts. Additionally, the type **C** H-phosphonates are unique in that they are capable of numerous chemical transformations (See **1.2.3**) that are inaccessible when using type **A** or **B** phosphorus building blocks.

1.2 H-Phosphonate chemistry

1.2.1 H-phosphonate stability and reactivity

In terms of ON synthetic chemistry, the ideal situation would be for a phosphorus compound to be stable on the shelf, convenient to manipulate, and highly reactive when

placed in a reaction flask. Indeed, all of these needs are met with H-phosphonate compounds. Type **C** phosphorus has been shown to be stable to silica gel chromatography.[4] Their coupling rates are comparable to phosphoramidite derivatives (< 5 min), and they are amenable to a variety of condensing agents.[4,5] Furthermore, there is no need to install phosphate protecting groups which are requisite for phosphoramidite compounds (**1.9**). This is due to phosphonate-phosphite equilibrium in solution which favors the phosphonate tautomer, shown in Figure 1.2.[6] The phosphonate tautomer **1.12** possesses a hydrogen atom that can be considered the smallest possible effective protecting group.



P(III) tetra-coordinated **1.12** P(III) tri-coordinated **1.13**

Figure 1.2 Tautomeric equilibrium in H-phosphonate compounds

1.2.2 Preparation of H-phosphonate monoesters

There are several simple, general, and efficient methods for the preparation of H-phosphonate monoesters for DNA synthesis. The salts of the monoesters are stable solids that are resistant to air oxidation as well as to hydrolysis.[5] In general, commercially available reagents such as **1.14-1.16** can form the nucleoside H-phosphonate **1.18** in high yields (70-95%) from **1.17** (Figure 1.3).[4,7,8] Additionally, these transformations can be completed in relatively short time frames (~ 3 hr) using mild reaction conditions at room temperature with pyridine/acetonitrile mixtures as solvent.

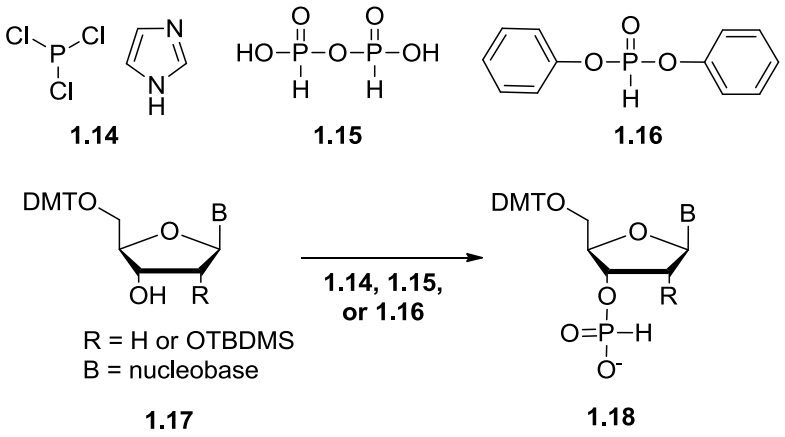
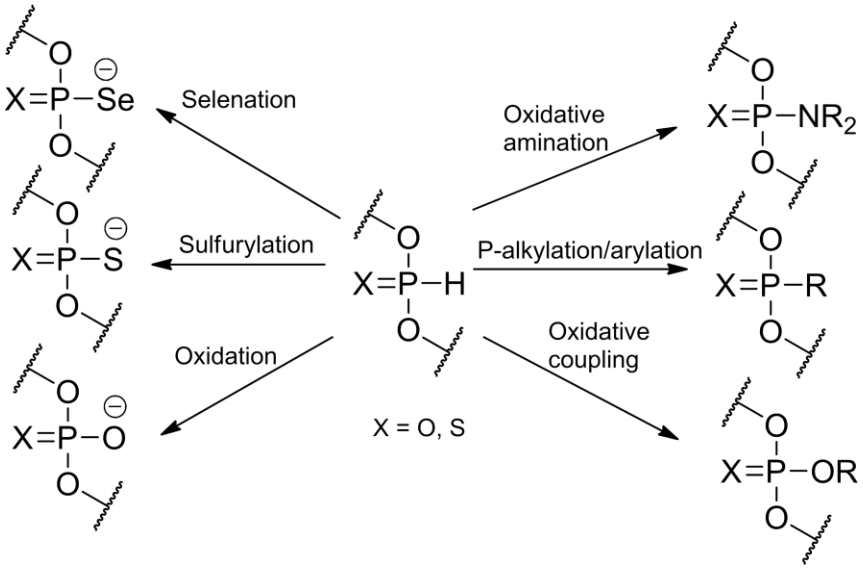


Figure 1.3 Synthesis of H-phosphonate monoesters

1.2.3 Oxidative transformations of H-phosphonate diesters

Once the H-phosphonate monoester is successfully formed and coupled to a growing ON polymer, an extensive number of oxidative transformations can be performed as shown in Scheme 1.3.

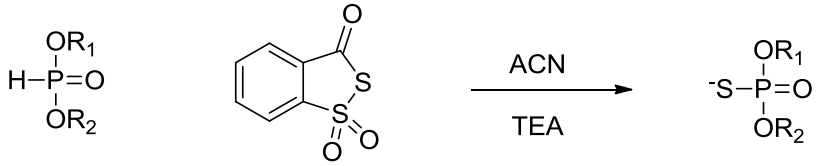


Scheme 1.3 Oxidative transformation of H-phosphonate diesters

Oxidative amination yields the P-N functionality which has led to a number of interesting compounds. Letsinger and colleagues initially explored cationic phosphoramidates in the late 80's and demonstrated that under appropriate conditions

(low pH; low salt concentration) the cationic probes bind more effectively than their natural counterparts.[9] It has also been shown that it is possible to synthesize nucleoside phosphoramidate libraries via treatment of H-phosphonates with acyclic, cyclic, heterocyclic, and aromatic amines. A representative 600 member library was assembled and screened for antiviral activity against the hepatitis B virus (HBV) with some members displaying potent antiviral activity.[10]

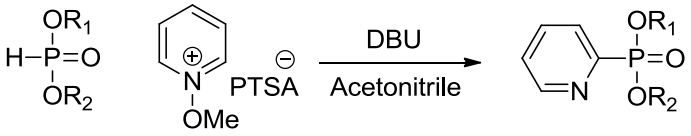
Oxidative sulfurylation and selenation have been demonstrated in the literature. The development of sulfurylation compounds such as the Beaucage reagent (3H-1,2-benzodithiol-3-one-1,1-dioxide) 1.20 (Scheme 1.4) has led to the development of one of the most successful classes of modified oligodeoxynucleotides (ODNs) utilized for gene therapy.[11,12] Indeed, the phosphorothioate ODN (PS-ODN) Vitravene is the only antisense drug that has been approved by the FDA to treat cytomegalovirus-induced retinitis in AIDS patients.[13] These types of ODN analogs have demonstrated the ability to form stable duplexes by Watson-Crick base pairs, acceptable stability against nucleases, and activation of RNase H. Unfortunately, these PS-ODNs display high affinity for certain proteins, and have been shown to cause cellular toxicity.[14] Oxidation to the phosphoroselenoate diesters has also been accomplished using transformation reagents similar to the Beaucage reagent, although practical applications of these diesters have been minimal.



1.19 **1.20** **1.21**

Scheme 1.4 Sulfurylation of H-Phosphonates[12]

P-alkyl and P-aryl functionalities have also been successfully installed using H-phosphonates as starting materials. Stawinski and colleagues have demonstrated that C-phosphonate pyridylphosphonate derivatives (**1.23**) can be obtained rapidly from H-phosphonate diesters in high yields (> 90%).[15] These P-C bond formations (Scheme 1.5) also make this class of phosphate esters resistant to enzymatic hydrolysis, making them interesting for antisense/antigene agents.



1.19 **1.22** **1.23**

Scheme 1.5 Arylation of of H-Phosphonates

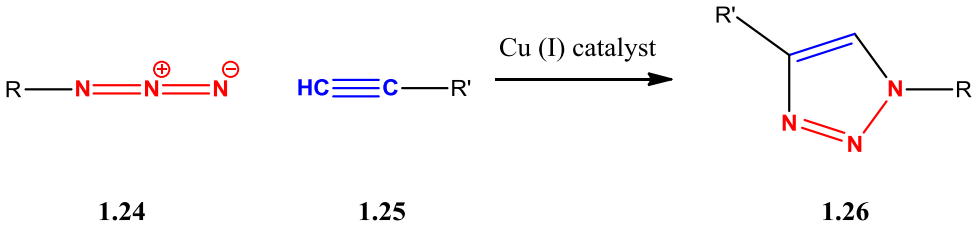
This section has presented a general synopsis of methodologies for ON synthesis which was then followed with information regarding H-Phosphonate utilization for construction of ONs with interesting structural motifs.

1.3 Click chemistry

1.3.1 Criteria and early development

Perhaps the best known and most exploited "click" reaction is the Cu(I) catalyzed [3+2] azide-alkyne cycloaddition (CuAAC), which was developed independently by the Sharpless and Meldal labs (Figure 1.4).[16,17] In order for chemical processes to be classified under this category, a stringent set of criteria must be met in regards to the reaction's efficiency, conditions, and purification requirements. Click reactions must be wide in scope, give high yields with stereospecific products, and generate only inoffensive byproducts. Reaction conditions must utilize readily available starting materials, none or benign solvent, and above all else, simple reaction conditions. Finally,

if any purification is to be performed, it must be by non-chromatographic methods and the final product has to be stable under physiological conditions.



1.24 **1.25** **1.26**

Figure 1.4 Generic CuAAC reaction

With such a demanding checklist, it seems amazing that any chemical transformation could be classified as a "click" reaction. However, several general types of reactions conform to the click checklist and are able to do so in general because of a high thermodynamic driving force usually greater than 20 kcal·mol^{-1}.

Several interesting advances with Click chemistry in the form of the Huisgen 1,3-cycloaddition have been demonstrated. One of the first steps forward was the development of a copper (I) catalyst, which was found to drastically accelerate triazole formation with high regioselectivity of the 1,4- over 1,5- product.[18] Another was the application of Click Chemistry by the Sharpless group to select for a femtomolar inhibitor of acetylcholinesterase (AChE).[19] This study provided an unprecedented example of a target enzyme synthesizing its own inhibitor, as well as a demonstrated ability of AChE to preferentially select between the 1,4- and 1,5- products.

An impressive extension to this preliminary study was the demonstration by the Bertozzi group of Cu-free click chemistry using strained hydrophilic azacyclooctynes.[20] These cyclooctynes were attached to biotin for biochemical binding assays and were shown to bind effectively to cells incorporating the corresponding azides with no recognizable cytotoxicity.

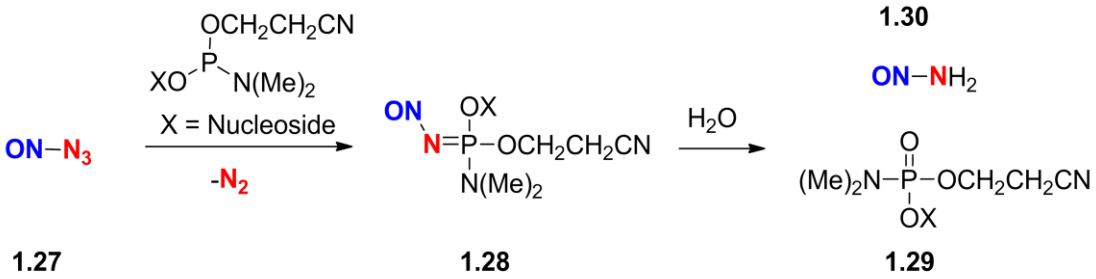
1.3.2 DNA modification by Click Chemistry

In addition to the general appeal of click chemistry, the CuAAC reaction has shown great potential in the nucleic acids fields for several reasons. First, azides and alkynes can be attached to nucleic acids without greatly disturbing biophysical properties. Secondly, both of these moieties are almost entirely unreactive to functional groups found in nature. Finally, the resulting triazole is extremely stable and non-toxic.[21]

When copper is not used as a catalyst, the CuAAC reaction is referred to simply as the Huisgen 1,3-dipolar cycloaddition. One of the earliest applications of this copper free cycloaddition in the nucleic acids field was by performed by Ju *et al* to attach an alkynyl 6-carboxyfluorescein (FAM) fluorophore to an azido-labeled single-stranded (ss) DNA.[22] This pioneering study demonstrated that this ligation strategy could be utilized to synthesize fluorescent oligonucleotides (ON's) with high selectivity, yield, and purity. Unfortunately, this initial study required high temperatures (80 ^{o}C), extended reaction times (3 days), and produced a mixture of 1,4- and 1,5- regioisomers.

The initial report by Ju and colleagues demonstrated the general appeal to the nucleic acids field. However, there were several incompatibilities of the CuAAC reaction to ON chemistry that first needed to be addressed. One of the most unfortunate of these was the incompatibility of azides with traditional ON synthesis. It is well known that azides react with P(III) reagents via the Staudinger reaction.[23] This is problematic because the major coupling step for ON synthesis takes place via P(III) (Phosphoramidite / H-Phosphonate) chemistry. If the azide functionality is already present within the growing ON structure (**1.27**), it is efficiently reduced to the corresponding amine (**1.30**) upon the subsequent coupling step. (Scheme 1.6) Consequently, the general methodology

has been reversed to modifying ON's with alkynes followed by subsequent conjugation with azide-functionalized groups.



Scheme 1.6 Reduction of azide to amine via Staudinger reduction

A second, and perhaps more unfavorable incompatibility of CuAAC chemistry with ON chemistry is DNA/RNA's rapid degradation in the presence of copper. Indeed, Cu(I) is known to damage DNA by promoting radical-mediated processes.[24] Additional studies have also reported an extremely short half-life (~10 min) of DNA under Cu(I)-conditions.[25] The solution to this problem was presented by the Sharpless group in a study reporting the construction of Cu^{I} stabilizing ligands.[26] Fortunately, several ligands such as (*tris*-(benzyltriazolylmethyl)amine) TBTA and *tris*-(hydroxypropyltriazolylmethyl)amine (HPTA) (Figure 1.5) were found to be quite efficient at accelerating the cyclization reaction itself as well as abolishing redox processes which led to ON degradation. These ligands paved the way towards the routine modification of ON's via the CuAAC pathway.

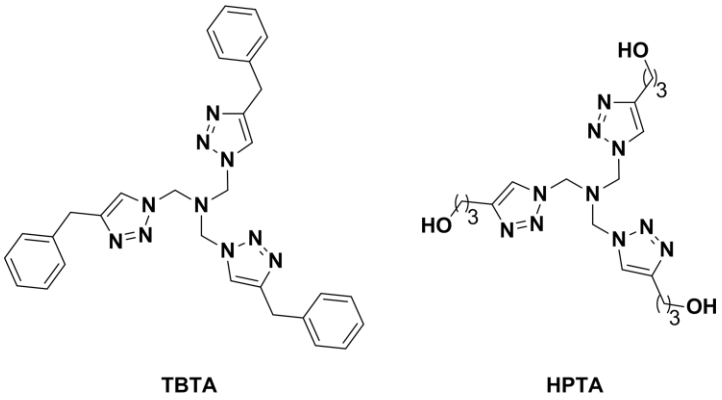


Figure 1.5 Cu(I) Stabilizing ligands

1.3.3 Post-synthetic modification of DNA

Post-synthetic functionalization (*e.g.*, attachment of fluorescent tags) of small reporter groups incorporated into DNA is very appealing for life science research as well as the diagnosis of genetic disorders. This type of post-synthetic approach is required because enzymatic incorporation of fluorescently tagged nucleoside building blocks into long genetic sequences is quite challenging.[27] Unfortunately, this limitation requires that the post-synthetic functionalization be highly efficient and specific.

To explore whether the CuAAC reaction was suited for this task, Carell and colleagues synthesized a series of 16-mer ODNs with ethynyl (**1.31**) as well as octadiynyl (**1.32**) uracils (Figure 1.6).[28] The octadiynyl uracil was incorporated in the event that sterics interrupted high density labeling on the ethynyl moiety. It was determined that cyclization efficiency of the azide tags (**1.33-1.35**) with the short ethynyl unit was unacceptable with only 6 adjacent units, presumably due to steric blockade (Figure 1.7). However, the less restricted octadiynyl functionalized ODNs displayed quantitative conversion even with 6 adjacent units. As a further testament to the efficiency of this

reaction, quantitative cyclization was obtained even on PCR fragments up to 2000 bases that were constructed from primers containing two click sites.

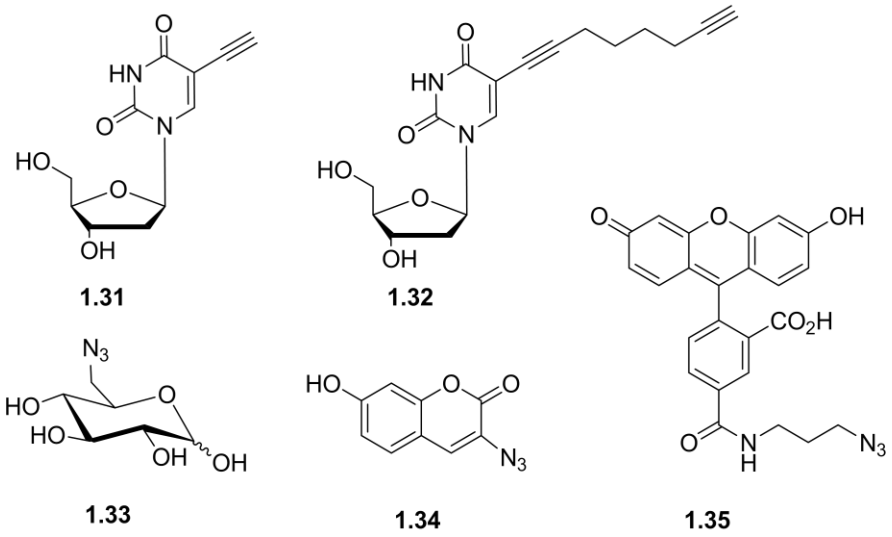


Figure 1.6 Alkyne modified uracils and azide reporter groups

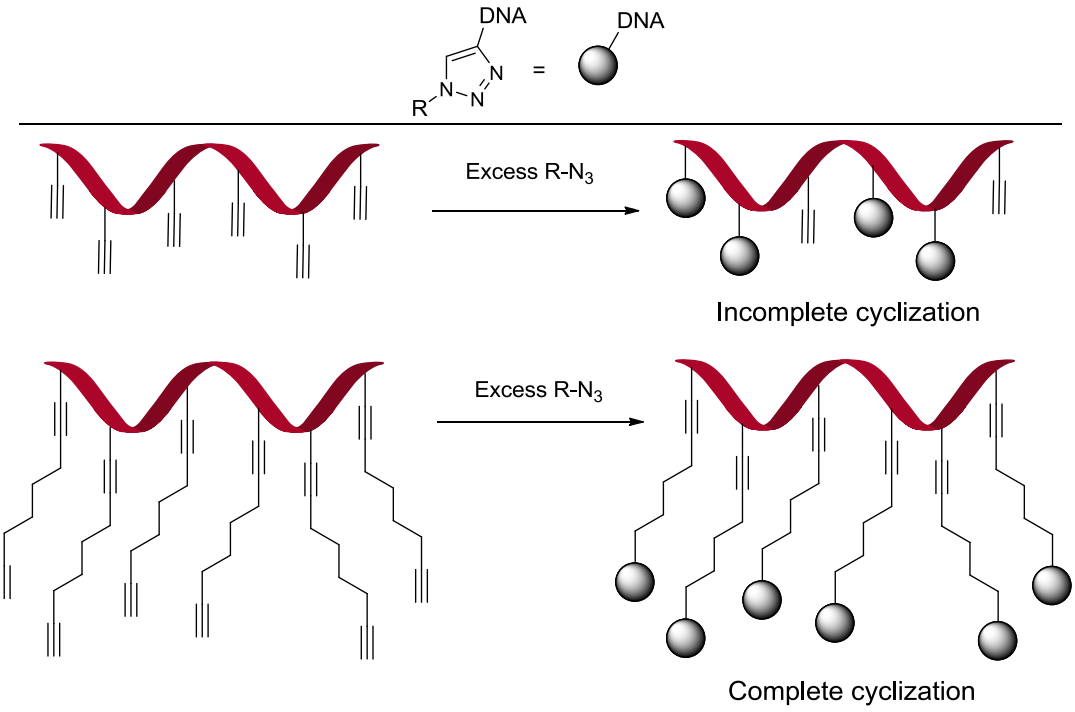


Figure 1.7 Cyclization efficiency of ethynyl (1.31) vs. octadiynyl uracils (1.32)

Seela and colleagues have also reported the functionalization of U, C, G, and A nucleobases with octadiynyl side chains from the corresponding iodo-nucleosides by Sonogashira cross coupling (Figure 1.8).[29,30] These nucleosides were converted to the phosphoramidite building blocks, and a series of oligonucleotides were prepared by solid-phase synthesis. It was determined that the octadiynyl resulted in a 1-2 ^{o}C T_m increase per modification when incorporated in the 5-position of 2'-deoxyuridine (**1.36b**) and a 2-3 ^{o}C T_m increase when incorporated in the 5-position of 2'-deoxycytidine (**1.37b**) or in the 7-position of 7-deaza-2'-deoxyguanosine (**1.39b**)

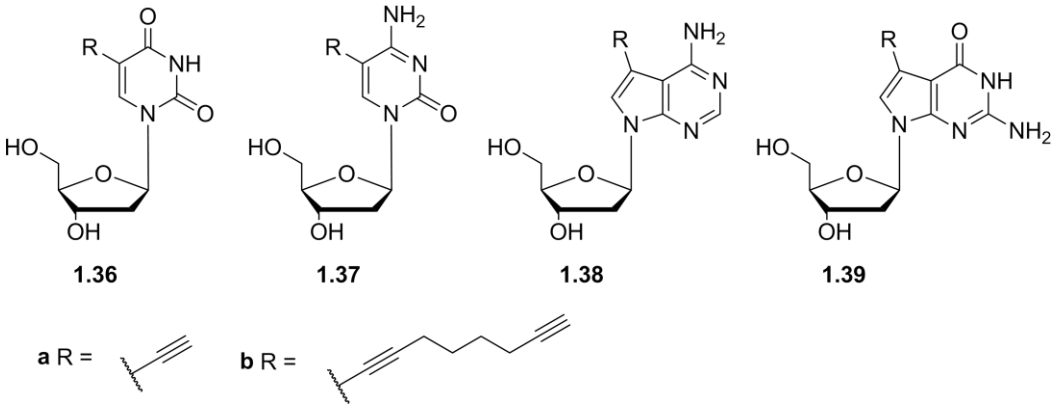
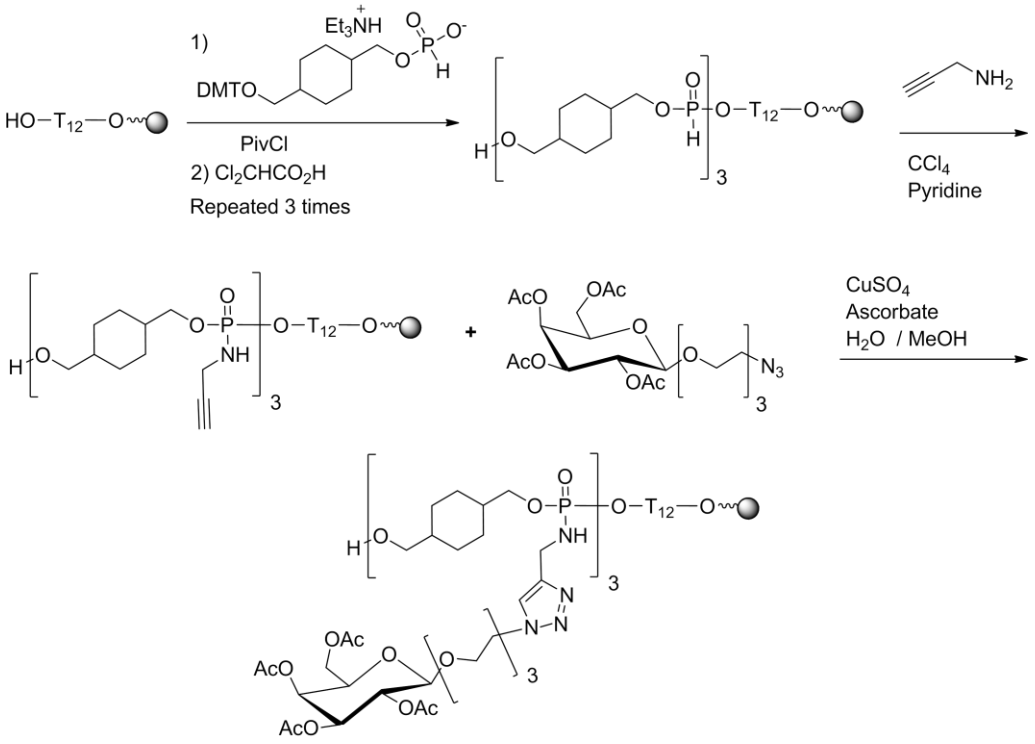


Figure 1.8 Alkyne functionalized U, C, G, and A nucleobases

In 2006, Morvan reported the utilization of microwave promoted click chemistry to attach carbohydrates to solid-supported oligonucleotides.[31] Three H-phosphonate diester linkages were introduced into a cyanoethyl-protected dodecathymidine (T_{12}) followed by oxidation with CCl_4/propargyl amine to yield the alkyne-functionalized oligonucleotide (Scheme 1.7). Quantitative conversion of the tri-alkyne ODN to the carbohydrate-labeled ODN was accomplished with microwave irradiation at 60^{o}C within 20 minutes. This study demonstrated that rapid and high-yielding multiple cycloadditions could be promoted by microwave irradiation without hydrolysis of phosphoramidate

bonds. This method has since been extended to orthogonal installation of mannose as well as galactose residues onto an ODN structure.[32]



Scheme 1.7 Microwave promoted click chemistry ligation of sugars[31]

Due to bio-incompatibility of Cu(I), methods to perform the azide/alkyne click reaction in efficient and quantitative fashion are highly desirable. Interestingly, it has been demonstrated that the click reaction is exquisitely suited for the purposes of microcontact printing (μCP).[33] This printing technique facilitates chemical synthesis between surface immobilized substrates and activated reagents in solution. In this study, Reinhoudt and colleagues demonstrated that the Huisgen 1,3-dipolar cycloaddition could be completed in only 15 min without any catalyst and only 35 $g \cdot cm^{-2}$ of pressure applied to the stamping apparatus. This concept was then in turn successfully applied to μCP immobilization of ON's onto an azide functionalized solid surface (Figure 1.9).[34] It was

then demonstrated that these immobilized ON probes were able to efficiently hybridize with the fluorescently tagged complementary strand.

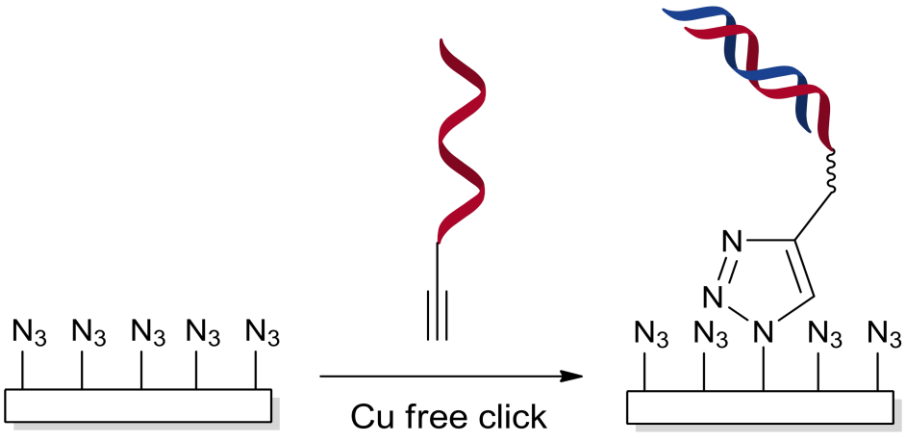


Figure 1.9 Immobilization and effective DNA hybridization

1.3.4 Artificial nucleobases and phosphate backbones

Attempts to modify the nucleobases of ON's is attracting much attention to increase base discrimination and to enhance the stability of duplex or triplex nucleic acids. Obika and colleagues reported the modification of 1-ethynyl-2-deoxy-B-D-ribofuranose (**1.40**) followed by conversion into various 1,2,3-triazoles to produce novel nucleobase analogs.[35] Unfortunately, of the 8 ONs synthesized that incorporated a novel nucleobase, all were found to display lower duplex stability than the native duplex. However, the artificial triazole **1.41** performed as a universal nucleobase (Figure 1.10).

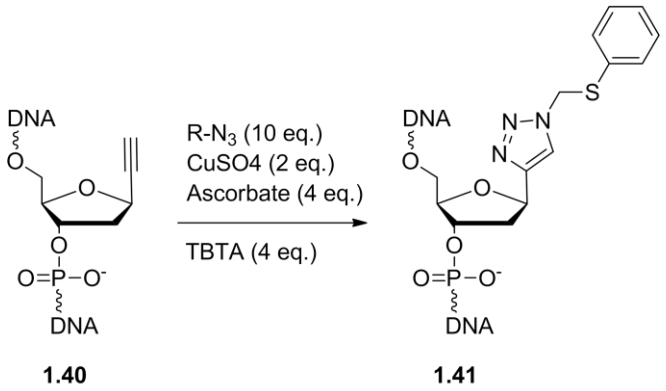


Figure 1.10 Synthesis of artificial triazole nucleobase[35]

Unfortunately, natural oligonucleotides are poor candidates for biological applications such as antisense gene therapy due to low stability towards cellular nucleases and poor cell penetration. It was hypothesized that nucleotides containing the triazole moiety would possess high stability toward chemical and enzymatic degradation and would therefore be interesting antisense candidates. In an effort to develop a new route to novel nucleotide analogues, Dondoni *et al.* demonstrated the approach of replacing natural phosphate backbones with triazole internucleoside linkages.[36] They were able to demonstrate the viability of this concept by synthesizing an artificial tri-nucleoside via solution phase CuAAC ligation of thymidine residues (Figure 1.11).[36]

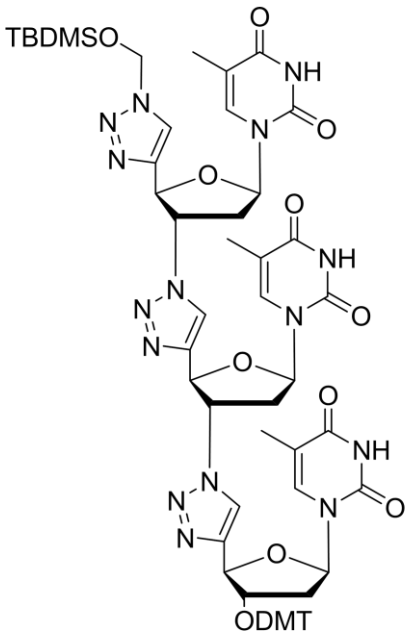
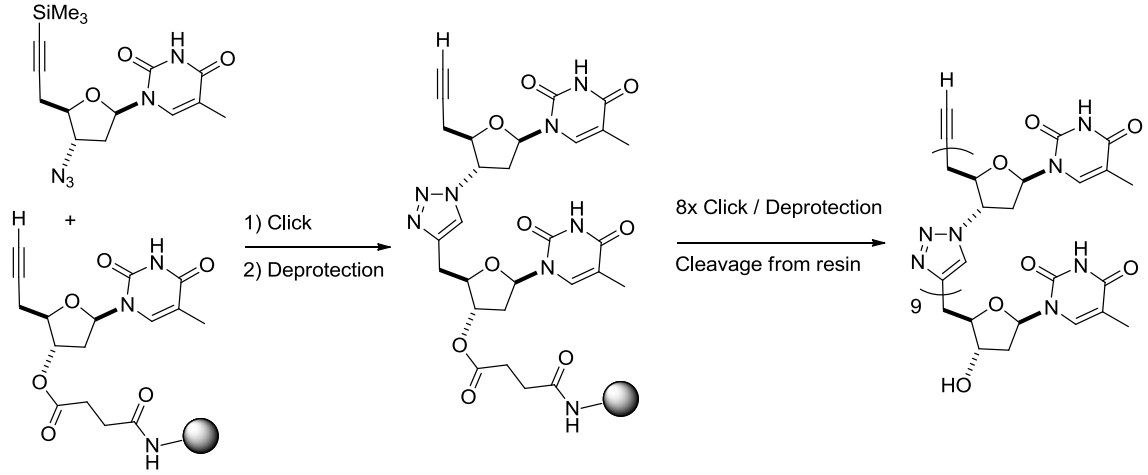


Figure 1.11 Trinucleoside with triazole backbone[36]

In 2008, Isobe and colleagues reported the efficient solid phase synthesis of a 10-mer triazole linked DNA (^{TL}DNA) utilizing a thymine base decorated with an azide at the

3’ position and a trimethylsilyl-protected acetylene moiety at the 5’ position (Scheme 1.8).[37] Initial couplings indicated the construction of the 2-mer in quantitative yield. However, a modified copper-mediated desilylation protocol was required because TBAF resulted in cleavage of the ^{TL}DNA from the resin. The 3-mer DNA was obtained in 84% yield after reaction in THF under microwave irradiation in only 1.5 hours. Finally, the 10-mer ^{TL}DNA was obtained in an isolated yield of 0.61% after 19 reaction steps and HPLC purification. Impressively, the ^{TL}DNA effectively formed a duplex with the natural $d(T)_2(A)_{10}(T)_2$ with a much higher melting temperature (T_m) of 61.1 ^oC than that of the control natural DNA $d(T)_{10}$ (20.0 ^oC). This was the first report of a highly efficient and selective route to solid phase synthesis of ON’s decorated with unnatural triazole backbones. The same authors have since developed a one-pot procedure for desilylation of a masked acetylene and CuAAC reaction to yield ^{TL}DNA.[38]



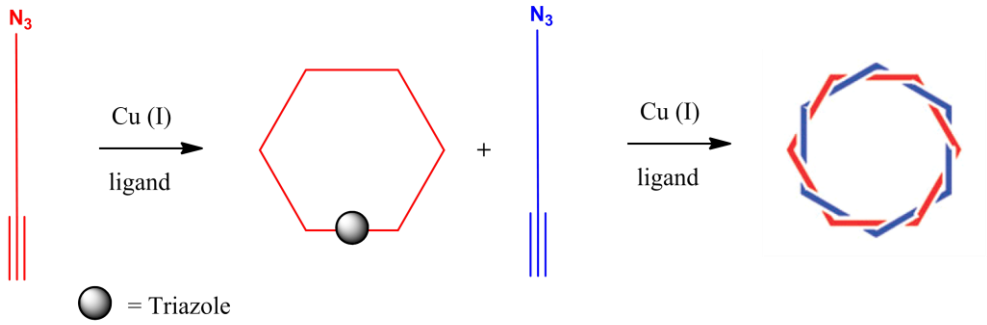
Scheme 1.8 Solid phase synthesis of 10-mer ^{TL}DNA[37]

1.3.5 DNA Cyclization and catenane formation

One of the latest examples of click chemistry on DNA is the formation of triazole-linked dumbbell ODNs by Ichikawa et al. as potential decoy molecules.[39] These decoys contained a binding sequence for NF-κB transcription factors, and thus prevented a

percentage of protein from interacting with the true transcription factor contained within a genetic sequence. These compounds have shown thermal stability with retention of the global structure associated with non-cross-linked ODNs. In addition, these dumbbells have shown excellent stability against snake venom phosphodiesterase, a 3'-exonuclease. ODN dumbbells hold the potential as powerful decoy molecules in their ability to down-regulate the inflammatory and immunological responses associated with the overexpression of the NF-κB transcription factor in *in vitro* and *ex vivo* experiments.

Click chemistry has also been applied to the assembly of single strand ON constructs starting from a 5'-alkyne and a 3'-azide. Using this installation method, cyclic ONs were formed using CuAAC conditions with Cu(I) stabilizing ligand (Scheme 1.9).[39] The formation of cyclic products was confirmed by reduced electrophoretic mobility when run through a denaturing 8% polyacrylamide gel. If the alkyne and azide are located on the same ODN strand, circularization can occur even in the absence of a template ODN. This is a very interesting result, as well as a testament to the power of the CuAAC reaction, because non-template-mediated circularization of long oligonucleotides has proven difficult.[40]



Scheme 1.9 Formation of Catenane via triazole formation

Alternative solid phase cyclization protocols have also been successful at constructing cyclic, branched, and bicyclic oligonucleotides. The microwave-assisted CuAAC reaction was more efficient in terms of cyclization efficiency, both on solid phase as well as in solution.[41]

1.4 Conclusions

This chapter examined methodologies for the synthesis of artificial ON constructs with a focus on the utilization of H-phosphonates for ON synthesis and subsequent modification. Also, a detailed account on the application of click chemistry to the modification and synthesis of ONs was examined. This background material was presented because it was relevant to and extremely valuable in efforts towards the completion of the work presented in Chapter 2 of this dissertation.

1.5 References

1. Brown, D. M. A brief history of oligonucleotide synthesis *Methods Mol. Biol.* **1993**, *20*, 1-17.

2. Beaucage, S. L., Iyer, R. P. Advances in the synthesis of oligonucleotides by the phosphoramidite approach *Tetrahedron.* **1992**, *48*, 2223-2311.

3. Stawinski, J.; Kraszewski, A. How to get the most out of two phosphorus chemistries. Studies on H-phosphonates *Acc. Chem. Res.* **2002**, *35*, 952-960.

4. Jankowska, J., Sobkowski, M., Stawinski, J., Kraszewski, A. Studies on aryl H-phosphonates: I, efficient method for the preparation of deoxyribo- and ribonucleoside 3'-H-phosphonate monoesters by transesterification of diphenyl H-phosphonate. *Tetrahedron Lett.* **1994**, *35*, 3355-3358.

5. Stawinski, J.; Stromberg, R. Di- and oligonucleotide synthesis using H-phosphonate chemistry *Methods Mol. Biol.* **2005**, *288*, 81-100.

6. Guthrie, J. P. Tautomerization equilibria for phosphorus acid and its ethyl esters, free energies of formation of phosphorus and phosphonic acids and their ethyl esters, and pKa values for ionization of the P-H bond in phosphonic acid and phosphonic esters. *Can. J. Chem.* **1979**, *57*, 236-239.

7. Stawinski, J., Thelin, M. Nucleoside H-phosphonates: XI a convenient method for the preparation of nucleoside H-phosphonate. *Nucleosides Nucleotides.* **1990**, *9*, 129-135.

8. Froehler, B. C.;Ng, P. G.; Matteucci, M. D. Synthesis of DNA via deoxynucleoside H-phosphonate intermediates *Nucleic Acids Res.* **1986**, *14*, 5399-5407.

9. Letsinger, R. L., Singman, C. N., Histand, G., Salunkhe, M. Cationic Oligonucleotides *J. Am. Chem. Soc.* **1988**, *110*, 4470-4471.

10. Jin, Y.;Chen, X.;Cote, M.;Roland, A.;Korba, B.;Mounir, S.; Iyer, R. P. Parallel solid-phase synthesis of nucleoside phosphoramidate libraries *Biorg. Med. Chem. Lett.* **2001**, *11*, 2057-2060.

11. Iyer, R. P., Egan, W., Regan, J.B., Beaucage, S.L. 3H-1,2-Benzodithiole-3-one 1,1-Dioxide as an Improved Sulfurizing Reagent in the Solid-Phase Synthesis of Oligodeoxyribonucleoside Phosphorothioates *J. Amer. Chem. Soc.* **1989**, *112*, 1253-1254.

12. Stawinski, J.; Thelin, M. Nucleoside H-Phosphonates .13. Studies on 3h-1,2-Benzodithiol-3-One Derivatives as Sulfurizing Reagents for H-Phosphonate and H-Phosphonothioate Diesters *J. Org. Chem.* **1991**, *56*, 5169-5175.

13. Kurreck, J. Antisense technologies. Improvement through novel chemical modifications *Eur. J. Biochem.* **2003**, *270*, 1628-1644.

14. Eckstein, F. Phosphorothioate oligodeoxynucleotides: what is their origin and what is unique about them? *Antisense Nucleic Acid Drug Dev.* **2000**, *10*, 117-121.

15. Johansson, T., Kers, A., Stawinski, J. 2-Pyridylphosphonates: a new type of modification for nucleotide analogues *Tetrahedron Lett.* **2001**, *42*, 2217.

16. Tornoe, C. W.;Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides *J. Org. Chem.* **2002**, *67*, 3057-3064.

17. Kolb, H. C.;Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 2004-2021.

18. Rostovtsev, V. V.;Green, L. G.;Fokin, V. V.; Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 2596-2599.

19. Lewis, W. G.;Green, L. G.;Grynszpan, F.;Radic, Z.;Carlier, P. R.;Taylor, P.;Finn, M. G.; Sharpless, K. B. Click chemistry in situ: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 1053-1057.

20. Sletten, E. M.; Bertozzi, C. R. A hydrophilic azacyclooctyne for Cu-free click chemistry *Org. Lett.* **2008**, *10*, 3097-3099.

21. El-Sagheer, A. H.; Brown, T. Click chemistry with DNA *Chem. Soc. Rev.* **2010**, *39*, 1388-1405.

22. Seo, T. S.;Li, Z.;Ruparel, H.; Ju, J. Click chemistry to construct fluorescent oligonucleotides for DNA sequencing *J. Org. Chem.* **2003**, *68*, 609-612.

23. Staudinger, H., Meyer, J. *Helv. Chim. Acta.* **1919**, *2*, 635-646.

24. Burrows, C. J.; Muller, J. G. Oxidative Nucleobase Modifications Leading to Strand Scission *Chem. Rev.* **1998**, *98*, 1109-1152.

25. Kanan, M. W.;Rozenman, M. M.;Sakurai, K.;Snyder, T. M.; Liu, D. R. Reaction discovery enabled by DNA-templated synthesis and in vitro selection *Nature.* **2004**, *431*, 545-549.

26. Chan, T. R.;Hilgraf, R.;Sharpless, K. B.; Fokin, V. V. Polytriazoles as copper(I)-stabilizing ligands in catalysis *Org. Lett.* **2004**, *6*, 2853-2855.

27. Jager, S.; Famulok, M. Generation and enzymatic amplification of high-density functionalized DNA double strands *Angew. Chem.* **2004**, *43*, 3337-3340.

28. Gierlich, J.;Burley, G. A.;Gramlich, P. M.;Hammond, D. M.; Carell, T. Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA *Org. Lett.* **2006**, *8*, 3639-3642.

29. Seela, F.; Sirivolu, V. R. DNA containing side chains with terminal triple bonds: Base-pair stability and functionalization of alkynylated pyrimidines and 7-deazapurines *Chem. Biodivers.* **2006**, *3*, 509-514.

30. Seela, F.; Sirivolu, V. R. Nucleosides and oligonucleotides with diynyl side chains: Base pairing and functionalization of 2 '-deoxyuridine derivatives by the copper(I)-catalyzed alkyne-azide 'click' cycloaddition *Helv. Chim. Acta.* **2007**, *90*, 535-552.

31. Bouillon, C.;Meyer, A.;Vidal, S.;Jochum, A.;Chevolot, Y.;Cloarec, J. P.;Praly, J. P.;Vasseur, J. J.; Morvan, F. Microwave assisted "click" chemistry for the synthesis of multiple labeled-carbohydrate oligonucleotides on solid support *J. Org. Chem.* **2006**, *71*, 4700-4702.

32. Pourceau, G.;Meyer, A.;Vasseur, J. J.; Morvan, F. Synthesis of mannose and galactose oligonucleotide conjugates by bi-click chemistry *J. Org. Chem.* **2009**, *74*, 1218-1222.

33. Rozkiewicz, D. I.;Janczewski, D.;Verboom, W.;Ravoo, B. J.; Reinhoudt, D. N. "Click" chemistry by microcontact printing *Angew. Chem.* **2006**, *45*, 5292-5296.

34. Rozkiewicz, D. I.;Gierlich, J.;Burley, G. A.;Gutsmiedl, K.;Carell, T.;Ravoo, B. J.; Reinhoudt, D. N. Transfer printing of DNA by "click" chemistry *ChemBioChem.* **2007**, *8*, 1997-2002.

35. Nakahara, M.;Kuboyama, T.;Izawa, A.;Hari, Y.;Imanishi, T.; Obika, S. Synthesis and base-pairing properties of C-nucleotides having 1-substituted 1H-1,2,3-triazoles *Biorg. Med. Chem. Lett.* **2009**, *19*, 3316-3319.

36. Nuzzi, A.;Massi, A.; Dondoni, A. Model studies toward the synthesis of thymidine Oligonucleotides with triazole internucleosidic linkages via iterative Cu(I)-Promoted azide-alkyne ligation chemistry *QSAR Comb. Sci.* **2007**, *26*, 1191-1199.

37. Isobe, H.;Fujino, T.;Yamazaki, N.;Guillot-Nieckowski, M.; Nakamura, E. Triazole-linked analogue of deoxyribonucleic acid ((TL)DNA): design, synthesis, and double-strand formation with natural DNA *Org. Lett.* **2008**, *10*, 3729-3732.

38. Fujino, T.;Yamazaki, N.; Isobe, H. Convergent synthesis of oligomers of triazole-linked DNA analogue ((TL)DNA) in solution phase *Tetrahedron Lett.* **2009**, *50*, 4101-4103.

39. Kumar, R.;El-Sagheer, A.;Tumpane, J.;Lincoln, P.;Wilhelmsson, L. M.; Brown, T. Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry *J. Am. Chem. Soc.* **2007**, *129*, 6859-6864.

40. Kool, E. T. Recognition of DNA, RNA, and Proteins by Circular Oligonucleotides *Acc. Chem. Res.* **1998**, *31*, 502-510.

41. Lietard, J.;Meyer, A.;Vasseur, J. J.; Morvan, F. New strategies for cyclization and bicyclization of oligonucleotides by click chemistry assisted by microwaves *J. Org. Chem.* **2008**, *73*, 191-200.

Chapter 2 Site-specific incorporation of diamondoids on DNA using click chemistry

Contributions

This chapter represents a modified version of a published article focusing on the incorporation of azide modified diamondoid substrates onto DNA via click chemistry. Contributions from the authors are as follows: Jason B. Crumpton, author of this dissertation, performed the synthesis and modification of all alkyne modified DNAs studied and made significant contributions to the writing and editing of the manuscript. Dr. Webster Santos made significant contributions to the writing and editing of the manuscript as well significant intellectual contributions for this work. The alkyne modified diamondoids were a gift from Prof. Peter Schreiner and Natalia Fokina. This chapter was reproduced by permission of The Royal Society of Chemistry. The original document can be located by following the link below, http://pubs.rsc.org/en/content/articlelanding/2012/cc/c2cc16860j.

Crumpton, J. B.; Santos, W. L. Site-specific incorporation of diamondoids on DNA using click chemistry *Chem. Commun.* **2012**, *48*, 2018-2020.

2.1 Introduction

2.1.1 Click reactions on DNA

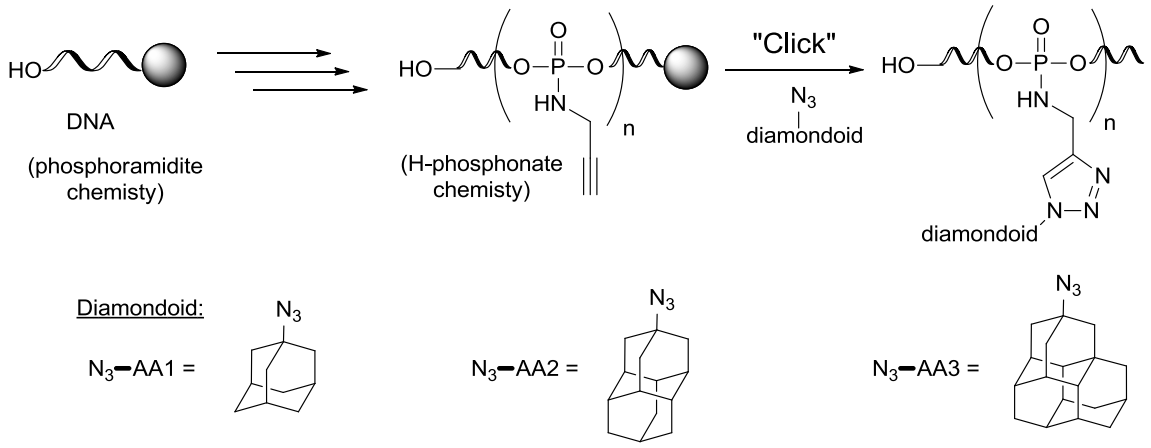
Exotic functionalization of DNA with various groups has widespread application in biotechnology, material science, diagnostics, and medicine.[1,2] Thus far, one of the most convenient, chemoselective, and compatible methods of conjugating additional functional groups to DNA is the copper (I)-catalyzed [3+2] azide-alkyne cycloaddition (CuAAC) reaction.[3,4] This reaction has recently found widespread application to the field of oligonucleotide ligation and modification. The CuAAC reaction has been utilized for intramolecular cyclization,[5] formation of nuclease resistant decoys,[6] synthesis of artificial triazole backbone,[7] oligonucleotide immobilization,[8] peptide conjugation,[9] and DNA crosslinking.[10] Indeed, the application of the CuAAC reaction to nucleic acid chemistry has shown extraordinary promise and versatility. Because of biocompatibility issues with copper, metal-free variations of the click reaction have been reported by Bertozzi utilizing high energy constrained cyclooctadiyne compounds.[10]

Ju and co-workers reported one of the earliest studies examining fluorescent labeling of azido tagged oligonucleotide on the 5'-end.[11] The CuAAC reaction under thermal conditions with alkynyl-6-carboxyfluorescein provided the triazole product in near quantitative yield and was successfully used as primer in DNA sequencing. Although the N-hydroxysuccinimide ester strategy used to arrive at the azido-functionalized DNA was efficient, functionalization on the 5'-end limited its practical utility. Carell and co-workers have since reported the high density functionalization of oligonucleotides with alkyne reporter groups, which were then reacted with azide functionalized coumarin, FAM, and sugar groups.[12] This type of high density labeling

was found to be an efficient source of alkyne-modified DNA. However, synthesis of requisite phosphoramidite monomers can present a significant limitation because four DNA building blocks modified with diynyl side chains at the 5-position of pyrimidine bases and the 7-position of 7-deazapurine or 8-aza-7-deazapurine residues are required.[13] Recently, attachment of carbohydrates to oligonucleotides using a microwave-assisted CuAAC on solid support was reported, wherein the alkyne moiety was installed using H-phosphonate chemistry.[14] In this particular case, three 'clickable' moieties were also placed on the 5'-end and required the synthesis of H-phosphonate monomers.

2.1.2 H-Phosphonate oxidations

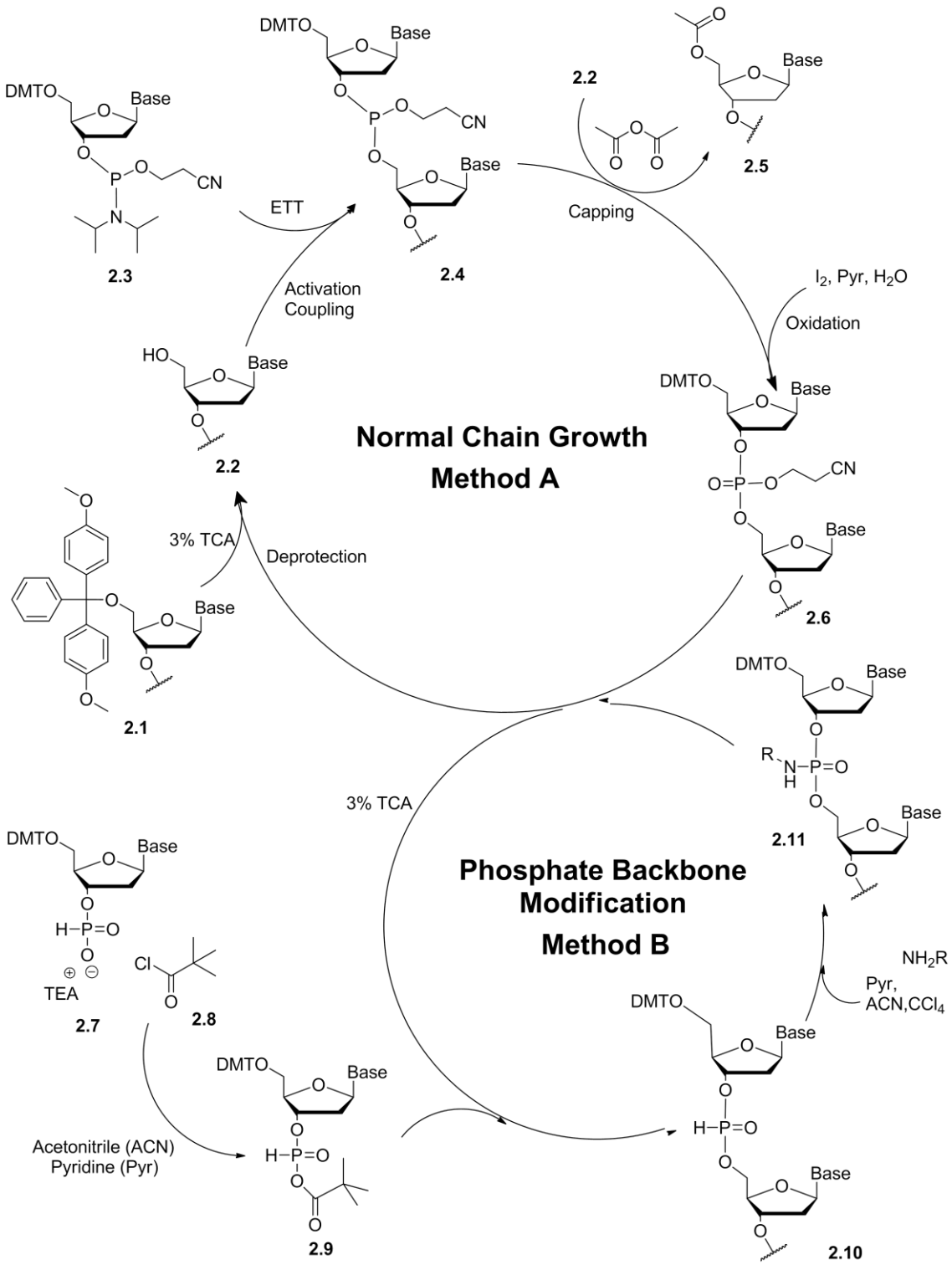
An approach that disposes the synthesis of alkyne-containing monomers and provides access to conjugated oligonucleotides with the requisite coupling partner is highly desirable. We hypothesized that commercially available H-phosphonate monomers coupled using H-phosphonate chemistry[15] could be oxidized with CCl_4/alkynylamine and generate oligonucleotides with a phosphoramidate alkyne tether that can be placed in any position along the phosphate backbone (Scheme 2.1). This type of modification is powerful not only because of the potential for high density and site-specific functionalization of oligonucleotides, but also because the tether length can be readily varied. Moreover, the phosphoramidate linkage formed is expected to be resistant to rapid hydrolysis by nucleases.[16] Nucleic acid polymers decorated with the desired organic pendant groups can result in properties suitable for a specific application.



Scheme 2.1 "Click" synthesis via H-phosphonate chemistry on solid support

2.1.3 Diamondoids

Of particular interest is incorporation of diamondoids[17] on DNA. Diamondoids are higher order derivatives of adamantane with great utility in medicine and excellent biocompatibility. For example, adamantane has been established as a building block for many pharmacologically active compounds.[18] Higher order derivatives of adamantane have also exhibited antiviral and anticancer activity[19-21] and have been utilized for the inhibition of cytochrome p450s.[22] Furthermore, adamantane has been demonstrated to form inclusion complexes with cationic transfection agents decorated with cyclodextrins.[23-25] Cyclodextrin based transfection agents are unique in that they display lower cytotoxicities than their corresponding polyamine transfection agents.[24] Therefore, adamantane conjugated ONs may yield favourable physiochemical properties, such as enhanced cellular uptake, when transfected with cyclodextrin decorated transfection agents. To date, site-specific and high density functionalization of DNA with diamondoids and their properties have not been published. Herein, we report a convenient, site-specific method of installing multiple diamondoids to DNA using CuAAC on solid support.



TCA = Trichloroacetic acid, ETT = 5-Ethylthio-1H-tetrazole, TEA = Triethylamine,

DMT = 4,4′-dimethoxytrityl

Scheme 2.2 Oligonucleotide synthesis on solid support

2.2 Results

2.2.1 Optimization of Reaction Conditions

We initiated our studies by modifying an automated phosphoramidite DNA/RNA synthesizer to accommodate H-phosphonate chemistry and subsequent oxidation with CCl_4/propargyl amine at the desired position (Scheme 2.2).[26] Traditional ON synthesis was performed using method **A**, which begins with the deprotection of **2.1** to yield **2.2**. The free 5'-hydroxyl is then coupled to phosphoramidite **2.3** to yield **2.4**, which is followed by oxidation to yield **2.6**. Any unreacted **2.2** is capped with acetic anhydride to yield **2.5**, which prevents further elongation. At this point, normal chain growth (method **A)** can be repeated as desired, or the synthetic cycle can branch off to the phosphate backbone modification cycle (method **B**). Compound **2.6** is deprotected and combined with activated H-phosphonate **2.9** to yield H-phosphonate modified ON **2.10**. The last step of method B is the oxidation of **2.10** to yield phosphoramidate **2.11**. Methods **A** or **B** may then be employed at any point during subsequent ON growth; some of the advantages of this protocol are the site-specific installation of the pendant group on the oligonucleotide backbone and the easy access to amino-alkynes. Additionally, the resulting process relieves the necessity for the cumbersome synthesis of alkyne-derivatized nucleic acid monomers. We also opted for a solid phase CuAAC ligation, in part because of the simple removal of the copper catalyst by filtration, but mostly because of the powerful capability of installing multiple and different azide coupling partners. In particular, a significant advantage is gained for azide coupling partners that are immiscible in water or mixtures of organic and aqueous solvents, such as **AA1-AA3**.

With the propargyl amidate-DNA conjugate in hand, we screened a series of

copper (I) sources for the formation of the triazole using azidotriadamantane **AA3** (Table 2.1). The traditional copper sulphate/ascorbic acid system[27] did not provide the desired product at rt (entry 1). Various combinations of solvents such as water, THF, acetonitrile and DMSO were also unproductive. However, when the temperature was raised to 40 or 60 °C, the reaction proceeded nearly to completion after 12 hours (entries 2-3). Substitution of the Cu (I) source with $CuBr{\cdot}SMe_2$ showed no reaction at room temperature (entry 4), but tetrakis(acetonitrile)copper(I) hexafluorophosphate ($[Cu(CH_3CN)_4]PF_6$) in anhydrous acetonitrile gave significant conversion to product at room temperature after 12 hours (entry 5). Increasing the temperature to 40 °C resulted in almost complete conversion after 12 hours (entry 6), and to our delight, a further increase in temperature to 60 °C afforded the product in quantitative yield after 3 hours (entry 7). In comparison to $[Cu(CH_3CN)_4]PF_6$, other Cu(I) sources (CuCl, CuCN, and CuBr) were significantly less efficient (entries 8-10). The use of $[Cu(CH_3CN)_4]PF_6$ as the copper source offers several advantages: excellent solubility in organic solvents such as acetonitrile, removal of possible interference of adventitious water during the phosphoramidite coupling reactions, and a reduced necessity of expensive copper-stabilizing ligands such as TBTA. Oligonucleotide strand breaks resulting from aqueous Cu(I)-mediated chemistry were not observed, possibly because the reaction was performed in nonaqueous conditions and sensitive groups on the oligonucleotide were still protected.[28,29] Since the "clicked" product is expected to be retarded because of the change in charge, molecular weight, and hydrophobicity, the progress of all click reactions was monitored by electrophoretic mobility shift assay (EMSA) on denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 2.1).

Table 2.1 Optimization of reaction conditions

	Reaction Time (hrs)		1	3	6	12
Entry	Temp (^{o}C)	Copper Source	Approximate Conversion[b] (%)			
1	rt	$CuSO_4$	NR	NR	NR	NR
2	40	CuSO4	40	50	80	90
3	60	$CuSO_4$	40	60	80	90
4	rt	$CuBr{\cdot}SMe_2$	NR	NR	NR	NR
5	rt	$[Cu(CH_3CN)_4]PF_6$	40	50	60	60
6	40	$[Cu(CH_3CN)_4]PF_6$	50	70	80	90
7	60	$[Cu(CH_3CN)_4]PF_6$	90	100	100	100
8	40	CuCl	NR	NR	NR	NR
9	40	CuCN	NR	NR	NR	NR
10	40	CuBr	5	40	50	50

[a] The DNA sequence 5'-TTT TTT TTT TG**x**T-3' was used. **x** = site of phosphoramidate modification. [b] Conversions are visual approximations. rt = room temperature. **AA3** was utilized for all "click" reactions. NR = no reaction.

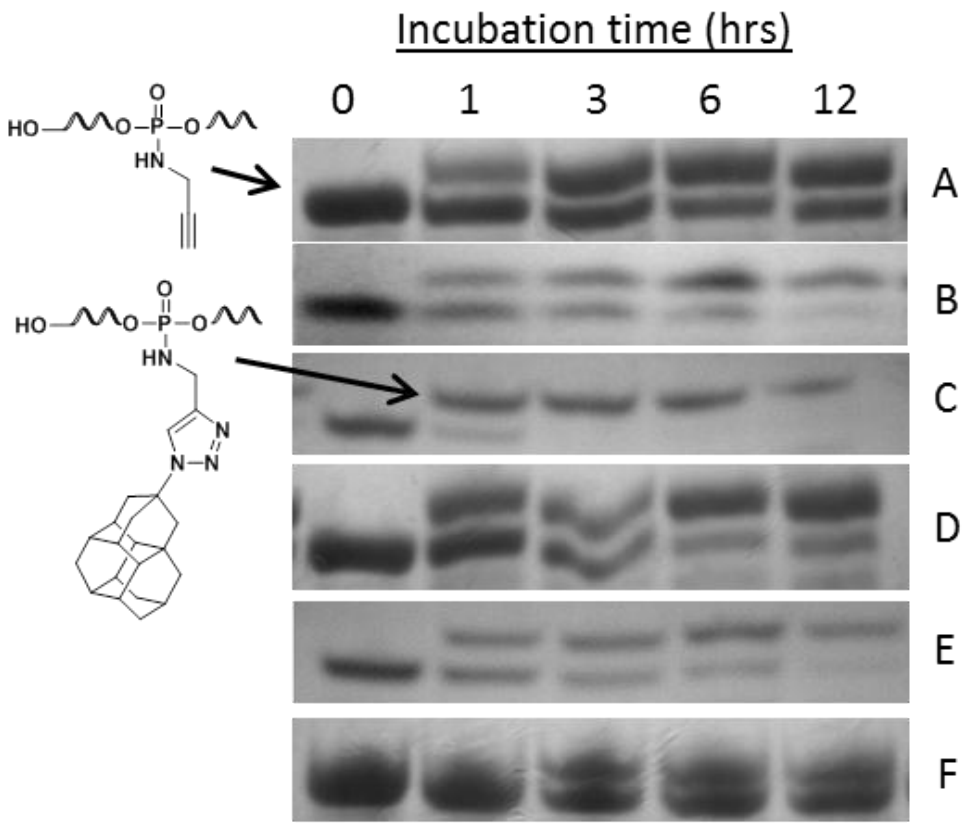


Figure 2.1 Time-course EMSA assays using $[Cu(CH_3CN)_4]PF_6$ at rt (A), 40 °C (B) and 60 °C (C). $CuSO_4$/ascorbate system assayed at 40 °C (D) and 60 °C (E). CuBr assayed at 40 °C (F)

2.2.2 Thermal Stability Studies

With the optimized conditions in hand, we synthesized $A_{10}GA$ (arbitrarily chosen sequence) 12-mer oligonucleotides containing a single phosphoramidate on the 3'-end of the DNA backbone. These strands were modified with adamantane (**AA1**), diamantane (**AA2**), and triamantane (**AA3**) to determine whether such modification had an effect on duplex stability (Table 2.2). See section **2.4.10** for experimental details for T_m data acquisition. ON **2.13**, which contains the propargyl group linked to the DNA via a phosphoramidate, had a slightly increased melting temperature ($\Delta T_m = + 0.8$) compared to the unmodified ON **2.12**. Surprisingly, phosphoramidate modification and ligation

with azidodiamondoid **AA1** (ON **2.14**) resulted in slightly increased T_m values compared with propargyl modification (ON **2.13**). Further, as the size of the diamondoid increased from adamantane (**AA1,** ON **2.14**) to triamantane (**AA3,** ON **2.16**), there is a corresponding increase in DNA duplex stability. Although the observed increase in duplex stability is minor, these effects were found to be reproducible and statistically significant. Since higher orders of azidodiamondoids were unavailable for study, we were unable to determine if T_m values would continue to increase with the size of the diamondoids.

Table 2.2 Thermal stability of diamondoid-clicked DNA.[a]

ON	Diamondoid	5'-Sequence-3' [b]	T_m (°C) [c]	Δ T_m (°C)
2.12	none	AAA AAA AAA AGA	35.9 ± 0.3	--
2.13	Propargyl	AAA AAA AAA AG**x**A	36.7 ± 0.3	+ 0.8
2.14	AA1	AAA AAA AAA AG**x**A	36.3 ± 0.3	+ 0.4
2.15	AA2	AAA AAA AAA AG**x**A	37.7 ± 0.4	+ 1.8
2.16	AA3	AAA AAA AAA AG**x**A	38.9 ± 0.3	+ 3.0

[a] Each double strand complex was prepared as 10 μM solutions in 200 mM NaCl and 10 mM potassium phosphate buffer. Each strand was hybridized to the unmodified complimentary TCT_{10} strand. [b] **x** denotes the site of modification. [c] Each measurement was performed in triplicate.

2.2.3 Duplex structure and HPLC analysis of Diamondoid Modified DNAs

To determine whether the DNA duplex structure was affected by the size and hydrophobicity stemming from the diamondoid pendant groups, the circular dichroism spectra of ONs **2.12** - **2.16** were measured. Each ON was hybridized to its complimentary sequence before data acquisition. See section **2.4.10** for details on CD spectra acquisition

conditions. When comparing the CD spectra of the natural duplex ON **2.12** to **AA3** modified duplex ON **2.16**, it was found that the DNA secondary structure was not perturbed (Figure 2.2).[30] This similar CD fingerprint was also detected with ONs **2.13-2.15**. After analysis by reversed-phase HPLC, it was found that the DNA with the propargyl modification **2.13** had the shortest retention time on the HPLC (4.8 min). As the adamantanes **AA1** - **AA3** were conjugated to the DNA, the retention time of the corresponding ONs increased. Specifically, the smallest and least hydrophobic adamantane unit **2.14** had a shorter retention time (6.4 min) than the bulkier diamondoids, diamantane **2.15** (7.4 min) and triamantane **2.16** (8.2 min) (Figure 2.3). The corresponding increase in retention time is presumably a consequence of the hydrophobicity of the diamondoid groups. See **2.4.8** for additional HPLC elution profiles.

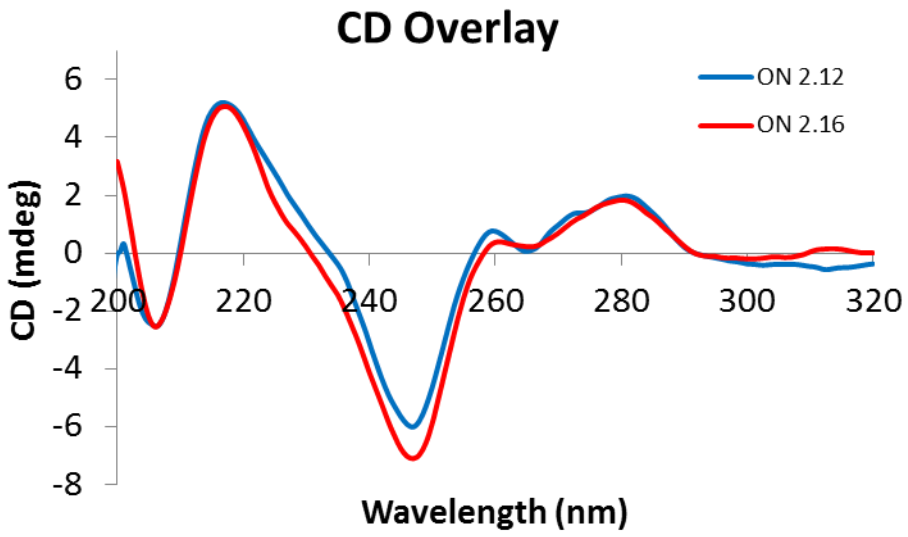


Figure 2.2 Circular Dichroism spectra of duplex **2.12** and **2.16** hybridized to TCT_{10}

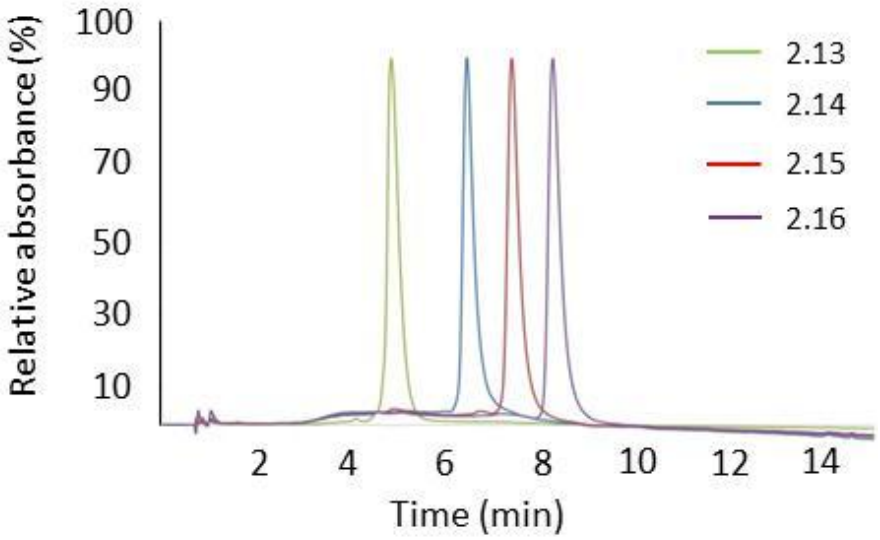


Figure 2.3 RP-HPLC analysis of purified d(A_{10}GxA) ON **2.13-2.16**. Mobile phase A = 0.1 M TEAA buffer (pH = 7.5); Mobile phase B = CH_3CN. Ramp: [time (%A)]: 0(90), 10(60), 13(20), 20(20), 23(90), 30(90). Flow rate = 0.5 mL/min. UV detection was at 260 nm. TEAA = triethylammonium acetate buffer

2.2.4 Click Protocol Compatibility Analysis

To test the generality and compatibility of the protocol, a series of oligonucleotide-diamondoid conjugates bearing 6-carboxyfluorescein (6-FAM) on the 3'-end were synthesized, since fluorescently labeled nucleic acid derivatives are important for diagnostic applications (Table 2.3).[31] Azido-adamantane, diamantane and triamantane derivatives readily "clicked" on the 5'-end of the DNA as determined by MALDI-MS after PAGE purification (ONs **2.17-2.19**). Further elaboration by introducing up to two triamantane derivatives internally proceeded smoothly to provide the desired product (ON **2.20**).

Table 2.3 Functionalization of 6-FAM labelled DNA with diamondoids

ON	DNA Sequence [a]	Diamondoid	Calcd MW	Obsd MW
2.17	A**x**CCCAACACTACTCGGC	AA1	5,867.9	5,868.6
2.18	A**x**CCCAACACTACTCGGC	AA2	5,919.9	5,921.5
2.19	A**x**CCCAACACTACTCGGC	AA3	5,972.0	5,971.6
2.20	A**x**CCCAA**x**CACTACTCGGC	AA3	6,290.3	6,289.0

[a] All modified DNA sequences were synthesized with a 6-FAM label on the 3'-end.

x = site of modification.

2.3 Conclusions

In summary, we developed a simple, robust, and site-specific solid-phase method for the introduction of diamondoids on the phosphate backbone of DNA using CuAAC. The procedure was most efficient using $[Cu(CH_3CN)_4]PF_6$ as the copper source. This procedure allows for high density functionalization and is compatible with attachment of multiple and distinct pendant groups on DNA.

2.4 Experimental

2.4.1 Bases used for DNA Synthesis

All DNA syntheses were performed on a BioAutomation MerMade synthesizer. Generic protocols were modified as shown in **2.4.3** to accommodate phosphoramidate synthesis. All phosphoramidites were purchased from Glen Research as follows:

Base A: 5'-Dimethoxytrityl-N-benzoyl-3'-deoxyAdenosine,2'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite

Base T: 5'- Dimethoxytrityl-3'-deoxyThymidine,2'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite

Base G: 5'-Dimethoxytrityl-N-dimethylformamidine-3'-deoxyGuanosine,2'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite

Base C: 5'-Dimethoxytrityl-N-acetyl-2'-deoxyCytidine,3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite

H-Phosphonate G: 5'-Dimethoxytrityl-N-isobutyryl-2'-deoxyGuanosine,3'-H-phosphonate, TEA salt

2.4.2 H-Phosphonate reagent preparation

1. H-Phosphonate dG: 52 mg in 1.5 mL of 1:1 ACN:Pyr (0.04 M solution).
2. Pivaloyl chloride *(Piv-Cl)*: 90 μL in 3.0 mL of 1:1 ACN:Pyr (0.24 M solution).
3. Propargyl amine: 48 μL in 3.0 mL of 1:1:1 ACN : Pyr : CCl_4 (0.25 M solution)

2.4.3 Procedure for H-Phosphonate Addition/Oxidation

1. ***Deblock***: To a dimethoxyl trityl (DMT) protected 1 μmol CPG column, 2 × 200 μL of 3% TCA in DCM solution was injected with 6 × 3.3 sec pulsed incubations to yield a total incubation time of 20 sec.
2. ***Acetonitrile wash***: To the deprotected column, 3 × 250 μL of acetonitrile was injected with a 1 × 1 sec pulsed incubation followed by a 55 × drain pulse to completely remove all acetonitrile from the reaction column.
3. ***Coupling***: To the deprotected column, 2 × 100 μL of H-Phosphonate reagent (0.04 M, 4 μmol) and 2 × 100 μL of Piv-Cl reagent (0.24 M, 24 μmol) were added as a staggered injection with 8 × 12.5 sec pulsed incubations to yield a total incubation time of 100 sec per injection.

4. ***Acetonitrile Wash***: To the column, 3 × 250 μL of acetonitrile was injected with a 1 × 1 sec pulsed incubation followed by a 55 × drain pulse to completely remove all acetonitrile from the reaction column.
5. ***Propargyl addition***: To the column, 5 × 200 μL of propargyl amine (0.25 M, 500 μmol) was injected with a 10 × 24 sec incubation to yield a total incubation time of 240 sec per injection.
6. ***Capping***: To the column, 1 × 125 μL of cap mix A (THF / 2,6-lutidine / acetic anhydride [8:1:1]) and 1 × 125 μL of cap mix B (16 % 1-methyl imidazole in THF) were added as a staggered injection with 9 × 5.5 sec pulsed incubations to yield a total incubation time of 50 sec per injection.
7. ***Acetonitrile Wash***: To the column, 3 × 250 μL of acetonitrile was injected with a 1 × 1 sec pulsed incubation followed by a 55 × drain pulse to completely remove all acetonitrile from the reaction column.

2.4.4 Preparation of Copper, Adamantane, and AMA solution

Dilution of Adamantane Sources: All adamantane solutions were diluted to a 0.1M concentration using THF. The dilutions were stored in a –78°C freezer between experiments.

Dilution of Copper Sources: Tetrakis(acetonitrile)copper(I)hexafluorophosphate (10 mg) was diluted in anhydrous acetonitrile (268 μL) to yield a 0.1 M solution immediately before each experiment. Copper sulfate pentahydrate (10 mg) was diluted in H_2O (400 μL) to yield a 0.1 M solution. Copper bromide dimethyl sulfide, copper bromide, copper cyanide, and copper chloride were all diluted to a 0.1 M solution immediately before each experiment.

Preparation of AMA solution: Ammonium hydroxide and 40% aqueous methylamine were mixed in a 1:1 ratio. http://www.glenresearch.com/GlenReports/GR7-12.html

2.4.5 EMSA Assays

To evenly distributed portions of CPG solid phase (~30 nmol), copper (100 eq) and adamantane (10 eq) were added and allowed to react for variable times. The beads were washed with 2 × 1 mL of DMSO, 1 × 1 mL H_2O, and 1 × 1 mL ACN. The beads were cleaved with AMA solution (1.0 mL) at 65 ^{o}C for 15 minutes. The supernatant was collected and dried on a GeneVac. The resulting pellet was dissolved in loading buffer, loaded onto a 20% denaturing PAGE gel, and run at 250 V for 4 hours. The resulting spots were visualized under a handheld UV lamp.

2.4.6 Preparation of MALDI Matrix and DNA sample for MALDI analysis

Zip Tip Desalt: Each DNA sample was desalted using a Harvard C-18 Zip-Tip (Cat # 74-3403)

Preparation of 2',4',6'-Trihydroxyacetophenone monohydrate (THAP) Matrix: In an Eppendorf tube, 22 mg of ammonium citrate dibasic (AC) was dissolved in 1mL of 1:1 ACN:H_2O. THAP matrix (10 mg) was stirred vigorously with 100 µL of the AC solution to prepare a saturated THAP solution. The insoluble particulates were spun down and the supernatant was collected for use as MALDI matrix.

2.4.7 Modified HPLC Conditions for Purity Determination

In order to confirm the purity of all the oligonucleotides post click reaction, an elongated HPLC run was performed to further confirm the lack of co-elution with any impurities. The doublets for each HPLC are indicative of the Rp and Sp diastereomers of each phosphoramidate modified oligonucleotide. All HPLCs were performed on an

Agilent Zorbax SB-C18 column (2.1 x 150 mm). Flow rate = 0.5 ml/min. Detection at 260 nm.

Mobile phase A = 0.1 M TEAA buffer (pH = 7.5); Mobile phase B = Acetonitrile.

Elution ramp is as follows [time(%A)] : 0(90), 20(60), 25(20), 30(20), 35(90), 40(90)

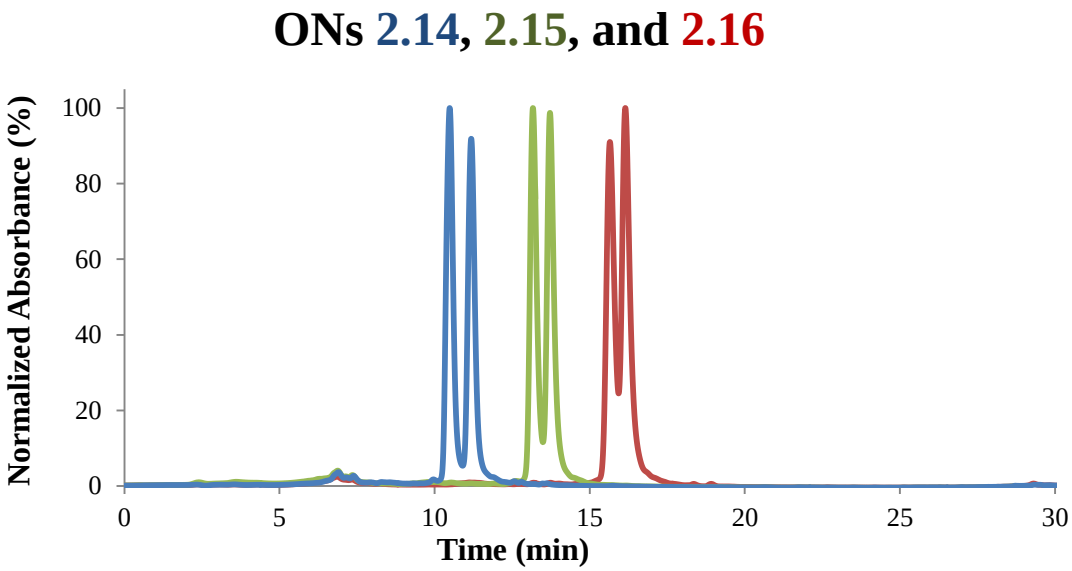


Figure 2.4 Elongated elution of ONs 2.14, 2.15, and 2.16.

2.4.8 Digestion and HPLC/MS analysis of ON 4

S1 nuclease buffer = 30 mM sodium acetate (pH 4.6 at 25°C), 50 mM NaCl, 1 mM $ZnCl_2$, 0.5 mg/ml denatured calf thymus DNA and 5% glycerol.

Digestion: 50 µL of oligonucleotide **2.15** (121.6 µM) was diluted with 9 µL of 10x S1 nuclease buffer followed by 1 µL of S1 nuclease (98 units/µL). The mix was incubated at 37 ^{o}C overnight. The next day, 40 µL of 0.1 M Tris Base (pH = 8.5) was added to the reaction buffer and 2 units (2 µL) of calf intestinal alkaline phosphatase was added and incubated at 37 ^{o}C for 2 hours. The solution was filtered and analyzed by HPLC using the following gradient.

HPLC Conditions: All HPLCs were performed on an Agilent Zorbax SB-C18 column (2.1 x 150 mm). Flow rate = 0.2 ml/min. Detection at 260 nm. Mobile phase A = 20 mM KH_2PO_4 buffer (pH = 5.5); Mobile phase B = 30% (v/v) buffer A in methanol. Elution ramp is as follows [time (%A)]: 0(100), 60(0), 65(0), 70(100), 80(100)

dA peak (rt = 27.6 min) : ESI-MS m/z calcd for $C_{10}H_{13}N_5O_3$ $[M+H]^+$ 252.10, found 252.10

dA/dG dimer peak (rt = 31.6 min) : ESI-MS m/z calcd for $C_{37}H_{47}N_{14}O_8P$ $[M+H]^+$ 846.34, found 846.30

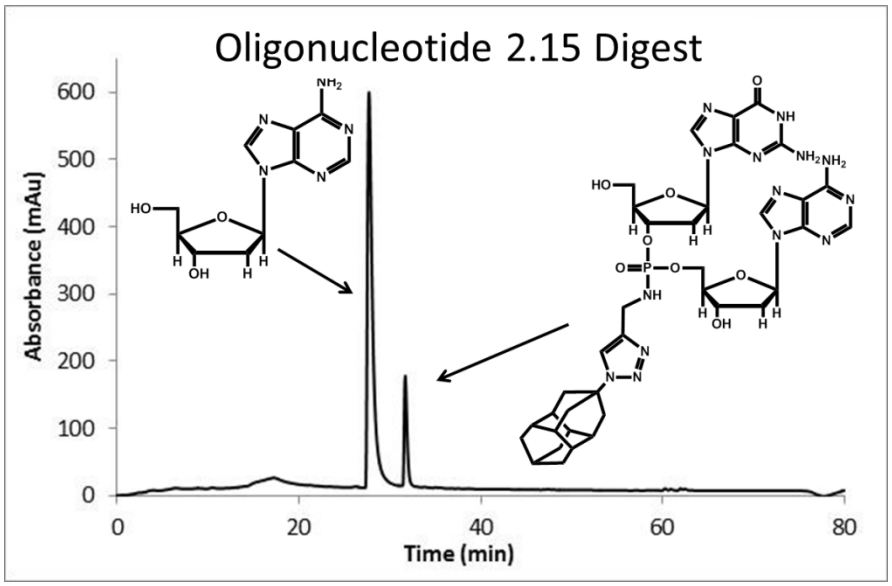


Figure 2.5 ON 2.15 digest

2.4.9 Phosphoramidite and H-phosphonate coupling efficiency

To confirm the successful incorporation of each nucleotide, internal sensors were set to detect the elution of the intensely red colored trityl groups formed during the deprotection steps. As shown in Figure 2.6, the deprotection steps of bases 1-12 were at least 90% complete, indicating successful incorporation of the previous nucleotide (See Scheme 2.2, conversion of **2.1** to **2.2**). Furthermore, comparison of the raw trityl data

after conventional phosphoramidite coupling (Figure 2.7) to the raw trityl data after H-phosphonate coupling (Figure 2.8), provides further evidence for highly efficient H-phosphonate coupling as well as evidence confirming the preservation of high efficiency coupling of future coupling steps after oxidation with propargyl amine.

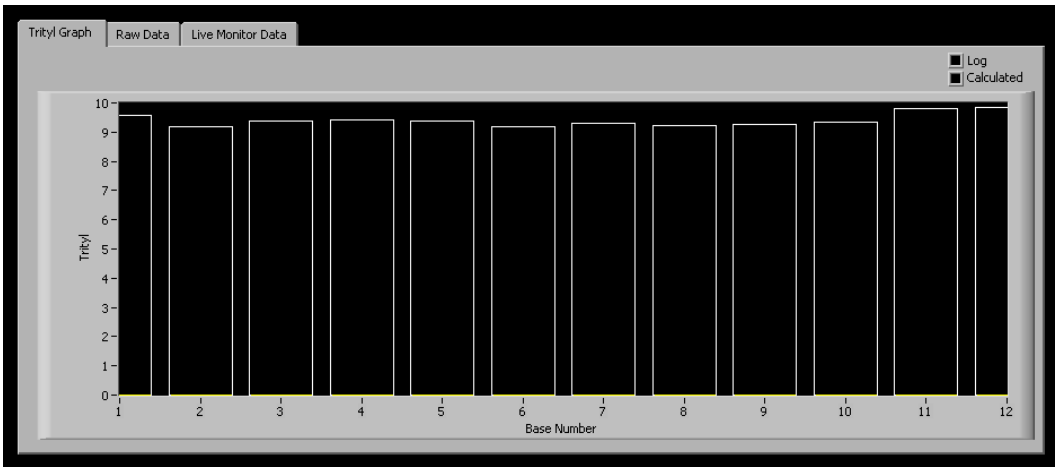


Figure 2.6 Trityl graphs for the synthesis of ON **2.13**

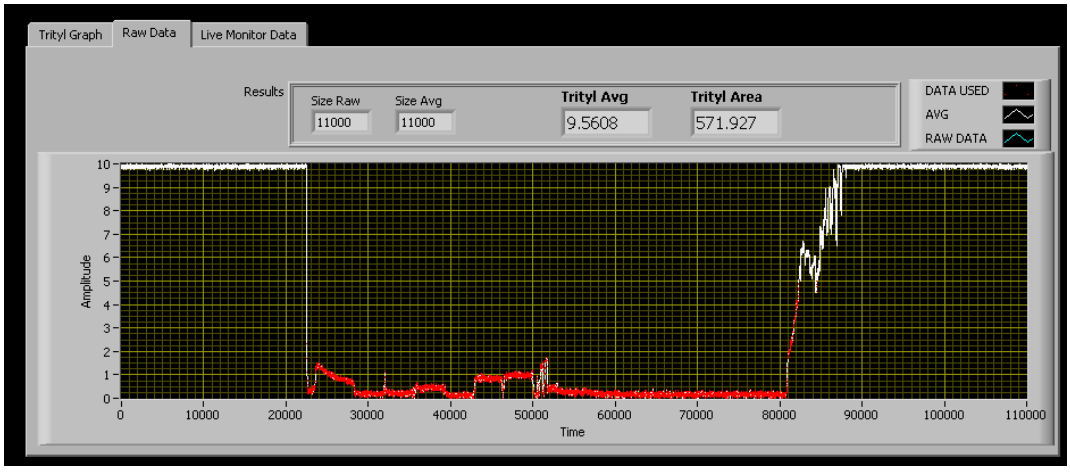


Figure 2.7 Raw trityl data for base 1 of ON **2.13**: Data is representative of conventional phosphoramidite coupling efficiencies for bases 1 and 3-12

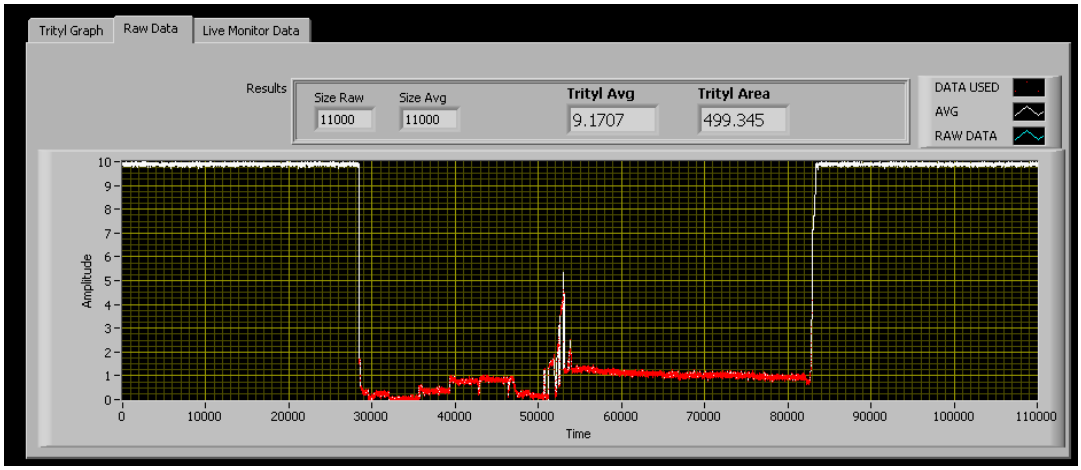


Figure 2.8 Raw trityl data for base 2 of ON **2.13**: Data is representative of H-phosphonate coupling efficiencies

2.4.10 CD and T_m analysis

The unmodified ON **2.12** as well as the modified ONs **2.13 – 2.16** were prepared as 10 µM solutions in 200 mM NaCl and 10 mM potassium phosphate buffer. Each ON was annealed to the complimentary unmodified strand (TCT_{10}) by incubating at 90 ºC for 5 minutes followed by slow cooling to room temperature. All CD and T_m data are representative of the hybridized duplexes and were acquired on a Jasco J-815 CD spectrometer. The CD spectra were acquired with standard sensitivity settings with a scanning range of 320 – 190 nm, a data pitch of 0.5 nm, and a scanning speed of 100 nm/min. Spectra for each duplex are reported as the average of 3 accumulations. The T_m data were acquired with a heating profile of 5 ºC – 80 ºC at 260 nm with a ramp rate of 2 ºC/min with data sampling every 2 ºC. Final T_m values were calculated from the data that were acquired in quadruplicate using Spectra Manager Version 2 software.

2.5 References

1. El-Sagheer, A. H.; Brown, T. Click chemistry with DNA *Chem. Soc. Rev.* **2010**, *39*, 1388-1405.

2. Weisbrod, S. H.; Marx, A. Novel strategies for the site-specific covalent labelling of nucleic acids *Chem. Commun.* **2008**, 5675-5685.

3. Wang, Q.;Chan, T. R.;Hilgraf, R.;Fokin, V. V.;Sharpless, K. B.; Finn, M. G. Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193.

4. Tornoe, C. W.;Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides *J. Org. Chem.* **2002**, *67*, 3057-3064.

5. Kumar, R.;El-Sagheer, A.;Tumpane, J.;Lincoln, P.;Wilhelmsson, L. M.; Brown, T. Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry *J. Am. Chem. Soc.* **2007**, *129*, 6859-6864.

6. Nakane, M.;Ichikawa, S.; Matsuda, A. Triazole-linked dumbbell oligodeoxynucleotides with NF-kappaB binding ability as potential decoy molecules *J. Org. Chem.* **2008**, *73*, 1842-1851.

7. Isobe, H.;Fujino, T.;Yamazaki, N.;Guillot-Nieckowski, M.; Nakamura, E. Triazole-linked analogue of deoxyribonucleic acid ((TL)DNA): design, synthesis, and double-strand formation with natural DNA *Org. Lett.* **2008**, *10*, 3729-3732.

8. Rozkiewicz, D. I.;Gierlich, J.;Burley, G. A.;Gutsmiedl, K.;Carell, T.;Ravoo, B. J.; Reinhoudt, D. N. Transfer printing of DNA by "click" chemistry *ChemBioChem.* **2007**, *8*, 1997-2002.

9. Gogoi, K.;Mane, M. V.;Kunte, S. S.; Kumar, V. A. A versatile method for the preparation of conjugates of peptides with DNA/PNA/analog by employing chemo-selective click reaction in water *Nucleic Acids Res.* **2007**, *35*, e139.

10. Laughlin, S. T.;Baskin, J. M.;Amacher, S. L.; Bertozzi, C. R. In vivo imaging of membrane-associated glycans in developing zebrafish *Science.* **2008**, *320*, 664-667.

11. Seo, T. S.;Li, Z.;Ruparel, H.; Ju, J. Click chemistry to construct fluorescent oligonucleotides for DNA sequencing *J. Org. Chem.* **2003**, *68*, 609-612.

12. Gierlich, J.;Burley, G. A.;Gramlich, P. M.;Hammond, D. M.; Carell, T. Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA *Org. Lett.* **2006**, *8*, 3639-3642.

13. Seela, F.;Sirivolu, V. R.; Chittepu, P. Modification of DNA with Octadiynyl Side Chains: Synthesis, Base Pairing, and Formation of Fluorescent Coumarin Dye Conjugates of Four Nucleobases by the Alkyne−Azide “Click” Reaction *Bioconjugate Chem.* **2007**, *19*, 211-224.

14. Bouillon, C.;Meyer, A.;Vidal, S.;Jochum, A.;Chevolot, Y.;Cloarec, J. P.;Praly, J. P.;Vasseur, J. J.; Morvan, F. Microwave assisted "click" chemistry for the synthesis of multiple labeled-carbohydrate oligonucleotides on solid support *J. Org. Chem.* **2006**, *71*, 4700-4702.

15. Froehler, B.;Ng, P.; Matteucci, M. Phosphoramidate analogues of DNA: synthesis and thermal stability of heteroduplexes *Nucleic Acids Res.* **1988**, *16*, 4831-4839.

16. Peyrottes, S.;Vasseur, J. J.;Imbach, J. L.; Rayner, B. Synthesis, Base-Pairing Properties and Nuclease Resistance of Oligothymidylate Analogs Containing Methoxyphosphoramidate Internucleoside Linkages *Nucleos Nucleot.* **1994**, *13*, 2135-2149.

17. Schwertfeger, H.;Fokin, A. A.; Schreiner, P. R. Diamonds are a chemist's best friend: diamondoid chemistry beyond adamantane *Angewandte Chemie. International Ed. In English.* **2008**, *47*, 1022-1036.

18. Cote, M.;Misasi, J.;Ren, T.;Bruchez, A.;Lee, K.;Filone, C. M.;Hensley, L.;Li, Q.;Ory, D.;Chandran, K.; Cunningham, J. Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection *Nature.* **2011**, *477*, 344-348.

19. Wang, J. J.;Huang, K. T.; Chern, Y. T. Induction of growth inhibition and G1 arrest in human cancer cell lines by relatively low-toxic diamantane derivatives *Anti-Cancer Drugs.* **2004**, *15*, 277-286.

20. Wang, J. J.;Chang, Y. F.;Chern, Y. T.; Chi, C. W. Study of in vitro and in vivo effects of 1,6-Bis[4-(4-amino-3-hydroxyphenoxy)phenyl]diamantane (DPD), a novel cytostatic and differentiation inducing agent, on human colon cancer cells *Br. J. Cancer.* **2003**, *89*, 1995-2003.

21. Wang, J. J.; Chern, Y. T. Biological activities of new poly(N-1-adamantylmaleimide) and poly(N-1-diamantylmaleimide) *Journal of Biomaterials Science, Polymer Edition.* **1996**, *7*, 905-915.

22. Hodek, P.;Bortek-Dohalska, L.;Sopko, B.;Sulc, M.;Smrcek, S.;Hudecek, J.;Janku, J.; Stiborova, M. Structural requirements for inhibitors of cytochromes P450 2B: assessment of the enzyme interaction with diamondoids *J. Enzyme Inhib. Med. Chem.* **2005**, *20*, 25-33.

23. Davis, M. E.; Bellocq, N. C. Cyclodextrin-Containing Polymers for Gene Delivery *J Incl. Phenom. Macro.* **2002**, *44*, 17-22.

24. Pun, S. H.;Bellocq, N. C.;Liu, A.;Jensen, G.;Machemer, T.;Quijano, E.;Schluep, T.;Wen, S.;Engler, H.;Heidel, J.; Davis, M. E. Cyclodextrin-modified polyethylenimine polymers for gene delivery *Bioconjugate Chem.* **2004**, *15*, 831-840.

25. Li, J.; Loh, X. J. Cyclodextrin-based supramolecular architectures: syntheses, structures, and applications for drug and gene delivery *Adv. Drug Del. Rev.* **2008**, *60*, 1000-1017.

26. The H-phosphonate-CCl4/amine oxidation to form the phosphoramidate by pressing back and forth using a 1 mL syringe is also a successful alternative.

27. Kolb, H. C.; Sharpless, K. B. The growing impact of click chemistry on drug discovery *Drug Discovery Today.* **2003**, *8*, 1128-1137.

28. Burrows, C. J.; Muller, J. G. Oxidative Nucleobase Modifications Leading to Strand Scission *Chem. Rev.* **1998**, *98*, 1109-1152.

29. Kanan, M. W.;Rozenman, M. M.;Sakurai, K.;Snyder, T. M.; Liu, D. R. Reaction discovery enabled by DNA-templated synthesis and in vitro selection *Nature.* **2004**, *431*, 545-549.

30. Gray, D. M.;Hung, S. H.; Johnson, K. H. Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes *Methods Enzymol.* **1995**, *246*, 19-34.

31. Ranasinghe, R. T.; Brown, T. Fluorescence based strategies for genetic analysis *Chem. Commun.* **2005**, 5487-5502.

Chapter 3 HIV structure, function, and design of therapeutic targets

3.1 HIV

3.1.1 Statistics and structure

As of 2010, the World Health Organization (WHO) estimated that the total number of people living with HIV worldwide was around 34 million people. In 2010, 2.7 million people were infected by the HIV virus, which was down from 3.1 million newly infected in 2001. Although the annual number of people dying from AIDS-related causes worldwide has decreased from 2.2 million in 2005, an estimated 1.8 million people died from the disease in 2010.[1] As a perspective comparison, approximately 1.2 million people were estimated to have died in road traffic accidents in 2008. The WHO has categorized the HIV/AIDS epidemic as one of the top 10 leading causes of death worldwide. Thus, it is no surprise that herculean efforts have been put forth to restrict the proliferation of the virus as well as to develop novel therapeutics to treat individuals already infected.

The human immunodeficiency virus type 1 is a retrovirus that has been extensively scrutinized and studied by scientists for over 25 years.[2] An understanding of HIV's structural components as well as how these components interact with host cells have paved the way for the construction of many effective antiviral agents. Although none of these agents have offered a cure, HIV has effectively become much more of a manageable disease in recent years. When broken down to its most fundamental components, the virus itself can be considered as an assembly of 15 proteins and one RNA. The actual structure of the HIV-1 genome is encoded by an ~ 9 kb RNA, which

encompasses 9 open reading frames, each of which encodes for various structural, enzymatic, and regulatory functions (Figure 3.1).[3]

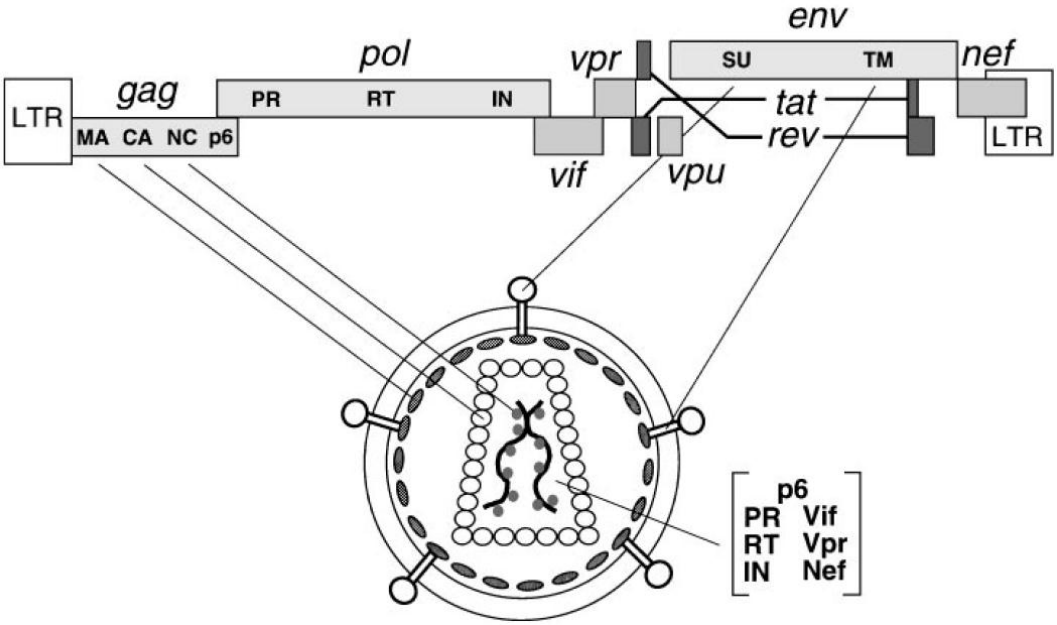


Figure 3.1 Organization of the HIV-1 genome and virion. Reproduced with permission of Annual Reviews in the format Journal via Copyright Clearance Center (Reference 3)

3.1.2 Virus life cycle

A mature virion begins its replication cycle with infection of healthy cells by targeting the surface gp 120 (SU) proteins and CD4 and CC chemokine co-receptors (Figure 3.2). After this binding event occurs, the transmembrane gp41 (TM) protein undergoes a conformational change that allows virus-cell membrane fusion, which results in the entry of the virus envelope.[3] The core of the virus is uncoated to expose a viral nucleoprotein complex that is transported to the nucleus, where the genomic RNA is reverse transcribed by reverse transcriptase (RT) into duplex DNA. The enzyme integrase (IN) catalyzes the integration of the viral DNA into the host's chromosome. Following this step, the viral genome is transcribed beginning at the 5' long terminal repeat (LTR).

Initially, this transcription is quite slow, but the rate of transcription is enhanced after the binding of the Tat protein to the TAR region of the viral transcript.

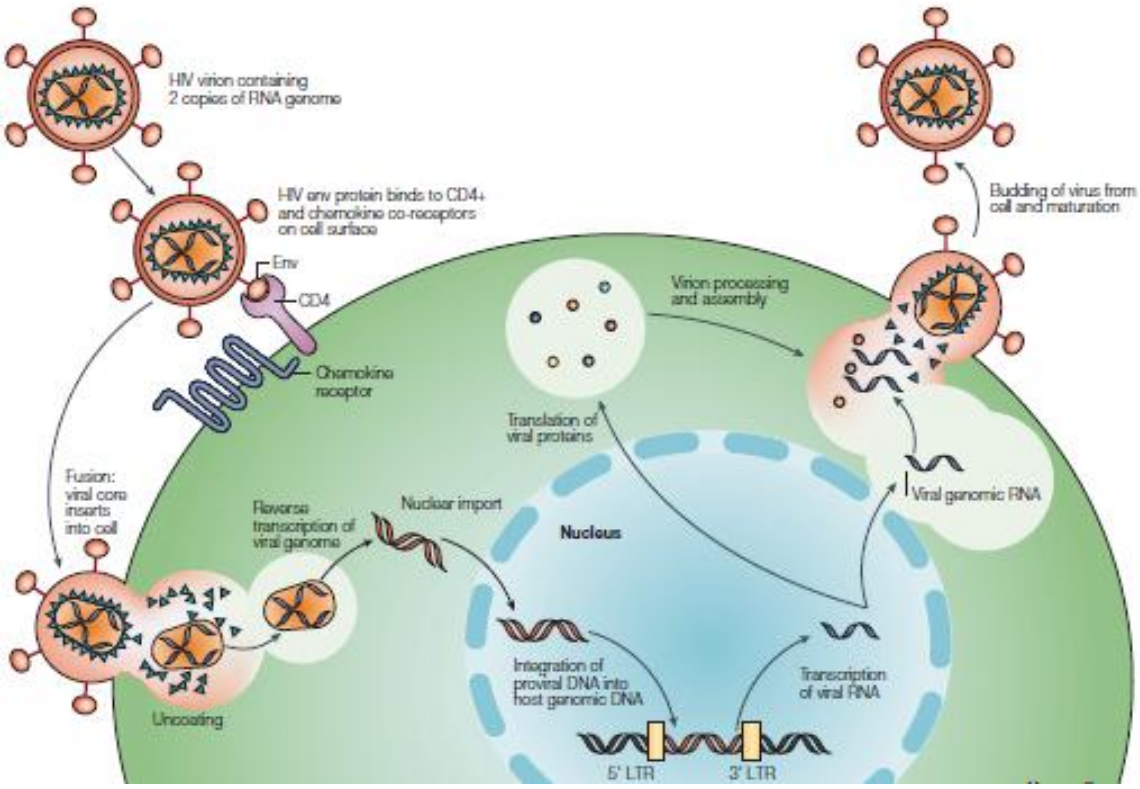


Figure 3.2 Basic life cycle of HIV-1. Reprinted by permission from Macmillan Publishers Ltd: *Nat. Rev. Genet.* (Reference 2), copyright 2004.

The next step is regulated by the Rev protein, which promotes the effective transport of longer RNA strands along with spliced HIV-RNA to the cytoplasm. After these RNAs are transported to the cytoplasm, they are translated or packaged for the final virion particle. The Gag and Gag-Pol polyproteins (later cleaved by HIV-1 protease (PR) enzyme to yield MA, CA, NC, p6, PR, RT, and IN) are translated and shuttled to the cell membrane while the Env mRNA, which codes for the surface (SU) and transmembrane proteins (TM), is translated at the endoplasmic reticulum (ER). The core virion particle is assembled from the Gag and Gag-Pol proteins, Vif, Vpr, Nef, and genomic RNA.

At this point, a virion begins to bud from the cell surface, but before the virion can detach itself from the host cell, it must first be coated with the SU and TM proteins for future binding to host receptor proteins. For the Env to be incorporated onto the outer shell of the virion, it must be released from complexes that are formed with CD4, which is expressed in the ER. Vpu promotes CD4 degradation, which releases Env for transport to the cell surface. The Nef protein promotes endocytosis and degradation of surface CD4 receptors.

Finally, the virion particle can bud and be released from the cell surface with a coat of SU and TM proteins. After processing of the Gag and Gag-Pol polyproteins by the PR, the virion is then ready to infect the next cell and the cycle continues.

3.2 Current anti-retroviral therapy (ART) drugs

Therapeutics that aim to combat the proliferation of HIV are limited to seven broad classes. Each of these classes has been designed to inhibit the successful infection of the host cell, replication within the host, or proliferation of the complete virion from the cell.[4] These will be discussed in the following sections.

3.2.1 Nucleoside/Nucleotide Reverse Transcriptase inhibitors (NRTIs)

NRTIs produce anti-HIV effects by inhibiting the activity of the HIV reverse transcriptase.[5] These compounds are generally nucleoside analogues that are designed by modification of a sugar, nucleic base, or a simultaneous modification of both moieties. Interestingly, the first anti-HIV drug approved for AIDs patients was the NRTI 3'-azido-2',3'-dideoxythymidine (**AZT**) (Figure 3.3).[6] Once inside the cell, the NRTI is effectively converted to the triphosphate derivative that is incorporated into the growing viral DNA by DNA polymerases. Since all NRTIs lack a 3'-hydroxyl group, the growing

DNA sequence is terminated; no other nucleotides can be incorporated, and viral replication is inhibited. Unfortunately, the virus has developed a repair mechanism that involves the removal of the chain-terminating residues.

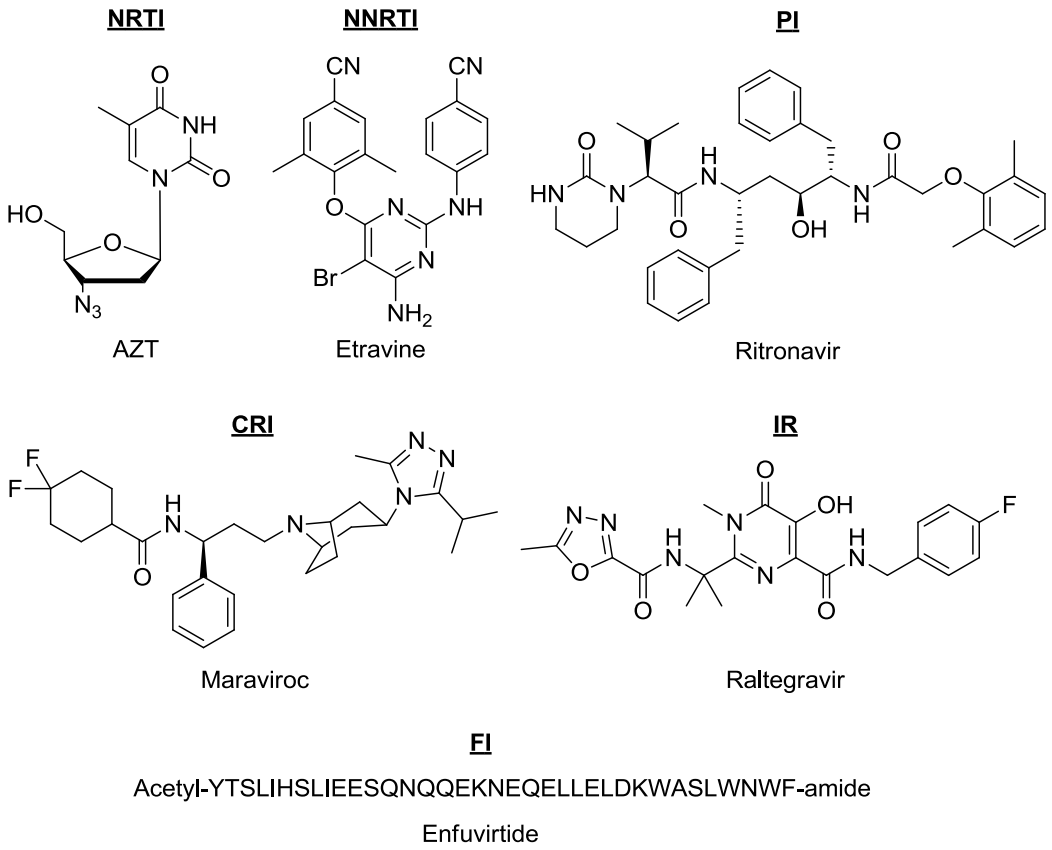


Figure 3.3 Selected examples of ART drugs

3.2.2 Non-nucleoside Reverse Transcriptase inhibitors (NNRTIs)

NNRTIs were first reported in the early 1990's and, like NRTIs, are designed to inhibit the process of HIV-reverse transcription. However, unlike NRTIs, these therapeutics attempt to interrupt the action of the transcriptase protein itself by binding to an allosteric site located very close to the catalytic site.[7] Unfortunately, of the four NNRTIs which have been approved for clinical use, three have been plagued by a rapid development of resistance. **Etravine** (Figure 3.3) is a second generation NNRTI that

binds to conserved residues of the reverse transcriptase such as W229 and was approved for clinical use in 2008. Unlike other approved NNRTIs, **Etravine** has displayed improved potency against several NNRTI resistant mutants.[8]

3.2.3 Protease Inhibitors (PIs)

The HIV protease is responsible for the cleavage of precursor polyproteins to structural proteins, functional proteins, reverse transcriptase, and integrase.[9] Consequently, its function is essential for the successful maturation of the virus as well as its ability to proliferate and spread to other host cells. Nine out of the ten currently FDA-approved protease inhibitors are peptidomimetic transition state analogues that interfere with the cleavage of these precursor polyproteins (*e.g.*, **Ritonavir**[10] (Figure 3.3)). Unfortunately, viral resistance, poor bioavailability, and toxicity are problems associated with this class of inhibitors.

3.2.4 Fusion Inhibitors (FIs)

In order to successfully enter a host cell, the viral glycoprotein gp120 must interact with a CD4 receptor on the host cell surface. A cascade of interactions between viral proteins and co-receptors on the cell surface results in the activation of the viral transmembrane fusion protein, gp41. This conformational activation results in the viral and cellular membranes to be drawn together for fusion.[11] Since gp41's interaction with the cellular membrane is essential for effective fusion, it is not surprising that the only approved fusion inhibitor, enfuvirtide, is a synthetic 36-amino acid peptide aimed at mimicing structural components of gp41.[12]

3.2.5 Co-Receptor Inhibitors (CRIs)

Before virus-cell fusion takes place, a secondary interaction between the viral gp120 and cellular chemokine co-receptors must take place. Two major co-receptors that have been identified as essential for viral fusion are CCR5 and CXCR4.[13] Although numerous small molecule CCR5 and CXCR4 antagonists have been identified, the compound **Maraviroc** (Figure 3.3) is the only CCR5 antagonist that has been approved for the treatment of HIV.[14]

3.2.6 Integrase Inhibitors (IRs)

For viral proliferation to take place, the proviral DNA must first be integrated into the host DNA. The integrase enzyme is essential for the successful catalysis of this process and is considered an extremely promising target for anti-HIV therapeutics. One of the reasons for this is that the catalytic core domain (CCD) is composed of a highly conserved DDE motif for amino acids D64, D116, and E152.[15] These residues are essential for the activity of the enzyme, and as such are exquisite targets for inhibition. At the moment, the only FDA-approved HIV-integrase inhibitor is **Raltegravir** (Figure 3.3) and is available only for use in combination with other anti-HIV drugs.[16]

3.2.7 Anti-HIV Drug Combinations

Although each class of inhibitor has demonstrated utility in combating the HIV virus, the introduction of combination approach termed highly active antiretroviral therapy (HAART) in the mid-90's has significantly decreased morbidity and mortality rates of HIV patients. Most therapeutic approaches now include two or more drugs that each targets a different stage of the virus' life cycle.[17]

3.3 RNA as a "drugable" target

The structural and functional diversity of HIV-therapeutics is a testament to the difficulty associated with attempting to interrupt the successful replication of the virus. Unfortunately, even with the most effective therapies, the emergence of drug resistance is always a threat, and the interruption of drug administration results in reemergence of viral activity.[18,19] Therefore, the development of novel therapeutic targets and methods are of the utmost importance.

Recently, HIV-RNA structures essential for successful replication and proliferation of the virus have become extremely attractive targets. Interestingly, RNA can adopt folded structures with a high degree of complexity that can approach the complexity of folded protein structures.[20] Therefore, these tertiary structures can be addressed as sites for drug recognition and binding. Indeed, the three-dimensional folding of RNA is already utilized in the cell as protein and small molecule binding platforms.[21] The construction of synthetic compounds that can recognize and bind to RNA structures with high affinity and specificity will be extremely useful next generation therapeutics.

3.3.1 Small and large molecules for RNA binding

According to Lipinski's rule of five, small molecules are generally considered those with MW's < 500 Daltons.[22] When inspecting various examples of ART's, such as those in Figure 3.3, it becomes apparent that, with the exception of enfuvirtide, the majority of the drugs that target the HIV-1 virus are small molecule protein inhibitors/antagonists. Alternative small molecules such as the aminoglycosides, oxazolidinones, and intercalators have all been demonstrated to effectively bind to RNA structures.[20,23] Although these binders may be designed to display good affinities towards

RNA structure, this feat is generally accomplished at the cost of promiscuous binding (poor selectivity) amongst a variety of RNA structures. Large molecules such as those used in RNAi have demonstrated the necessary selectivity and affinity, but they suffer from poor cellular uptake.

Compounds that can bridge the gap between these two extremes in terms of bioavailability, stability, cell permeability, affinity, and selectivity will be welcome additions as a new class of RNA binders. Interestingly, there has been an increasing interest within the last decade in peptides and peptide analogs for targeting RNA structures. Indeed, these molecules seem to have shown potential to compromise between the two current RNA targeting strategies, and have displayed impressive selectivity as well as affinity towards high impact RNA targets such as HIV-TAR and HIV-RRE.

3.4 Peptides and peptidic analogs

3.4.1 Backbone Cyclic Peptides

In 2007, Chaloin reported a series of Backbone Cyclic Peptides (BCP) analogs that contains a conformationally constrained arginine rich motif (ARM) of Rev.[24] Sixteen Rev-BCPs were synthesized with rings that contained between 2 and 5 carbons for the "m" chain and 2 and 6 carbons for the "n" chain (Figure 3.4). These compounds were tested for *in vitro* inhibition of HIV-1 replication (Table 3.1), and it was determined that all BCP's assayed were more active than the linear form of the Rev-ARM peptide (**2.9**). Compounds **2.1** and **2.5** displayed similar IC_{50}'s but disparate (m + n) values, and were chosen as representative Rev-BCP models. Both peptides were able to completely block HIV-1 replication when incubated at 50 μM in human H9 T-lymphocytes and

demonstrates that Rev-BCPs can be used to inhibit virual replication in $CD4^+$ and $CXCR4^+$ human T cells.

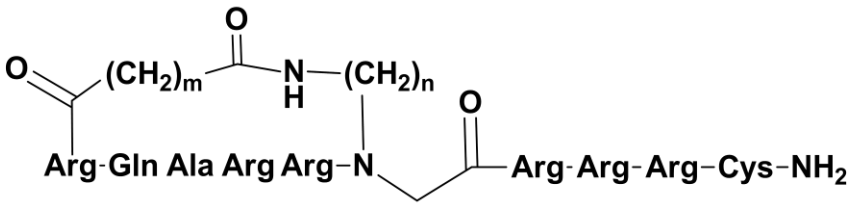


Figure 3.4 Rev-BCP

Table 3.1 Representative sampling of Rev-BCPs

Compound	m	n	Sum (m + n)	IC_{50} (μM)
2.1	2	2	4	8.5
2.2	3	2	5	12.6
2.3	4	2	6	9.7
2.4	3	4	7	14.6
2.5	2	6	8	10
2.6	3	6	9	6.2
2.7	4	6	10	13.9
2.8	5	6	11	16.2
2.9 (Lin-Rev)	-	-	-	24.3

3.4.2 Cyclic β-hairpin peptidomimetics

A second example of cyclic peptides used to target specific RNA structures comes from Varani and co-workers. They targeted the Tat protein using peptidomimetic chemistry combined with a structure-based approach. There is no structural information regarding the HIV Tat-TAR complex, so they attempted to understand recognition principles regarding the well characterized and structurally similar Bovine Immunodeficiency Virus (BIV) Tat-TAR (Figure 3.5).[25,26] They hypothesized that

information gained from studying peptidomimetic interactions with the BIV could also be applied to the HIV.

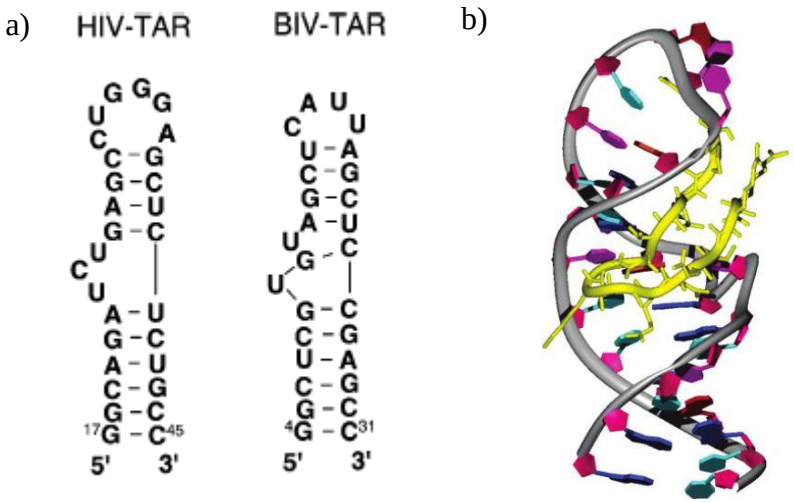


Figure 3.5 a) HIV-TAR and BIV-TAR. b) BIV Tat-TAR complex. Reprinted (adapted) with permission from reference 28. Copyright (2004) American Chemical Society

Structural information from the BIV Tat-TAR complex suggests that Tat adopts a β-hairpin conformation upon binding to TAR. Hence, they decided to introduce a cyclic β-hairpin conformation to their peptide library via a D-Pro-L-Pro template.[27] After evaluating this small library of β-hairpin peptidomimetics, they elucidated the structure of the cyclic peptide [cyclo-D-Pro-L-Pro-Arg^{1}-Val^{2}-Arg^{3}-Thr^{4}-Arg^{5}-Gly^{6}-Lys^{7}-Arg^{8}-Arg^{9}-Ile^{10}-Arg^{11}-Val^{12})] (**2.10)** by mutational analysis of positions 1-3, 11, and 12.[28] Peptide **2.10** was determined to display nanomolar affinity (150 nM) to BIV TAR even with a large excess of competing tRNA (10,000x). In comparison, the BIV Tat was determined to have a 50 nM binding constant under the same conditions.

Thereafter, ~100 peptides were constructed by performing mutations on all residues of **2.10**, and 3 sequences stood out in regards to their potency.[29] As shown in Table 3.2, **2.11**, **2.12**, and **2.13** displayed activity in the low nM range. These compounds

were able to withstand competition from a 10,000-fold excess of tRNA.[30] Also, compounds **2.11** and **2.13** were able to discriminate between HIV-1 and BIV TAR.

Table 3.2 High affinity binders to HIV and BIV TAR

	Position													
Compound	1	2	3	4	5	6	7	8	9	10	11	12	K_d (HIV), nM	K_d (BIV), nM
2.11	R	V	R	T	R	K	G	R	R	I	R	I	30	5
2.12	R	V	R	T	R	G	K	R	R	I	R	R	1	1
2.13	R	T	R	T	R	G	K	R	R	I	R	V	5	50

In 2011, Varani *et al.* reported a series of conformationally constrained cyclic peptides that act as structural mimics of the HIV-1 Tat protein at nanomolar concentrations *in vitro*.[31] These compounds were also able to block transcription in cell-free as well as cell-based assays. These compounds were also cell permeable, displayed low toxicity, and inhibited the replication of a number of HIV-1 strains, including CXCR4-tropic and CCR5-tropic HIV-1 isolates. The most effective compound, **2.12**, exhibited an IC_{50} value of ~250 nM in human peripheral blood mononuclear cells.

3.4.3 α-helical peptides

Recently, Yu and co-workers reported a study on the effects of N,N-dimethyl-Lys containing α-helical peptides on RRE RNA binding.[32] This study was inspired by observations that suggested that methylated Lys/Arg residues in certain proteins were responsible for specificity and selectivity.[33,34] Binding affinities for a series of synthetic peptides were determined by fluorescence anisotropy (Table 3.3).

Interestingly, even though the peptides have similar α-helical content to the Rev peptide, most peptides displayed an order of magnitude decrease in binding affinity. When comparing peptides with only two **K*** modifications (**2.18-2.20**), it appears that the

location of **K*** is more influential on affinity than the number of modifications or α-helical content. In addition to binding RRE RNA with similar affinity to the Rev peptide, peptide **2.20** was able to discriminate between RRE RNA and $tRNA^{mix}$ as well as TAR RNA with higher selectivity than the Rev peptide itself.

Table 3.3 Helicity and binding of N,N-dimethyl-Lys derivatives

Compound	Sequences[a]	Helicity (%) [b]	K_d (nM)
2.14	LKKLLKLLKKLLKLKG	26/57	22
2.15	LKKLLKLLKKLL**K***LKG	5/45	79
2.16	LKKLLKLL**K***KLLKLKG	9/48	74
2.17	LK**K***LLKLLKKLLKLKG	8/50	75
2.18	LKKLLKLL**K***KLL**K***LKG	7/49	30
2.19	LK**K***LLKLLKKLL**K***LKG	6/43	69
2.20	LK**K***LLKLL**K***KLLKLKG	8/52	9.1
2.21	LK**K***LLKLL**K***KLL**K***LKG	6/47	87
2.22 (Rev)	TRQARRNRRRRWRERQRAAAAR	33/73	8.5

[a] **K*** = N^{ε},N^{ε}-dimethyl Lys. [b] In 10mM H_3PO_4 / 50% TFE in 10mM H_3PO_4 at pH 7.4.

In a subsequent report, Yu reported the affinity of amphiphilic helical peptides that included acridinyl moieties via covalent attachment to the N^{ε}-position of Lys residues.[35] They hypothesized that the inclusion of the acridine intercalator would increase binding affinity via π-π interactions while retaining the selectivity imparted by their previous design.[32] After scanning a series of mono- and di-acridinylated peptides utilizing the same parent peptide sequence shown in Table 3.3, the tightest binder

displayed an affinity of 610 pM to RRE. Unfortunately, it was also discovered that none of the acridine modified peptides displayed adequate selectivity towards TAR or RREs.

Guy and colleagues recently described a new class of conformationally constrained peptidomimetics that resulted in an extremely specific binder to the HIV-1 RRE. In contrast to reports of peptides designed with α-turn conformational restraints,[30,36] this was the first report of peptides with macrolactam-induced (i, i+4) α-helical conformations being used for targeting viral RNA.[37] Their design was based on the highly specific RRE binding peptide R_6QR_7 (**2.26**) that was identified in a genetic selection experiment.[33] A representative sample of R_6QR_7 peptidomimetics are shown in Table 3.4. Of the 12 macrolactam constrained peptides, only 3 were found to have significant α-helical content (**2.23**, **2.24**, and **2.25**) as determined by circular dichroism (CD) (dual minima at 222 and 208 nm). Also, only one peptide (**2.24**) displayed specificity towards the wild-type RRE IIB against C46-G74 mutant RRE, which has been shown to not specifically recognize the Rev or R_6QR_7 peptides. K_d's were determined by electrophoretic mobility shift assays. Specificity was defined as the ratio of dissociation constants of wild-type RRE IIB and C46-G74 mutant RRE.

Table 3.4 Amino acids in red were utilized for the macrolactam constraint. RRE binding

Compound	Sequence	K_{sp} (nM)	K_{nonsp} (nM)	specificity
2.23	Ac-RRR**E**RRQ**K**RRRRRR-OH	700	1200	1.7
2.24	Ac-RRRR**E**RQR**K**RRRRR-OH	45	1200	26
2.25	Ac-RRRRR**E**QRR**K**RRRR-OH	700	1200	1.7
2.26	R_6QR_7	150	1200	8
2.27	Rev17	100	1600	16

3.4.4 Multivalent binding oligomers (MBOs)

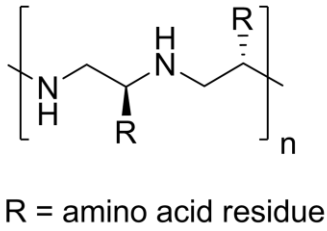


Figure 3.6 Generalized structure of MBO polyamines

In 2009 Appella *et al.* described the synthesis of multivalent binding oligomers (MBOs), which are composed of multiple amino acid side chains that extend from a polyamine backbone (Figure 3.6).[38] While derived from amino acids, MBOs are composed of an amine linkage instead of an amide linkage. The hypothesis is that non-ionic sidechains influence the specificity of binding while the amines in the backbone contribute ionic interactions that contribute to binding affinity toward the anionic RNA backbone.

A series of MBO derivatives were investigated for their ability to inhibit Tat-TAR interactions. As shown in Table 3.5, as the length of the MBO was increased from **2.28-2.33**, a significant increase in inhibitory activity was observed. An increase in chain length from the trimer (**2.28**) to the octamer (**2.33**) resulted in two orders of magnitude increase in EC_{50} values. To determine the importance of side chain residues, single and multiple point alanine mutations were introduced on the MBO sequence (entries **2.31a-2.31f**). Although single point mutations did not significantly affect the binding of the MBO to the TAR RNA, a five point mutation caused almost an order of magnitude increase in EC_{50} value (**2.31f**). Similarly, sidechain mutations to lysine and tryptophan

residues (**2.231g-2.31j**) did not impart any significant decrease in inhibition in the Tat-TAR interaction when compared to **2.32a** and **2.33a.**

Table 3.5 Variation on MBO sequence and length

Length variation		Alanine scan		Sidechain variation	
Compound	EC_{50} (μM) in vitro	Compound	EC_{50} (μM) in vitro	Compound	EC_{50} (μM) in vitro
YYY (2.28)	83.8 ± 3.7	YAYYYY (2.31a)	1.9 ± 0.3	YYYKYY (2.31g)	0.5 ± 0.1
YYYY (2.29)	8.9 ± 1.0	YYAYYY (2.31b)	3.2 ± 0.4	YYYKWY (2.31h)	0.3 ± 0.1
YYYYY (2.30)	3.0 ± 0.7	YYYAYY (2.31c)	1.6 ± 0.3	YKYKYY (2.31i)	0.4 ± 0.1
YYYYYY (2.31)	1.7 ± 0.4	YYYYAY (2.31d)	3.9 ± 0.5	YKYKWY (2.31j)	0.2 ± 0..1
YYYYYYY (2.32)	0.5 ± 0.2	YYYYYA (2.31e)	3.3 ± 0.5	YYYAYYY (2.32a)	0.5 ± 0.1
YYYYYYY Y (2.33)	0.2 ± 0.1	YAAAAA (2.31f)	14 ± 2	YYYAYYYY (2.33a)	0.20 ± 0.02

After comparing activities and selectivities of MBOs against Tat and Rev function in HeLa cell systems, the polyamines **2.31**, **2.31c**, and **2.32a** were selected for tests against antiviral activity due to their high activity and selectivity compared to other MBOs. All three compounds demonstrated activity in the low micromolar range with toxicity profiles significantly higher than the concentration required to show activity against HIV-1. The MBOs in Table 3.5 were also found to retain their inhibitory activity over a variety of different HIV clades.

3.4.5 Peptoids

In 1997, Hamy and coworkers reported the synthesis of a hybrid peptoid/peptide oligomer of 9 residues (CGP64222) that was able to block the formation of the Tat/TAR RNA complex in vitro at nanomolar concentrations (Figure 3.7).[39] Peptoids are isomers of peptides wherein side chains are attached to amide nitrogens. These are interesting substrates for binding due to their unique secondary structures and resistance towards enzymatic degradation. **CGP64222** was discovered from a combinatorial chemistry approach where each of the 5 peptoid residues were optimized in a sequential manner. In this library, the four C-terminal residues were held constant with a composition of D-Lys-D-Lys-D-Arg-D-Pro-amide (k-k-r-p*) in order to limit complexity of the library deconvolution. Antiviral activity of **CGP64222** was tested in a cellular assay system referred to as the fusion-induced gene stimulation (FIGS) assay to determine its ability to inhibit HIV-LTR function.[40] Along with a control peptide, **Arg5** (RRRRR-k-k-r-p*), **CGP6422** was found to inhibit HIV replication with no cytotoxic changes up to 100 μM. In human lymphocytes, incubation with 30 μM of **CGP64222** was able to completely suppress the production of reverse transcriptase (RT) without inhibiting cell viability while the **Arg5** control was able to achieve only ~50% inhibition.

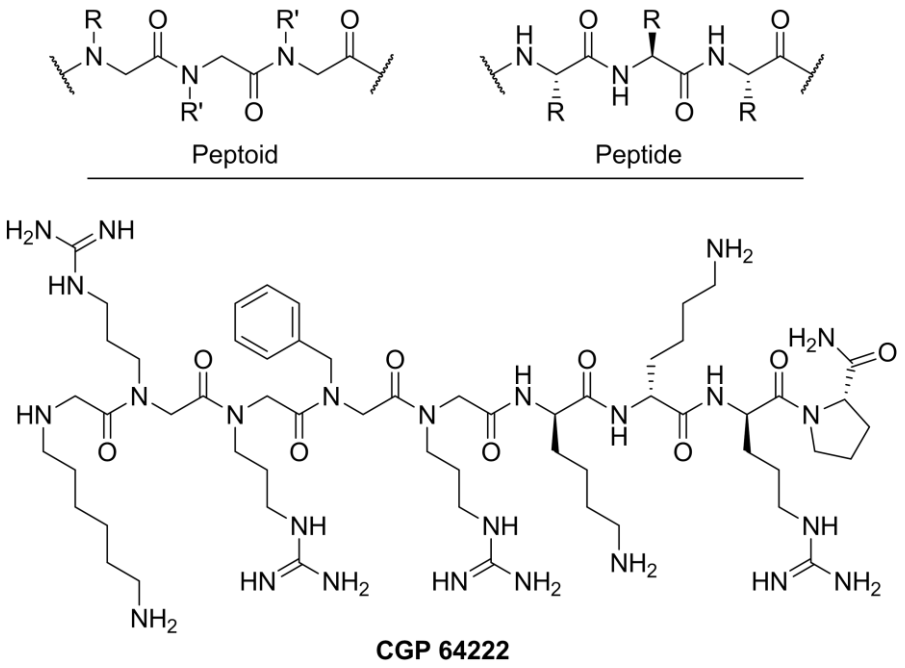


Figure 3.7 CGP64222 Peptoid/Peptide hybrid

3.4.6 β-peptides, oligocarbamates, and oligoureas

In 2003, Rana and coworkers reported the construction of a series of β-peptide homologues of the Arg-rich region of Tat 47-57 (Figure 3.8).[41] Dissociation constants for these homologues were measured by fluorescence anisotropy as K_d values for similar peptides can be significantly affected by varying salt concentrations and experimental conditions.[42,43]

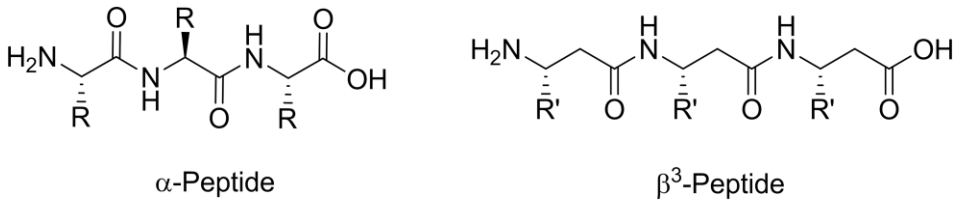


Figure 3.8 α-peptide vs. β^3-peptide structure

An 11-residue β-peptide (**2.34**) bound to TAR with nanomolar affinity, although its affinity for wild type TAR was surprisingly 2-fold weaker than its Lys-counterpart

(**2.35**) (Table 3.6). Interestingly, this decrease in binding affinity was also observed when compared to the α-amino acid analogue of **2.36**. Although the affinity for TAR was decreased for β-peptide **2.34**, specificity for the wt TAR was between 5-20 fold greater than any of the other α- or β-peptides. This feature is worth noting since the bulgeless hairpin is a common secondary structural motif in RNA. Thus, specificity for the bulged stem-loop will be essential for any effective therapeutic directed at TAR.

Table 3.6 Peptide affinity and specificity

Compound	Peptide Sequence	K_d (nM) wt TAR	K_d (nM) bulgeless TAR	Specificity for bulge
2.34[a]	β^3-(YGRKKRRQRRR)	29 ± 4	291 ± 68	10
2.35[a]	β^3-(YG**K**KK**KK**Q**KKK**)	16 ± 2	33 ± 2	2
2.36	YGRKKRRQRRR	2.1 ± 0.8	1.1 ± 0.2	0.5
2.37	YG**K**KK**KK**Q**KKK**	32 ± 4	72 ± 8	2

K = Arginine to Lysine mutation. [a] All amino acids are composed of β^3H-amino acids.

In addition to β-peptides, Rana has also reported oligomers formed with oligocarbamate and oligourea linkages positioned throughout the backbone (Figure 3.9).[42,44] The sequences for these oligocarbamate and oligourea homologues were derived from the Arg-rich region of Tat 47-57 in a similar fashion to their β-peptide work.[41] To account for the variation in measured K_d values, dissociation constant ratios (K_{rel}), which represent the binding of the analogue relative to the natural peptide **2.36** (larger constants correspond to higher affinity binding), were reported. Interestingly, the K_{rel} of the β-peptide **2.34** (Table 3.6) is only 0.072, whereas the alteration of the backbone (keeping

the amino acid sequence constant) to the oligocarbamate gives K_{rel} = 0.69 and alteration to the oligourea gives K_{rel} = 7.1.

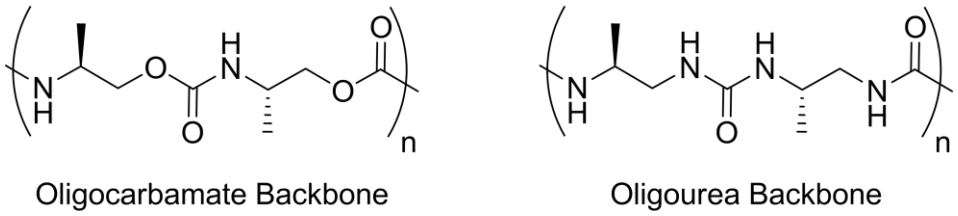


Figure 3.9 Oligocarbamate and Oligourea peptide analogues

3.5 Conclusions

This chapter presented a cursory examination of the structure and replicative cycle of the HIV virus. In addition, a synopsis of the current anti-retroviral therapy (ART) drug categories was presented along with representative examples of each major class. Due to issues associated with viral drug resistance towards ARTs, there is an urgent need for the expansion and examination of new methods that target novel facets of HIV proliferation. Currently, there is significant interest in the development of compounds that interrupt essential RNA/protein interactions by binding directly to RNA tertiary structures. Examples of small and large molecule binding motifs were presented as well as a detailed account of binding motifs presented by peptide and peptidomimetic compounds. This background material was presented as it was relevant to and extremely valuable in efforts towards the completion of the work presented in Chapters 4 and 5 of this dissertation.

3.6 References

1. *Global HIV/AIDS Response: Epidemic Update and Health Sector Progress Towards Universal Access*; World Heath Organization, UNICEF, and UNAIDS. Progress Report 2011: Geneva, Switzerland, **2011**.

2. Rambaut, A.;Posada, D.;Crandall, K. A.; Holmes, E. C. The causes and consequences of HIV evolution *Nat. Rev. Genet.* **2004**, *5*, 52-61.

3. Frankel, A. D.; Young, J. A. T. HIV-1: Fifteen Proteins and an RNA *Annu. Rev. Biochem.* **1998**, *67*, 1-25.

4. Mehellou, Y.; De Clercq, E. Twenty-Six Years of Anti-HIV Drug Discovery: Where Do We Stand and Where Do We Go? *J. Med. Chem.* **2009**, *53*, 521-538.

5. Balzarini, J. Metabolism and mechanism of antiretroviral action of purine and pyrimidine derivatives *Pharm. World. Sci.* . **1994**, *16*, 113-126.

6. Mitsuya, H.;Weinhold, K. J.;Furman, P. A.;St Clair, M. H.;Lehrman, S. N.;Gallo, R. C.;Bolognesi, D.;Barry, D. W.; Broder, S. 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 7096-7100.

7. Pauwels, R. New non-nucleoside reverse transcriptase inhibitors (NNRTIs) in development for the treatment of HIV infections *Curr. Opin. Pharmacol.* **2004**, *4*, 437-446.

8. Ludovici, D. W.;De Corte, B. L.;Kukla, M. J.;Ye, H.;Ho, C. Y.;Lichtenstein, M. A.;Kavash, R. W.;Andries, K.;de Bethune, M. P.;Azijn, H.;Pauwels, R.;Lewi, P. J.;Heeres, J.;Koymans, L. M.;de Jonge, M. R.;Van Aken, K. J.;Daeyaert, F. F.;Das, K.;Arnold, E.; Janssen, P. A. Evolution of anti-HIV drug candidates. Part 3: Diarylpyrimidine (DAPY) analogues *Biorg. Med. Chem. Lett.* **2001**, *11*, 2235-2239.

9. De Clercq, E. New approaches toward anti-HIV chemotherapy *J. Med. Chem.* **2005**, *48*, 1297-1313.

10. Mangum, E. M.; Graham, K. K. Lopinavir-Ritonavir: a new protease inhibitor *Pharmacotherapy.* **2001**, *21*, 1352-1363.

11. Matthews, T.;Salgo, M.;Greenberg, M.;Chung, J.;DeMasi, R.; Bolognesi, D. Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes *Nat. Rev. Drug. Discov.* . **2004**, *3*, 215-225.

12. Wild, C.;Greenwell, T.; Matthews, T. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion *AIDS Res. Hum. Retroviruses.* **1993**, *9*, 1051-1053.

13. Berger, E. A.;Murphy, P. M.; Farber, J. M. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease *Annu. Rev. Immunol.* **1999**, *17*, 657-700.

14. Dorr, P.;Westby, M.;Dobbs, S.;Griffin, P.;Irvine, B.;Macartney, M.;Mori, J.;Rickett, G.;Smith-Burchnell, C.;Napier, C.;Webster, R.;Armour, D.;Price, D.;Stammen, B.;Wood, A.; Perros, M. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity *Antimicrob. Agents Chemother.* **2005**, *49*, 4721-4732.

15. Pommier, Y.;Johnson, A. A.; Marchand, C. Integrase inhibitors to treat HIV/AIDS *Nat. Rev. Drug Discov.* **2005**, *4*, 236-248.

16. Summa, V.;Petrocchi, A.;Bonelli, F.;Crescenzi, B.;Donghi, M.;Ferrara, M.;Fiore, F.;Gardelli, C.;Gonzalez Paz, O.;Hazuda, D. J.;Jones, P.;Kinzel, O.;Laufer, R.;Monteagudo, E.;Muraglia, E.;Nizi, E.;Orvieto, F.;Pace, P.;Pescatore, G.;Scarpelli, R.;Stillmock, K.;Witmer, M. V.; Rowley, M. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection *J. Med. Chem.* **2008**, *51*, 5843-5855.

17. Fauci, A. S. The AIDS epidemic--considerations for the 21st century *N. Engl. J. Med.* . **1999**, *341*, 1046-1050.

18. Richman, D. D.;Margolis, D. M.;Delaney, M.;Greene, W. C.;Hazuda, D.; Pomerantz, R. J. The challenge of finding a cure for HIV infection *Science.* **2009**, *323*, 1304-1307.

19. Menendez-Arias, L. Molecular basis of human immunodeficiency virus drug resistance: an update *Antiviral Res.* **2010**, *85*, 210-231.

20. Thomas, J. R.; Hergenrother, P. J. Targeting RNA with Small Molecules *Chem. Rev.* **2008**, *108*, 1171-1224.

21. Batey, R. T.;Rambo, R. P.; Doudna, J. A. Tertiary Motifs in RNA Structure and Folding *Angew. Chem.* **1999**, *38*, 2326-2343.

22. Lipinski, C. A.;Lombardo, F.;Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings *Adv. Drug Del. Rev.* **2001**, *46*, 3-26.

23. Luedtke, N. W.;Liu, Q.; Tor, Y. RNA-ligand interactions: affinity and specificity of aminoglycoside dimers and acridine conjugates to the HIV-1 Rev response element *Biochemistry.* **2003**, *42*, 11391-11403.

24. Chaloin, L.;Smagulova, F.;Hariton-Gazal, E.;Briant, L.;Loyter, A.; Devaux, C. Potent inhibition of HIV-1 replication by backbone cyclic peptides bearing the Rev arginine rich motif *J. Biomed. Sci.* **2007**, *14*, 565-584.

25. Ye, X.;Kumar, R. A.; Patel, D. J. Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex *Chem. Biol.* **1995**, *2*, 827-840.

26. Puglisi, J. D.;Chen, L.;Blanchard, S.; Frankel, A. D. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex *Science.* **1995**, *270*, 1200-1203.

27. Favre, M., Moehle, K, Jiang, L, Pfeiffer, B, Robinson, J Structural Mimicry of Canonical Conformations in Antibody Hypervariable Loops Using Cyclic Peptides Containing a Heterochiral Diproline Template *J. Am. Chem. Soc.* **1999**, *121*, 2679-2685.

28. Athanassiou, Z.;Dias, R. L.;Moehle, K.;Dobson, N.;Varani, G.; Robinson, J. A. Structural mimicry of retroviral tat proteins by constrained beta-hairpin peptidomimetics: ligands with high affinity and selectivity for viral TAR RNA regulatory elements *J. Am. Chem. Soc.* **2004**, *126*, 6906-6913.

29. Athanassiou, Z.;Patora, K.;Dias, R. L. A.;Moehle, K.;Robinson, J. A.; Varani, G. Structure-Guided Peptidomimetic Design Leads to Nanomolar B-Hairpin Inhibitors of the Tat-TAR Interaction of Bovine Immunodeficiency Virus *Biochemistry.* **2006**, *46*, 741-751.

30. Davidson, A.;Leeper, T. C.;Athanassiou, Z.;Patora-Komisarska, K.;Karn, J.;Robinson, J. A.; Varani, G. Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11931-11936.

31. Lalonde, M. S.;Lobritz, M. A.;Ratcliff, A.;Chamanian, M.;Athanassiou, Z.;Tyagi, M.;Wong, J.;Robinson, J. A.;Karn, J.;Varani, G.; Arts, E. J. Inhibition of both HIV-1 reverse transcription and gene expression by a cyclic peptide that binds the Tat-transactivating response element (TAR) RNA *PLoS Path.* **2011**, *7*, 1-17.

32. Hyun, S.;Kim, H. J.;Lee, N. J.;Lee, K. H.;Lee, Y.;Ahn, D. R.;Kim, K.;Jeong, S.; Yu, J. Œ±-Helical Peptide Containing N,N-Dimethyl Lysine Residues Displays Low-Nanomolar and Highly Specific Binding to RRE RNA *J. Am. Chem. Soc.* **2007**, *129*, 4514-4515.

33. Tan, R.; Frankel, A. D. A novel glutamine-RNA interaction identified by screening libraries in mammalian cells *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 4247-4252.

34. Liu, Q.; Dreyfuss, G. In vivo and in vitro arginine methylation of RNA-binding proteins *Mol. Cell. Biol.* **1995**, *15*, 2800-2808.

35. Yeongran Lee, S. H., Hyun†Jin Kim, Jaehoon Yu, Amphiphilic Helical Peptides Containing Two Acridine Moieties Display Picomolar Affinity toward HIV-1 RRE and TAR13 *Angew. Chem. Int. Ed.* **2008**, *47*, 134-137.

36. Davidson, A.;Patora-Komisarska, K.;Robinson, J. A.; Varani, G. Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein *Nucleic Acids Res.* **2011**, *39*, 248-256.

37. Mills, N. L.;Daugherty, M. D.;Frankel, A. D.; Guy, R. K. An Œ±-Helical Peptidomimetic Inhibitor of the HIV-1 Rev-RRE Interaction *J. Am. Chem. Soc.* **2006**, *128*, 3496-3497.

38. Wang, D.;Iera, J.;Baker, H.;Hogan, P.;Ptak, R.;Yang, L.;Hartman, T.;Buckheit Jr, R. W.;Desjardins, A.;Yang, A.;Legault, P.;Yedavalli, V.;Jeang, K.-T.; Appella, D. H. Multivalent binding oligomers inhibit HIV Tat-TAR interaction critical for viral replication *Biorg. Med. Chem. Lett.* **2009**, *19*, 6893-6897.

39. Hamy, F.;Felder, E. R.;Heizmann, G.;Lazdins, J.;Aboul-ela, F.;Varani, G.;Karn, J.; Klimkait, T. An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 3548-3553.

40. Wyatt, J. R.;Vickers, T. A.;Roberson, J. L.;Buckheit, R. W., Jr.;Klimkait, T.;DeBaets, E.;Davis, P. W.;Rayner, B.;Imbach, J. L.; Ecker, D. J. Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 1356-1360.

41. Gelman, M. A.;Richter, S.;Cao, H.;Umezawa, N.;Gellman, S. H.; Rana, T. M. Selective Binding of TAR RNA by a Tat-Derived B-Peptide *Org. Lett.* **2003**, *5*, 3563-3565.

42. Rana, T. M., Huq, I., Wang, X. HIV-1 TAR RNA Recognition by Unnatural Biopolymer *J. Amer. Chem. Soc.* **1997**, *119*, 6444-6445.

43. Calnan, B. J.;Biancalana, S.;Hudson, D.; Frankel, A. D. Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition *Genes Dev.* **1991**, *5*, 201-210.

44. Rana, T. M., Huq, I., Tamilarasu, N. High Affinity and Specific Binding of HIV-1 TAR RNA by a Tat-Derived Oligourea *J. Am. Chem. Soc.* **1999**, *121*, 1597-1598.

Chapter 4 Facile analysis and sequencing of linear and branched peptide boronic acids by MALDI mass spectrometry

Contributions

This chapter represents a modified version of a published article focusing on the development of a methodology for the analysis and sequencing of peptide boronic acids by MALDI mass spectrometry. Contributions from the authors are as follows: Jason Crumpton (author of this dissertation) performed the synthesis of the modified Fmoc-L-borono-phenylalanine amino acid (F_{BPA}) used for peptide synthesis as well as significant intellectual contributions to experimental design, acquisition of MALDI spectra, and the writing and editing of the manuscript. Synthesis of the majority of the peptides used for this study was performed by Wenyu Zhang. Dr. Webster Santos made significant contributions to the writing and editing of the manuscript as well as aiding in the experimental design. This chapter is reprinted (adapted) with permission from the reference below.

Crumpton, J. B.;Zhang, W. Y.; Santos, W. L. Facile Analysis and Sequencing of Linear and Branched Peptide Boronic Acids by MALDI Mass Spectrometry *Anal. Chem.* **2011**, *83*, 3548-3554.

4.1 Introduction

4.1.1 Peptides incorporating boron as therapeutics

An increasing number of boron containing peptides are being investigated for their potential use in the treatment of various diseases. For example, peptides incorporating boronic acids have been utilized to target the proteasome,[1,2] HIV-1 protease,[3,4] thrombin,[5,6] and human ClpXP.[7] Additionally, a wide variety of artificial boron containing lectins (boronolectins) are being developed, taking advantage of the strong complexation of boronic acids with 1,2- and 1,3- diols.[8] These boronolectins may prove capable of mediating biological processes such as immune response,[9,10] fertilization,[11] embryonic development,[12] and apoptosis.[13] Peptides, proteins, and small molecules that feature boronic acid functionality have been utilized as fluorescent reporter groups,[14] sensors for glucose responsive insulin release,[15] and for detection of cancer related glycoproteins.[16]

4.1.2 Mass spectrometry techniques for boronic acid detection

As the uses of organoboron compounds continue to expand in scope, analytical techniques such as mass spectrometry are increasingly essential for compound characterization and structure elucidation. Boronic acid derivatives have been traditionally analyzed using a wide variety of gas-phase ionization and desorption/ionization techniques including direct insertion probe electron ionization (DIP/EI),[17] electron capture chemical ionization (ECCI),[18] liquid secondary ionization (LSI),[18,19] fast atom bombardment (FAB),[20] and electrospray ionization (ESI). However, there are limited examples in the literature concerning the analysis of boronic acids via MALDI-MS.[16,21,22]

Free boronic acids undergo thermally induced dehydrations[7,23,24] or cyclizations to boroxines[20,21,25] via dehydration/trimerization reactions, which can hinder detection by mass spectrometry (Figure 4.1). When dehydration/boroxine formation interferes with MS data analysis, additional steps such as laborious analysis of cyclization products and/or derivatization to the boronic ester may be required. The former has been identified as undesirable for the characterization of peptides incorporating multiple boronic acid functionalities.[26] Conversely, the derivatization of boronic acids with diols and simple sugars to the cyclic boronic esters has proven to reliably eliminate boroxine formation (Figure 4.2).[17,19,27,28]

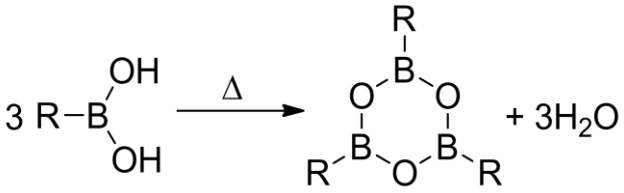


Figure 4.1 Thermally induced boroxine formation

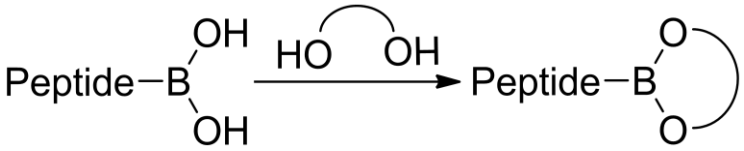


Figure 4.2 Boronic acid peptide functionalization with a 1,2-diol.

The first to report of the successful characterization of small molecular weight boronic ester peptides (Figure 4.3), Haas *et al.* used positive-ion ammonia chemical ionization (CI) as well as positive-ion LSI-MS.[18] Detection of various boronic esters of peptides **4.5** and **4.6** demonstrated that boronic acid peptides were compatible with diol functionalization protocols.[18]

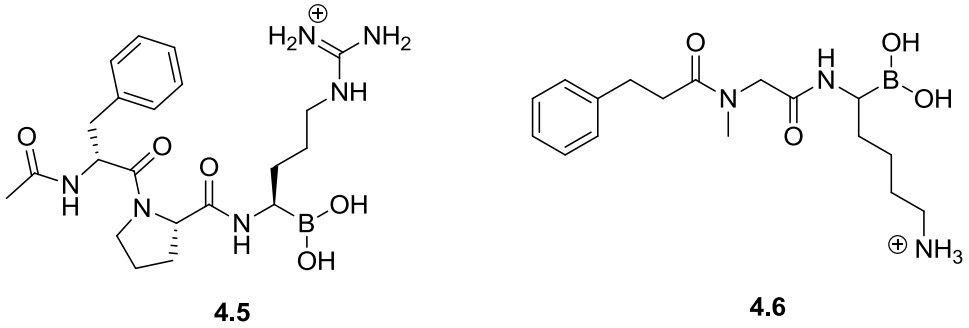
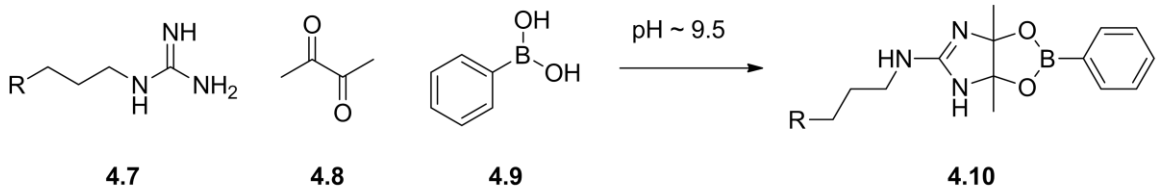


Figure 4.3 Small molecular weight boronic ester peptides[18]

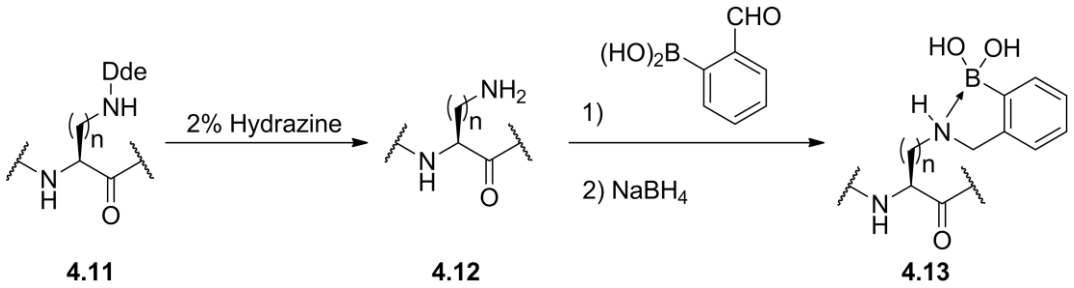
In regards to sequencing peptide boronic acid libraries, Hall has shown that under appropriate conditions, boroxine formation is not necessarily detrimental to library deconvolution.[25] Specifically, when three internal boronic acids are distributed at known positions within the peptide sequence, intramolecular dehydration occurs preferentially over intermolecular processes. This predictable dehydration pattern greatly simplifies library deconvolution.

More recently, Leitner and colleagues reported a technique for the specific labeling of arginine residues (**4.7)** of peptides using 2,3-butanedione (**4.8)** and phenylboronic acid **(4.9)** under alkaline conditions to form bicyclic **4.10** (Scheme 4.1) and subsequent analysis using MALDI-MS.[21] Their work demonstrated that boronic acid-peptide conjugates can be utilized for the identification of peptides based on their mass fingerprint.



Scheme 4.1 Arginine labeling with phenylboronic acid

Finally, Lavigne and Thompson have reported the design and synthesis of a peptide boronolectin (PBL) library that showed promise for the detection of cancer related targets.[16] The boronic acid moiety employed in the library was formed by reductive amination between **4.12** and (2-formylphenyl) boronic acid to yield **4.13** (Scheme 4.2), and MALDI-MS was successfully employed for PBL sequence determination.[16] However, compound **4.13** features an ortho-amino methyl group that can complex to boron, protecting it from trimerization/dehydration processes. Unfortunately, the scope of the detection method presented by Lavigne and Thompson is severely limited to this structural feature for detection by MALDI-MS.



Scheme 4.2 Formation of MS compatible ortho-amino methyl residue 4.13[16]

As part of our long range goal of identifying novel cell permeable molecules that can selectively bind to RNA tertiary structure, branched peptide libraries were initially investigated as RNA ligands.[29] As an extension of this approach, branched peptides containing boronic acid functionalities are particularly interesting, in part because of the possibility of forming Lewis acid-base complexes. Thus, the necessity for a combinatorial library of branched peptide boronic acids (i.e., one bead-one structure) required a rapid, efficient and cost-effective method for deconvolution of hit compounds. Herein, we report a convenient strategy to identify and sequence branched peptides containing boronic acid moieties using MALDI-MS. Derivatization of peptide boronic

acids with 1,2-diols followed by MALDI-MS allowed the successful detection by MS. Fortunately, we also discovered that using DHB as a matrix efficiently converts the peptide boronic acid to a DHB adduct, which provides high quality spectra without the need for prior derivatization. Ultimately, it was determined that the use of DHB for analysis is superior than pretreatment with 1,2-diols. Using this protocol, we successfully identified and sequenced a branched peptide containing five boronic acid moieties released by photocleavage from a single Tentagel bead (Figure 4.4).

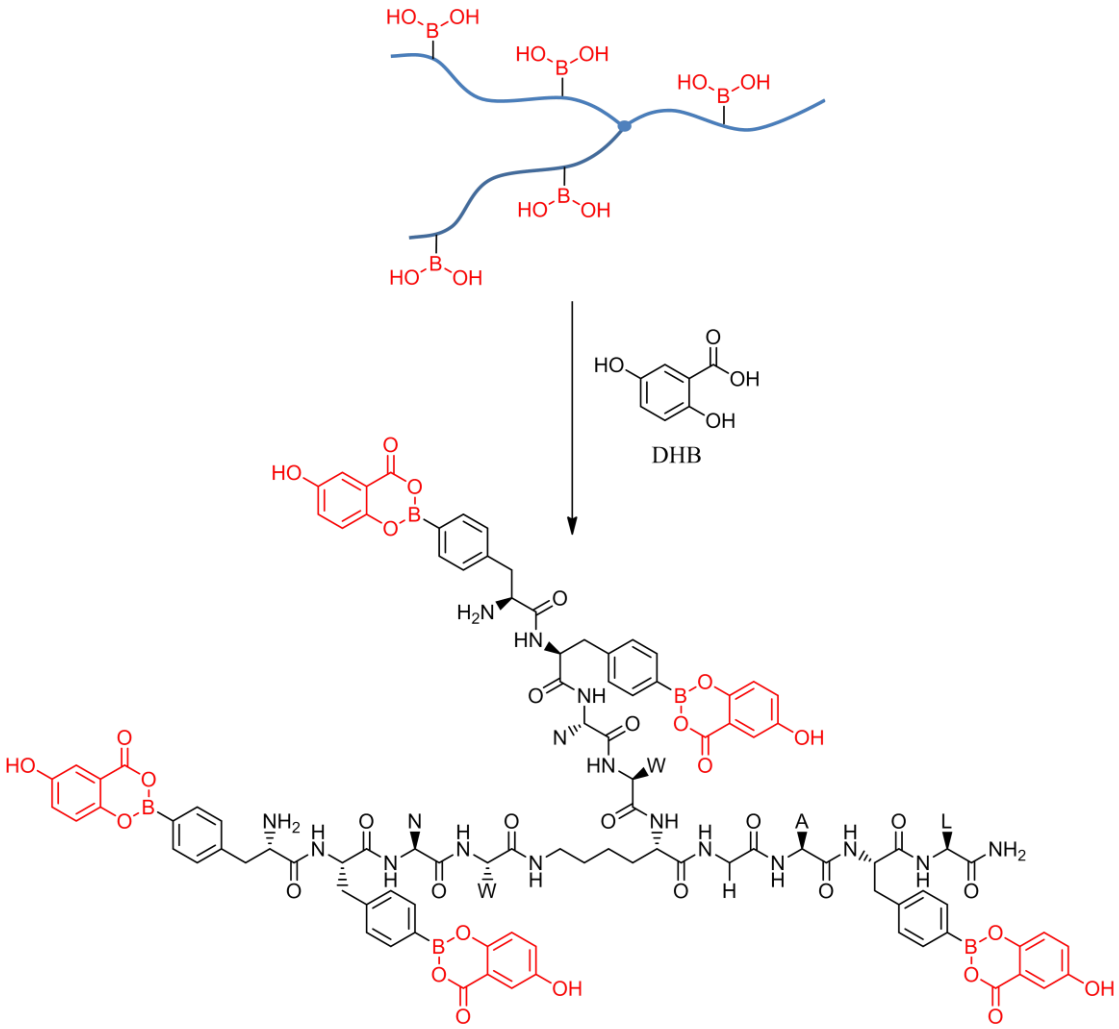


Figure 4.4 Derivatization of peptide boronic acid with DHB.

4.2 Results

4.2.1 Screening of MALDI matrices for peptide boronic acid detection

The suitability of MALDI as a desorption/ionization technique to detect branched peptides containing multiple boronic acids was initially explored by examining a fluorescein isothiocyanate (FITC) labeled peptide containing a single boronic acid, **FLBP1** (Table 4.1). The structure of all matrices utilized for this study can be found in **Figure 4.5**. The molecular ion peak corresponding to the free boronic acid was not observed using CHCA (α-cyano-4-hydroxycinnamic acid) matrix as expected (entry 1). Since modification of boronic acids with 1,2-diols such as pinacol has been reported to reliably improve molecular ion detection, we next combined **FLBP1** with 0.1 M pinacol using a 5 min incubation time (entry 2). Unfortunately, neither the desired molecular ion peak nor any identifiable fragment ions were observed. This result was surprising because previous studies indicate that abbreviated incubation times (10 min) with excess diol were sufficient for boronic ester formation and detection using LSIMS and ammonia CI.[18] It was hypothesized that these results could be attributed to incomplete esterification, ionization interference by excess pinacol or matrix incompatibility. To determine if the absence of the expected molecular ion was due to incomplete esterification, incubation time was increased to 30 min (entry 3), which again failed to detect the expected pinacol protected peptide boronic acid.[30]

Since effective MALDI ionization is highly dependent on effective energy transfer of matrix to analyte, high quantities of competing molecules such as pinacol can interfere with ionization. Therefore, on-plate water washes were performed in an effort to remove excess pinacol and reduce any possible ionization interference; again no $(M+H)^+$ ions were detected. We suspected that matrix incompatibility was responsible for ineffective ionization of the molecular

ion. Therefore, alternative matrices were surveyed to check for compatibility with the esterification protocol. Since HPA (3-hydroxypicolinic acid) is traditionally utilized for DNA detection, it was not surprising that this matrix did not provide the desired molecular ion peak (entry 4). Although the $(M+H)^+$ ion was not observed for DABP (3,4-diaminobenzophenone) and CMBT (5-chloro-2-mercaptobenzothiazole), $(M+H)^+$ minus FITC fragments had excellent relative abundance (entries 5-6). Decreasing the laser intensity to minimize fragmentation of the sensitive FITC moiety did not produce the intact molecular ion. Conversely, CHIQ (1-chloro-4-hydroxyisoquinoline) and DHAP (2',6'-dihydroxyacetophenone) were found to be more efficient at ionization and detection of the $(M+H)^+$ ion without cleaving the FITC label, although 100% relative abundance was not always observed (entries 7-8). An unusual peak at *m/z* 2163.0 appeared with high intensity and percent abundance when DHB (2,5-dihydroxybenzoic acid) was utilized as matrix (entry 9). This peak was 36.8 Da heavier than the expected molecular ion of the pinacol derivatized peptide and upon closer inspection was determined to correspond to the DHB-boronic acid adduct shown in Figure 4.4. Boronic acids have been shown to interact strongly with α-hydroxycarboxylates, and therefore this result is not entirely surprising.[31] The

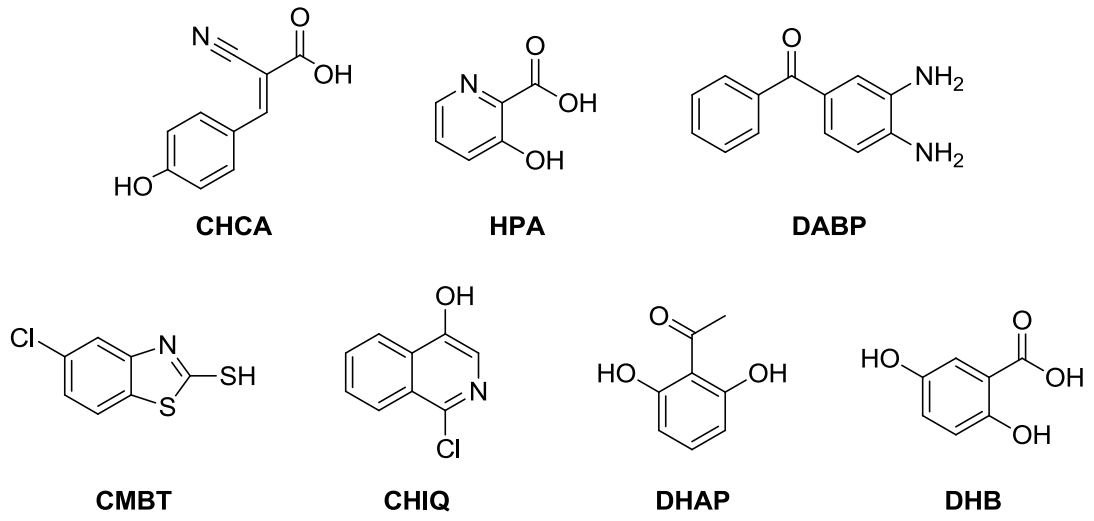


Figure 4.5 Maldi matrices used for study

Table 4.1 Optimization and screening of matrix.[a]

Entry	Additive	Incubation Time (min)	Predicted *m/z*	Matrix	Signal *m/z* (% relative abundance)
1	none	-	1,653.9 / 2,043.0	CHCA	1,304.9 (100), 1,342.8 (74.4), 2,200.1 (28.9)
2	0.1 M pinacol	5	**1,736.0 / 2,125.1**	CHCA	1,025.9 (21.4), 1,028.9 (24.0), 1,174.0 (27.4), 1,304.8 (77.7), 1,342.8 (100)
3	0.1 M pinacol	30	**1,736.0 / 2,125.1**	CHCA	1,304.9 (97.6), 1,342.9 (100)
4	0.1 M pinacol	30	**1,736.0 / 2,125.1**	HPA	No Signal
5	0.1 M pinacol	30	**1,736.0 / 2,125.1**	DABP	1,304.8 (100), **1,736.2** (86.3),[b] 2,093.3 (48.8)
6	0.1 M pinacol	30	**1,736.0 / 2,125.1**	CMBT	1,304.7 (100), **1,736.0** (84.1)[b]
7	0.1 M pinacol	30	**1,736.0 / 2,125.1**	CHIQ	**2,126.2** (100)[c]
8	0.1 M pinacol	30	**1,736.0 / 2,125.1**	DHAP	1,304.8 (27.0), **2,126.2** (100)[c]
9 [a]	0.1 M pinacol	30	**1,736.0 / 2,125.1**	DHB	1,304.7 (100), 2,163.0 (79.2)[d]

[a] Sample utilized was **FLBP1** [FITC-(F$_{BPA}$RW)(FRW)*VRD]; F$_{BPA}$ = 4-borono-L-phenylalanine; * = branching lysine. All samples were subjected to a 10 second on plate water wash before analysis. Signals < 20% abundance not reported. [b] **[(M+H)$^+$ - FITC]** . [c] **(M+H)$^+$**. [d](M+DHB+H)$^+$

Table 4.2 Transesterification of 1,2-diols with DHB.[a]

Entry	Diol	MALDI Matrix	Predicted *m/z* ([(M+diol+H)$^+$-FITC] / (M+diol+H)$^+$)	Signal *m/z* (%relative abundance)
10	Ethylene glycol	CHIQ	**1,680.0** / **2,069.0**	1,304.4 (100), **1,680.0** (28.0), **2,068.7** (42.5)
11[b]	Ethylene glycol	DHB	**1,771.9** / **2,161.0**	1,304.6 (100), **1,771.8** (48.0), 1,813.7 (20.4), **2,161.8** (99.2)
12	Glycerol	CHIQ	**1,710.0** / **2,099.0**	1,304.7 (100), **1,710.0** (42.3), **2,099.0** (58.0)
13[b]	Glycerol	DHB	**1,771.9** / **2,161.0**	1,304.7 (100), **1,772.0** (88.8), **2,162.0** (89.7)
14	Pinacol	CHIQ	**1,736.0** / **2,125.1**	1,304.8 (85.9), **1,736.0** (100), 2,093.1 (33.0), **2,125.1** (45.6)
15[b]	Pinacol	DHB	**1,771.9** / **2,161.0**	1,304.1 (68.7), **1,771.8** (45.2), **2,161.9** (100)
16[b]	-	DHB	**1,771.9** / **2,161.0**	1,304.4 (49.9), **1,771.6** (34.8), **2,162.5** (100)

[a] Sample utilized was **FLBP1**. All diols were prepared as 0.1 M solutions, mixed with FLBP1 in a 1:1 (v:v) ratio, and allowed to incubate for 30 min. All samples were subjected to a 10 second on-plate water wash before analysis. Signals < 20% abundance not reported. [b] Predicted *m/z* is for the DHB ester.

detection of DHB/boronic acid adducts was found to be reproducible over many experiments and suggested that the pinacol moiety was being exchanged with the matrix. Although DHB and HPA both contain α-hydroxycarboxylate moities, no HPA modification was observed for any peptide sample.

4.2.2 Transesterification of three diol protecting groups by DHB

To investigate whether transesterification with DHB was indeed occurring under these conditions, we performed an analysis of the **FLBP1** derivatized peptide with pinacol, ethylene glycol, and glycerol (Table 4.2). First, we confirmed the formation of all 3 esters with the corresponding diols with CHIQ as matrix (entries 10, 12, and 14). As expected, the $(M+H)^+$ ion along with $(M+H)^+$ minus FITC fragments were observed. To test whether the three boronic ester adducts could be exchanged with DHB during the spotting and crystallization process, aliquots of these esters were then spotted using DHB as matrix followed by MALDI-MS analysis. In every case, the boronic ester samples were detected as the DHB ester regardless of the starting diol (entries 11, 13, and 15). These data provided evidence to support our hypothesis that an *in situ* transesterification reaction was occurring. Moreover, when DHB was incorporated as the matrix, the transesterification products were consistently detected in > 95% excess to the original pinacol esters (Figure 4.6). This result implies that the DHB detection protocol is compatible with a wide variety of boronic acid protecting groups.

The success of this procedure provided the impetus to test a possible direct esterification with DHB from a boronic acid sample. When **FLBP1** was spotted, crystallized and shot on the MALDI instrument, the $[\mathbf{FLBP1}\cdot\text{DHB} + H]^+$ ion was detected in excellent relative abundance (entry 16). Thus, this method allowed the

detection of peptide boronic acids without a "pretreatment" step with a 1,2-diol, so all further MALDI experiments were performed using the direct method.

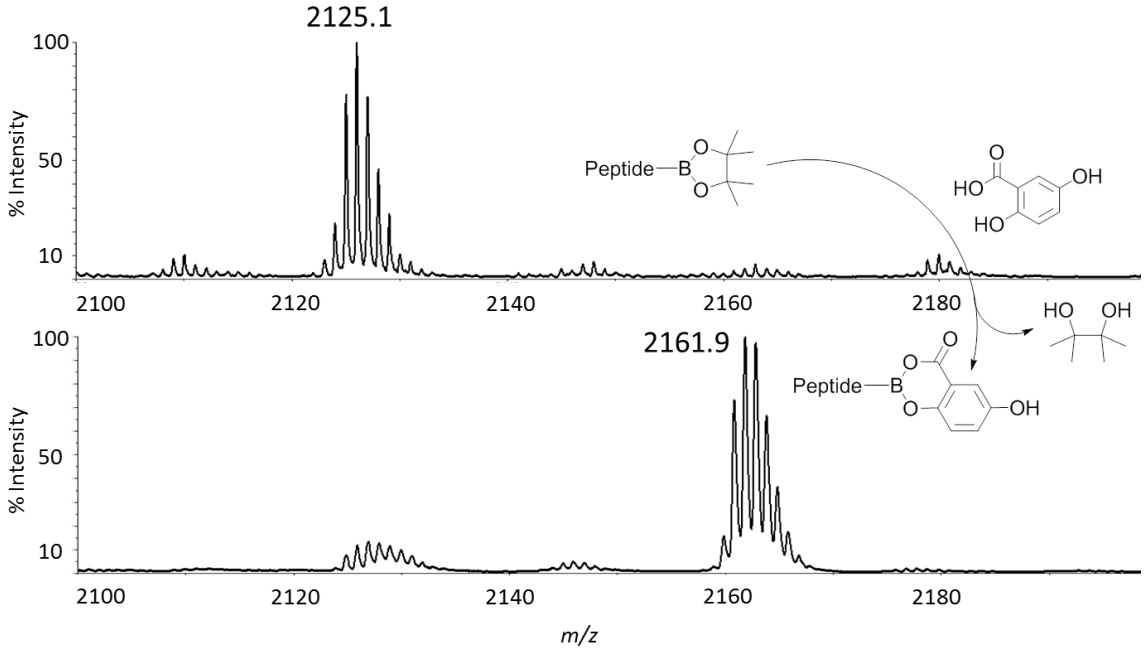


Figure 4.6 Displacement of pinacol protected peptide **FLBP1** by matrix to yield the DHB protected peptide

4.2.3 Scope of detection method

With the optimal protocol in hand, we investigated the robustness of the method with linear as well as branched peptides containing single and multiple boronic acid residues (Table 4.3). Because the boronic acid moiety was introduced onto peptides as an arylboronic acid (i.e. F_{BPA}) in prior samples, it was unclear whether this protocol was sensitive to other boronic acid types. Therefore, several linear peptides conjugated to alkylboronic acids were synthesized and analyzed using DHB (Table 4.3, entries 17-19). To our delight, all three peptides were readily ionized and detected as the DHB adduct with 100% relative abundance, suggesting that the method is compatible with both alkyl- and arylboronic acids. We then increased the number of boronic acid moieties to two

with a more complicated HPLC-purified peptide BP1. We also successfully detected this sample in 100% abundance as the DHB ester (Table 4.3, entry 20). To further test the robustness of the procedure, peptides released from a single Tentagel macrobead (140-170 μm) by photocleavage at 360 nm were analyzed (Table 4.3, entries 21-22).[29] Indeed, these would be the type of samples that will be obtained from screening libraries of branched peptide boronic acids. Unknowns derived from our biochemical assay may contain multiple boronic acid residues at different positions. Applicability of this protocol to libraries containing multiple boronic acid residues requires the effective identification and sequencing of peptides with a high density of boronic acid groups. Thus, a peptide with five BPA units was synthesized, photocleaved, and successfully detected from a single bead (Table 4.3, entry 23). This data suggests that *in situ* DHB transesterification is compatible with high density boronic acid functionalized peptides, and should be compatible with peptide library deconvolution. To the best of our knowledge, this is the first example of the utilization of MALDI-MS for the detection of polyboronic acid peptides as DHB esters.

During the coupling, acidic removal of protecting groups and photocleavage of peptides from the resin, it is important that the boronic acid functionality remain intact, and evidence that boron is present in the peptide is critical. Naturally occurring boron atoms have an isotopic distribution of 1:4 ^{10}B:^{11}B. This differential in atomic mass can be a basis for determining whether a boron atom is present in our peptides.[18] The MS spectrum for peptide $(RRW)_2$*HAL was missing the lower molecular weight isotopic peak corresponding to the ^{10}B isotope because this peptide does not contain a boron atom (Figure 4.7). These signals were clearly present in **BP1**, whether **BP1** was detected as the

pinacol or DHB adduct. The data suggests that the boronic acid moiety is stable to the conditions of the synthesis and MALDI characterization.

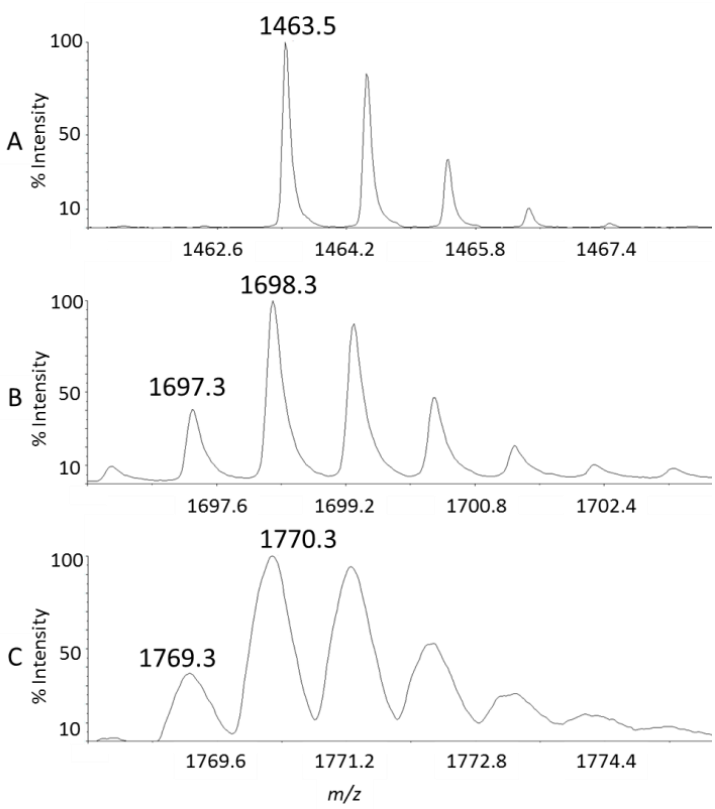
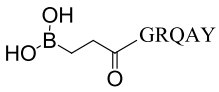
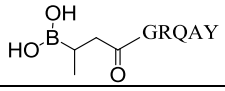
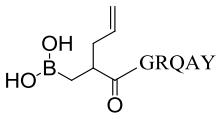


Figure 4.7 Comparison of ^{10}B isotope distribution of (A) arginine analogue [$(RRW)_2$*HAL / MW: 1462.9)], (B) pinacol ester of peptide **BP1** [$(F_{BPA}RW)_2$*HAL / MW: 1697.8)], and (C) DHB ester of peptide **BP1** [$(F_{BPA}RW)_2$*HAL / MW: 1769.8)]

Table 4.3 Detection of DHB ester of peptide boronic acids.

Entry	Sample	Sequence [a]	Calc. Mass of DHB adduct (g/mol)	Signal *m/z* (% relative abundance)
17	**LP1**		**812.3**	776.6 (52.4), 790.6 (34.6), 811.5 (30.5), **812.5 (100)**, 826.5 (68.2)
18	**LP2**		**826.4**	**826.2 (100)**, 840.8 (29.7)
19	**LP3**		**852.4**	408.2 (48.7), 692.4 (73.5), 816.5 (33.2), 851.5 (38.3), **852.4 (100)**, 866.4 (45.8)
20	**BP1**[b]	$(F_{BPA}RW)_2$*HAL	**1,769.8**	1,656.1 (12.4), 1,769.2 (53.2), **1,770.2 (100)**
21	**BP2**[c]	$(F_{BPA}RA)_2$*VRD	**1,588.8**	1,055.1 (74.2), 1,322.2 (100), **1,589.3 (55.5)**
22	**BP3**[c]	$(F_{BPA}RA)_2$*HRF	**1,658.8**	953.8 (41.4), 966.8 (55.9), 1,066.8 (100), 1,124.8 (68.2), 1,391.9 (94.5), 1,513.0 (36.3), **1,659.0 (63.3)**
23	**BP4**[c]	$(F_{BPA}F_{BPA}NW)_2$*HAF_{BPA}L	**2,613.0**	1,647.2 (54.3), 1,793.2 (93.4), 2,093.4 (30.7), 2,305.5 (16.5), 2,467.6 (20.1), **2,613.6 (25.7)**

[a] F_{BPA} = 4-borono-L-phenylalanine; * = branching lysine. [b] Sample was purified by HPLC. [c] Analysis of a small aliquot of unpurified peptide sample obtained from one Tentagel macrobead (theoretical yield < 100 pmol).

4.2.4 MS-MS sequencing of branched peptide boronic acids

To date, sequencing of peptide boronic acids using MS remains challenging. The current method will be valuable, in particular, if sequencing of unknown peptides can be achieved. To determine whether branched peptide boronic acids can be sequenced *de novo*, **BP4** containing five boronic acids was investigated. MS-MS fragmentation was induced on samples prepared using two techniques: prior derivatization with pinacol, and *in situ* adduct formation with DHB. The fragmentation patterns observed for both protocols were similar (data not shown). Hence the major advantage of the direct method is the ease and the elimination of a pre-derivatization step with pinacol. A representative MS-MS spectrum observed is shown in **Figure 4.8**. Major fragments corresponding to y-ions y_7–y_5 as well as b-ions b_5-b_7 were detected in high abundance as shown in Table 4.4.

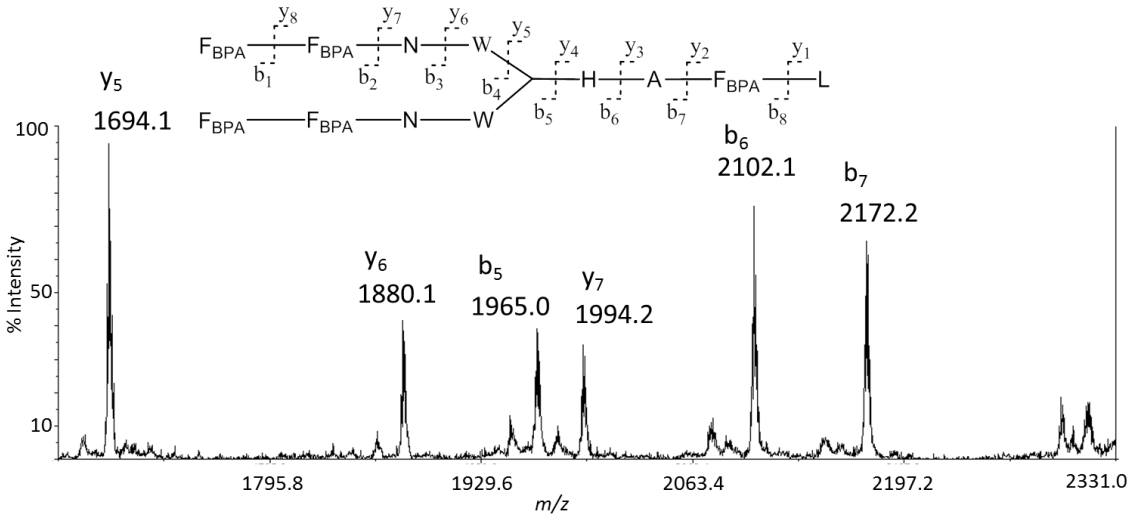
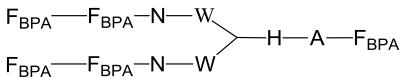
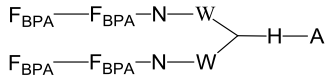
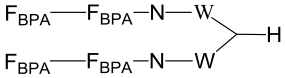
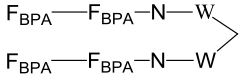
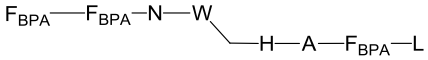
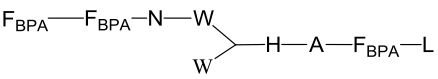
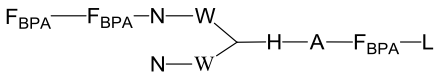
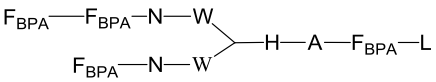


Figure 4.8 MS/MS of DHB modified peptide **BP4**.

Also, minor fragments corresponding to y_4, b_3, and b_4 were detected in low abundance. As expected, fragments encompassing the boronic acid residues were detected as DHB adducts. No signal was observed for fragmentation corresponding to the outermost amide bonds y_8 and b_8 regardless of laser power. Although it is not clear

why the y_8 and b_8 ions were not observed, the suppression or lack of signal is not expected to be a result of the boronic acid functionality, because fragmentation resulting in y_7 ion is observed. For example, fragmentation to provide the b_8 ion has been observed in other sequencing efforts in our group.[35] Since our branched peptide boronic acid libraries are synthesized by split and pool techniques and each amino acid position is restricted to four possible variants,[29] the information from this experiment was sufficient to unambiguously assign a sequence from a theoretical library of 65,536 members.

Table 4.4 MS/MS of DHB modified peptide BP4.

Fragment	Fragment Position	Predicted $(M+H)^+$ *m/z*	Observed $(M+H)^+$ *m/z*
	b_8	2482.8	Not Observed
	b_7	2173.8	2173.2
	b_6	2102.7	2102.1
	b_5	1965.7	1965.0
	y_5	1694.7	1694.1
	y_6	1880.8	1880.1
	y_7	1994.8	1994.2
	y_8	2303.9	Not Observed

4.3 Conclusions

We have developed a simple, rapid and robust on-plate derivatization method for MALDI-MS analysis of linear, branched, and FITC labeled peptide boronic acids using DHB as both the derivatizing agent and matrix. We also demonstrated that trapping the boronic acid moiety as pinacol, ethylene glycol, or glycerol adduct is effective for MALDI-MS analysis of peptide boronic acids. However, the major advantage of using DHB is the removal of a laborious, time-consuming and uneconomical derivatization step.

When 1,2-diols are used for derivatization, removal of excess reagents by a short water wash is necessary to promote effective ionization and subsequent analysis of the resulting spectrum. Under these circumstances, DABP, CMBT, CHIQ, and DHAP can be used as matrices. Indeed, we discovered that DHB appears universally compatible with the boronic acid moiety, because arylboronic acids (such as 4-borono-Phe) and alkylboronic acids were readily detected. Furthermore, the current method allows the identification and sequencing of peptides containing high density boronic acid groups released from single beads. We expect that the current method will be highly valuable not only in the identification of simple boronic acids but also in the deconvolution of peptide boronic acid libraries used in high throughput screening assays.

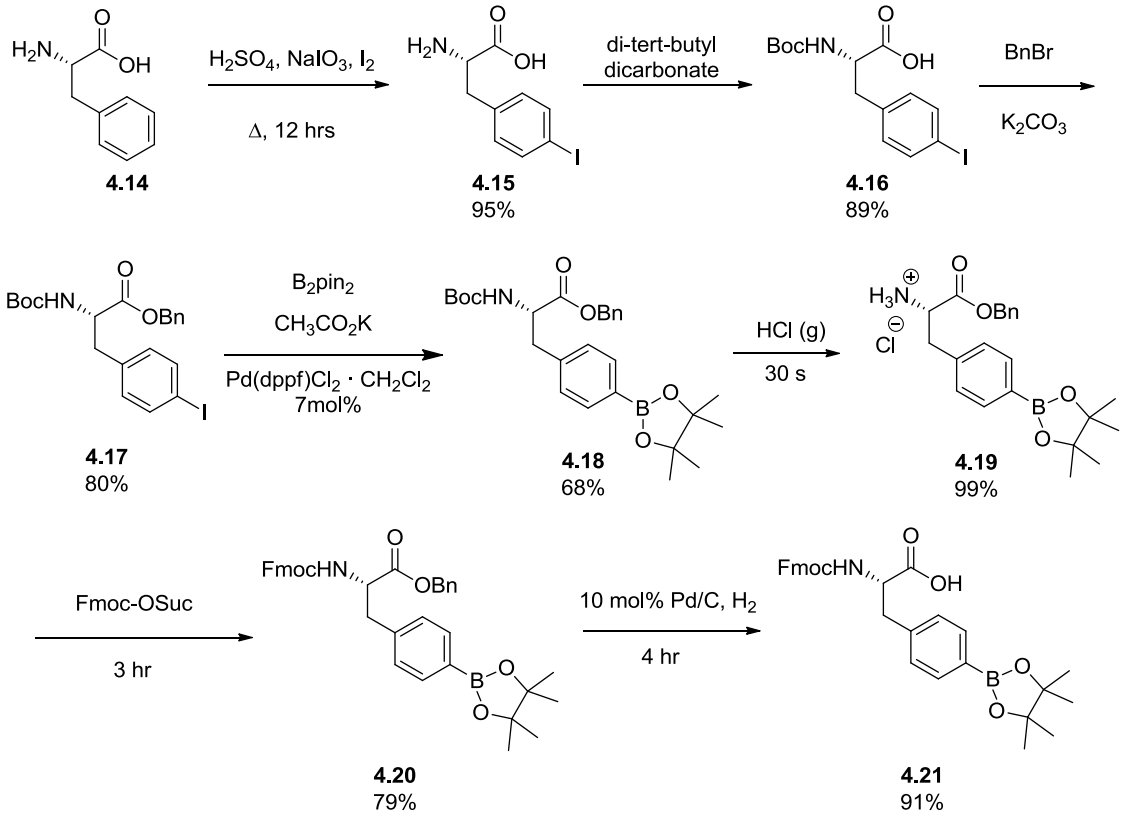
4.4 Experimental

4.4.1 Materials

MALDI matrices 5-chloro-2-mercaptobenzothiazole (CMBT), 1-chloro-4-hydroxyisoquinoline (CHIQ), and 2,5-dihydroxybenzoic acid (DHB) were purchased from Sigma Aldrich. 3-hydroxypicolinic acid (HPA) and 3,4-diaminobenzophenone

(DABP) were purchased from Acros, and 2',6'-dihydroxyacetophenone (DHAP) was purchased from Fluka. All matrices were greater than 98% purity. Milli-Q (MQ) grade H_2O was used for peptide dilution, and HPLC grade acetonitrile was purchased from Fischer Scientific. Unless specified otherwise, all matrices were prepared in a 10 $mg \cdot mL^{-1}$ concentration using a 1:1 (v/v) ACN : H_2O mixture. CHIQ and DHB were prepared as saturated solutions using a 1:1 (v/v) ACN : H_2O mixture.

4.4.2 Synthesis scheme for F_{BPA}



Scheme 4.3 Synthetic Scheme for $F_{BPA.}$ 33% overall yield

4.4.3 Peptide synthesis

Peptides were synthesized using traditional solid phase peptide synthesis (SPPS) using N-α-Fmoc protected L-amino acids (Novabiochem), HCTU ((2-(6-chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate) and DIEA (*N*,*N*-

diisopropylethylamine) in DMF. Solid phase synthesis was performed on a vacuum manifold (Qiagen) outfitted with 3-way Luerlock stopcocks (Sigma) in either Poly-Prep columns or Econo-Pac polypropylene columns (Bio-Rad). The resin was mixed in solution by bubbling argon during all coupling and washing steps. Beads were washed extensively with DMF between reactions, and the couplings were tested for completeness *via* Kaiser test.[32] FITC labeled peptides were prepared following literature protocol.[29] Boron was incorporated into branched peptides as 4-borono-L-phenylalanine (F_{BPA}) [26,33] or into linear peptides on the N-terminus.[7] FLBP1[34] was diluted to a concentration of 370 μM for initial optimization studies.

4.4.4 Generalized procedure for pinacol derivatization

A 0.1 M pinacol solution was prepared with a 10 mM NH_4OAc solution (pH = 9.5) to ensure the formation of the boronic ester. The alkaline pinacol solution was combined with an aliquot of the peptide in a 1:1 (v/v) ratio and allowed to incubate without shaking for 30 min at room temperature. After the incubation period, the MALDI plate was spotted with the matrix of choice, and the sample was then spotted on top of the matrix without mixing. The sample was allowed to dry and 1 μL of MQ water was spotted on top of the sample and allowed to incubate undisturbed for 30 seconds followed by vacuum aspiration. This washing procedure removes unreacted, excess pinacol.

4.4.5 Optimized procedure for DHB derivatization

A saturated solution of DHB was prepared in a 1:1 (v/v) ACN : H_2O solution, which was then spotted on the plate. After drying, the sample (1.0 μL, 300 μM) was spotted on top of the matrix and allowed to dry. Another 1 μL of 1:1 (v/v) ACN/H_2O was

spotted on top of the sample to generate a homogenous solution of the sample and matrix. Samples were then allowed to dry, yielding thin needlelike crystals.

4.4.6 Photocleavage of peptides from the ANP resin and MALDI-TOF sequencing

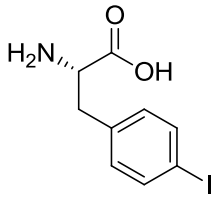
After solid phase peptide synthesis, the F_{BPA} pinacol group was removed via mixing with excess phenyl boronic acid (1.6 M) overnight. Thereafter, amino acid side-chains were deprotected by a 3-hour treatment with 95:2.5:2.5 trifluoroacetic acid/triisopropylsilane/water (v/v). After deprotection, the resin was washed extensively with DMF, DCM, and MeOH before drying and storage at –20 °C protected from the light. A single bead was placed into a clear non-stick 0.5 mL microfuge tube containing 20 μL of 1:1 (v/v) MeOH:H_2O and photocleaved in a foil-lined container by irradiation at 365 nm using a 4 W handheld UV lamp for 1 hour. A small aliquot of the supernatant was utilized for MALDI analysis without any further treatment.

4.4.7 Mass spectrometry

All spectra were acquired with an Applied Biosystems 4800 MALDI-TOF mass spectrometer equipped with a 355 nm Nd:YAG laser. All MALDI spectra were acquired using an Applied Biosystems 123x81 mm Opti-TOF 384 stainless steel plate (Part # 1016629) and acquired in positive reflector mode. A fixed laser intensity of 6500 (arbitrary units) was utilized to acquire 500 shots per spectra for MS samples and 1000-2000 shots per spectra for MS-MS samples. All MS-MS spectra were acquired via post source decay using the MS-MS 1kV positive mode with a selected precursor ion window set to a relative 200.000 FWHM (Full Width Half Maximal) Da resolution. All spectra were processed using Applied Biosystems 4000 series explorer software v.3.5.2., and settings for peak detection were restricted to a S/N threshold of 200.

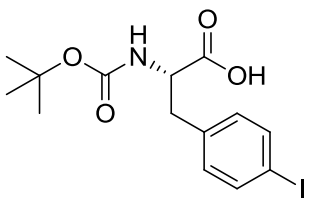
4.4.8 Synthesis of F_{BPA}

(*S*)-2-Amino-3-(4-iodophenyl)propanoic acid **(4.15)**



To a solution of acetic acid was added (*S*)-2-amino-3-phenylpropanoic acid (4.00 g, 24.2 mmol) followed by sulfuric acid (1.29 mL, 24.2 mmol). To the reaction mixture, diiodine (1.23 g, 4.84 mmol) was added followed by sodium iodate (1.01 g, 5.09 mmol). This mixture was refluxed for 12 hrs and the solvent was removed under reduced pressure and re-dissolved in a minimum volume of methanol. The pH of the solution was raised to ~12, which yielded a white precipitate. The mixture was left in the fridge for 4 hours followed by isolation of the white precipitate by vacuum filtration. The title compound **4.15** (6.7 g, 95%) was collected after several washings with methanol as a slightly yellow solid, mp 247.0 - 248.2 ^{o}C. ^{1}H NMR (500 MHz, MeOD) δ 7.59 (d, *J* = 8.3 Hz, 2H), 7.00 (d, *J* = 8.3 Hz, 2H), 3.65 (dd, *J* = 8.5, 4.6 Hz, 1H), 3.14 (dd, *J* = 14.6, 4.6 Hz, 1H), 2.88 (dd, *J* = 14.6, 8.5 Hz, 1H). ^{13}C NMR (126 MHz, MeOD) δ 169.8, 137.6, 135.1, 128.7, 83.9, 53.7, 36.2. HRMS (APCI-) *m/z* calcd for $C_9H_9INO_2$ $[M–H]^-$ 289.9678, found 289.9674.

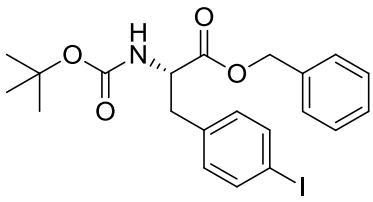
(*S*)-2-((Tert-butoxycarbonyl)amino)-3-(4-iodophenyl)propanoic acid **(4.16)**



To **4.15** (1.00 g, 3.44 mmol) in 3.80 mL of a 1.0 M NaOH solution was added tert-butyl alcohol (2.67 mL, 28.1 mmol). To this biphasic solution was added dropwise a melted solution of di-tert-butyl dicarbonate (0.825 g, 3.78 mmol). The reaction was stirred for 12 hours. The aqueous layer was extracted with 3 x 10 mL of hexanes, which was then back extracted with 3 x 10 mL of K_2CO_3. The combined aqueous layers were cooled in an ice

bath followed by drop-wise addition of HCl until the pH was adjusted to ~2.0 followed by extraction with 5 x 10 mL of diethyl ether. The organic layers were dried with sodium sulfate and concentrated under reduced pressure to yield **4.16** (1.2 g, 89%) as a viscous clear oil. The compound can also be observed as a conglomerate of white crystals if washed with hexanes. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.61 (d, *J* = 8.3 Hz, 2H), 7.01 – 6.89 (m, 2H), 5.02 (d, *J* = 8.0 Hz, 1H), 4.59 (dd, *J* = 12.8, 6.0 Hz, 1H, major), 4.36 (dd, *J* = 11.6, 7.3 Hz, 1H, minor), 3.23 – 3.06 (m, 1H), 3.00 (dd, *J* = 13.8, 6.1 Hz, 1H, major), 2.85 (dd, *J* = 13.0, 8.9 Hz, 1H, minor), 1.41 (s, 9H, major), 1.28 (s, 9H, minor). ^{13}C NMR (126 MHz, $CDCl_3$) δ 175.9 (major), 175.4 (minor), 156.4 (minor), 155.3 (major), 137.7, 136.2 (minor), 135.5 (major), 131.6 (minor), 131.4 (major), 92.7, 81.8 (minor), 80.5 (major), 55.7 (minor), 54.1 (major), 38.9 (minor), 37.4 (major), 28.3 (major), 28.0 (minor). HRMS (ESI+) *m/z* calcd for $C_{14}H_{18}INO_4$ $[M+Na]^+$ 414.0173, found 414.0170.

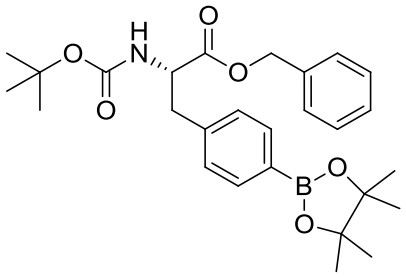
(*S*)-Benzyl 2-((tert-butoxycarbonyl)amino)-3-(4-iodophenyl)propanoate **(4.17)**



To a round bottom flask was added **4.16** (3.38 g, 8.65 mmol) dissolved in 30 mL of DMF. To this solution was added K_2CO_3 (1.20 g, 8.65 mmol), which was marginally soluble in the solvent. To the reaction mixture was added benzyl bromide (2.82 g, 8.65 mmol). The reaction mixture was stirred at room temperature for 6 hours. The solution was filtered, and poured into 30 mL of LiBr. The organic compound was extracted with 4 x 30 mL of diethyl ether, and the organic layer was washed with 1 x 30 mL brine. The organic layer was concentrated, and the crude product was purified on silica gel and concentrated to yield **4.17** as a chunky white solid (3.3 g, 80%). TLC: 5.5:1 /

Hexanes:EtOAc, Rf 0.4. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.51 (d, *J* = 8.1 Hz, 2H), 7.39 – 7.35 (m, 3H), 7.28 – 7.25 (m, 2H), 6.74 (d, *J* = 8.0 Hz, 2H), 5.17 (d, *J* = 12.1 Hz, 1H), 5.07 (d, *J* = 12.1 Hz, 1H), 4.97 (d, *J* = 8.0 Hz, 1H), 4.63 – 4.55 (m, 1H), 3.03 (dd, *J* = 13.7, 6.0 Hz, 1H), 2.98 (dd, *J* = 13.7, 5.7 Hz, 1H), 1.41 (s, 9H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.5, 155.1, 137.6, 135.6, 135.1, 131.5, 128.8, 92.6, 80.2, 67.3, 54.3, 37.9, 28.4. ESI+ *m/z* calcd for $C_{21}H_{24}INO_4$ $[M+Na]^+$ 504.1, found 503.9

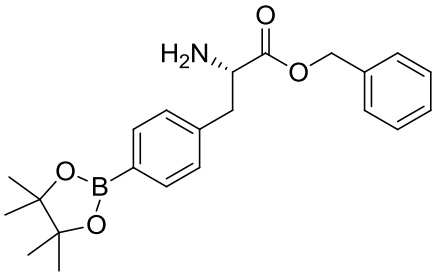
(S)-Benzyl 2-((tert-butoxycarbonyl)amino)-3-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)propanoate **(4.18)**



All glassware was flame dried before reaction. To a flask was added dichloro 1,1'-bis(diphenylphosphino)ferrocene palladium (II) (650 mg, 0.789 mmol), potassium acetate (4.65 g, 47.4 mmol), **4.17** (7.60 g, 15.8 mmol), and bis(pinacolato) diboron (4.41 g, 17.4 mmol). The flask was flushed with nitrogen, 150 mL of dry DMF was added, and the reaction mixture was heated to 80 ^{o}C with vigorous stirring for 12 hours. The sample was diluted with 150 mL of LiBr and extracted with 5 x 50 mL of diethyl ether. The organic layer was washed with 1 x 50 mL of LiBr and 1 x 50 mL of brine. The crude product was purified on silica, and the purified fractions were combined and concentrated under reduced pressure to yield **4.18** as a white foamy solid (5.2 g, 68%). TLC: 5.5:1 / Hexanes:EtOAc, Rf 0.3. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.69 (d, *J* = 7.6 Hz, 2H), 7.37 – 7.31 (m, 3H), 7.29 – 7.25 (m, 2H), 7.05 (d, *J* = 7.5 Hz, 2H), 5.15 (d, *J* = 12.2 Hz, 1H), 5.08 (d, *J* = 12.2 Hz, 1H), 4.99 (d, *J* = 8.1 Hz, 1H), 4.66 – 4.59 (m, 1H), 3.16 – 3.04 (m, 2H), 1.41 (s, 9H), 1.34 (s, 12H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.7, 155.2, 139.2, 135.3, 135.1, 128.9, 128.7, 128.59, 128.55, 83.9, 80.0,

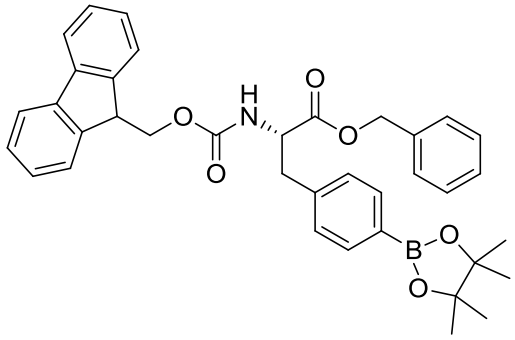
67.2, 54.5, 38.4, 28.4, 25.0. ^{11}B NMR (160 MHz, $CDCl_3$) δ 31.0. HRMS (ESI+) *m/z* calcd for $C_{27}H_{36}BNO_6$ $[M+Na]^+$ 504.2528, found 504.2514

(*S*)-Benzyl 2-amino-3-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl) phenyl)propanoate **(4.19)**



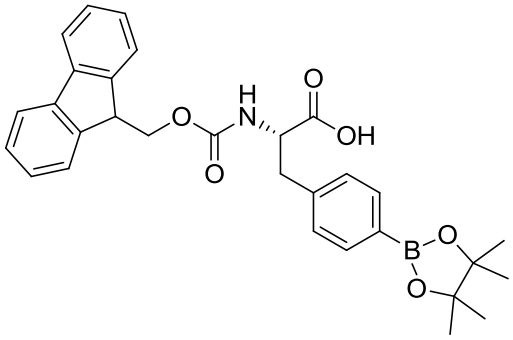
Compound **4.18** (5.05 g, 10.5 mmol) was dissolved in 50 mL of dry CH_3CN, and HCl gas was passed through the stirred solution for 60 seconds. Upon TLC analysis, it was determined that all of the starting material was consumed, yielding an intense UV active spot at baseline. The solvent was removed under reduced pressure, and the product was neutralized with K_2CO_3 followed by extraction with 3 x 50 mL of EtOAc. The organic layer was washed with 1 x 50 mL of brine, and the organic layer was concentrated under reduced pressure. The product, **4.19** (4.0 g, 99%) was used with no further purification. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (d, *J* = 7.9 Hz, 2H), 7.36 – 7.30 (m, 3H), 7.30 – 7.24 (m, 2H), 7.16 (d, *J* = 7.9 Hz, 2H), 5.13 (d, *J* = 12.2 Hz, 1H), 5.10 (d, *J* = 12.2 Hz, 1H), 3.77 (dd, *J* = 7.5, 5.7 Hz, 1H), 3.08 (dd, *J* = 13.4, 5.7 Hz, 1H), 2.90 (dd, *J* = 13.4, 7.5 Hz, 1H), 1.34 (s, 12H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.8, 140.4, 135.5, 135.1, 128.7, 128.6, 128.4, 128.4, 83.7, 66.7, 55.8, 41.3, 24.9. ^{11}B NMR (160 MHz, $CDCl_3$) δ 30.7, 22.2. HRMS (ESI+) *m/z* calcd for $C_{22}H_{28}BNO_4$ $[M+H]^+$ 382.2184, found 382.2150

(*S*)-Benzyl 2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)propanoate **(4.20)**



To a rb flask containing **4.19** (4.00 g, 10.5 mmol) was added 50 mL of DCM. To the solution was added Fmoc-O-succinimide (3.89 g, 11.5 mmol). The reaction was allowed to stir for 4 hours and diluted to 100 mL with DCM. The mixture was washed with 2 x 50 mL of brine, dried with Na_2SO_4, and concentrated under reduced pressure. The crude mixture was purified on silica gel to yield **4.20** as a foamy solid (5.0 g, 79%), mp 53.0 - 54.2 ^{o}C. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.78 (d, *J* = 7.6 Hz, 2H), 7.74 (d, *J* = 7.7 Hz, 2H), 7.57 (t, *J* = 7.6 Hz, 2H), 7.42 (t, *J* = 7.4 Hz, 2H), 7.39 – 7.26 (m, 7H), 7.08 (d, *J* = 7.6 Hz, 2H), 5.36 (d, *J* = 8.3 Hz, 1H), 5.19 (d, *J* = 12.2 Hz, 1H), 5.14 (d, *J* = 12.2 Hz, 1H), 4.81 – 4.72 (m, 1H), 4.43 (dd, *J* = 10.6, 7.1 Hz, 1H), 4.36 (dd, *J* = 10.6, 7.1 Hz, 1H), 4.22 (t, *J* = 7.1 Hz, 1H), 3.23 – 3.08 (m, 2H), 1.37 (s, 12H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.4, 155.7, 143.9, 141.4, 139.0, 135.2, 135.1, 128.9, 128.8, 128.7, 128.7, 127.8, 127.2, 125.2, 120.1, 83.9, 67.4, 67.2, 54.9, 47.2, 38.5, 25.0. ^{11}B NMR (160 MHz, $CDCl_3$) δ 30.8. HRMS (ESI+) *m/z* calcd for $C_{37}H_{39}BNO_6$ $[M+H]^+$ 604.2865, found 604.2901.

N-(9-Fluorenylmethoxycarbonyl)-4-pinacolatoborono-L-phenylalanine **(4.21)**



Compound **4.20** (4.70 g, 7.78 mmol) was dissolved in 100 mL of methanol followed by the addition of 10% palladium on carbon (0.300 g, 0.362 mmol). The flask was evacuated with nitrogen and fitted with a hydrogen balloon apparatus. The black solution was stirred for 4 hours upon which TLC indicated complete consumption of the starting material. The solution was filtered twice over a bed of celite and the solution was concentrated under reduced pressure. The product **4.21** (3.6 g, 91%) presents as a foamy off-white solid, mp 78.5 - 81.2 ^{o}C. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.82 – 7.73 (m, 4H), 7.57 – 7.49 (m, 2H), 7.39 (t, *J* = 7.4 Hz, 2H), 7.33 – 7.27 (m, 2H), 7.22 (d, *J* = 7.6 Hz, 2H), 5.41 (d, *J* = 8.1 Hz, 1H), 4.76 – 4.68 (m, 1H), 4.41 (dd, *J* = 10.4, 7.3 Hz, 1H), 4.31 (dd, *J* = 10.4, 7.3 Hz, 1H), 4.19 (t, *J* = 7.3 Hz, 1H), 3.27 (dd, *J* = 13.7, 5.0 Hz, 1H), 3.13 (dd, *J* = 13.7, 6.6 Hz, 1H), 1.35 (s, 12H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 175.6, 155.9, 143.7, 141.3, 139.2, 135.2, 128.8, 127.7, 127.1, 125.1, 120.0, 83.9, 67.2, 54.7, 47.0, 37.9, 24.8. ^{11}B NMR (160 MHz, $CDCl_3$) δ 30.8. HRMS (ESI+) *m/z* calcd for $C_{30}H_{32}BNO_6$ $[M+Na]^+$ 536.2213, found 536.2205

4.5 References

1. Adams, J.;Behnke, M.;Chen, S.;Cruickshank, A. A.;Dick, L. R.;Grenier, L.;Klunder, J. M.;Ma, Y. T.;Plamondon, L.; Stein, R. L. Potent and selective inhibitors of the proteasome: dipeptidyl boronic acids *Bioorg. Med. Chem. Lett.* **1998**, *8*, 333-338.

2. Tsubuki, S.;Kawasaki, H.;Saito, Y.;Miyashita, N.;Inomata, M.; Kawashima, S. Purification and characterization of a Z-Leu-Leu-Leu-MCA degrading protease expected to regulate neurite formation: a novel catalytic activity in proteasome *Biochem. Biophys. Res. Commun.* **1993**, *196*, 1195-1201.

3. Prusoff, W.;Lin, T. S.;Pivazyan, A.;Sun, A. S.; Birks, E. Empirical and rational approaches for development of inhibitors of the human immunodeficiency virus--HIV-1 *Pharmacol. Ther.* **1993**, *60*, 315-329.

4. Pivazyan, A. D.;Matteson, D. S.;Fabry-Asztalos, L.;Singh, R. P.;Lin, P. F.;Blair, W.;Guo, K.;Robinson, B.; Prusoff, W. H. Inhibition of HIV-1 protease by a boron-modified polypeptide *Biochem. Pharmacol.* **2000**, *60*, 927-936.

5. Wienand, A.;Ehrhardt, C.;Metternich, R.; Tapparelli, C. Design, synthesis and biological evaluation of selective boron-containing thrombin inhibitors *Bioorg. Med. Chem.* **1999**, *7*, 1295-1307.

6. Fevig, J. M.;Abelman, M. M.;Brittelli, D. R.;Kettner, C. A.;Knabb, R. M.; Weber, P. C. Design and synthesis of ring-constrained boropeptide thrombin inhibitors *Bioorg. Med. Chem. Lett.* **1996**, *6*, 295-300.

7. Knott, K.;Fishovitz, J.;Thorpe, S. B.;Lee, I.; Santos, W. L. N-Terminal peptidic boronic acids selectively inhibit human ClpXP *Org. Biomol. Chem.* **2010**, *8*, 3451-3456.

8. Yan, J.;Fang, H.; Wang, B. Boronolectins and fluorescent boronolectins: an examination of the detailed chemistry issues important for the design *Med. Res. Rev.* **2005**, *25*, 490-520.

9. Lis, H.; Sharon, N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition *Chem. Rev.* **1998**, *98*, 637-674.

10. Weis, W. I.;Taylor, M. E.; Drickamer, K. The C-type lectin superfamily in the immune system *Immunol. Rev.* **1998**, *163*, 19-34.

11. Sinowatz, F.;Topfer-Petersen, E.; Calvete, J. J. Glycobiology of Fertilization *Glycosciences.* **1997**, *595*.

12. Clark, G. F.;Oehninger, S.;Patankar, M. S.;Koistinen, R.;Dell, A.;Morris, H. R.;Koistinen, H.; Seppala, M. A role for glycoconjugates in human development: the human feto-embryonic defence system hypothesis *Hum. Reprod.* **1996**, *11*, 467-473.

13. Fukumori, T.;Takenaka, Y.;Yoshii, T.;Kim, H. R.;Hogan, V.;Inohara, H.;Kagawa, S.; Raz, A. CD29 and CD7 mediate galectin-3-induced type II T-cell apoptosis *Cancer Res.* **2003**, *63*, 8302-8311.

14. Arimori, S.;Bell, M. L.;Oh, C. S.;Frimat, K. A.; James, T. D. Modular fluorescence sensors for saccharides *Chem. Commun. (Camb).* **2001**, 1836-1837.

15. Hoeg-Jensen, T.;Ridderberg, S.;Havelund, S.;Schaffer, L.;Balschmidt, P.;Jonassen, I.;Vedso, P.;Olesen, P. H.; Markussen, J. Insulins with built-in glucose sensors for glucose responsive insulin release *J. Pept. Sci.* **2005**, *11*, 339-346.

16. Zou, Y.;Broughton, D. L.;Bicker, K. L.;Thompson, P. R.; Lavigne, J. J. Peptide borono lectins (PBLs): a new tool for glycomics and cancer diagnostics *Chembiochem.* **2007**, *8*, 2048-2051.

17. Longstaff, C.; Rose, M. E. Derivatization and mass spectrometric investigations of substituted benzeneboronic acids. The use of linked scanning during gas chromatography mass spectrometry *Org. Mass. Spectrom.* . **1982**, *17*, 508-518.

18. Haas, M.;Blom, K.; Schwartz, C. Positive-Ion Analysis of Boropeptides by Chemical Ionization and Liquid Secondary Ionization Mass Spectrometry *Anal. Chem.* **1999**, *71*, 1574-1578.

19. Snow, R. J.;Bachovchin, W. W.;Barton, R. W.;Campbell, S. J.;Coutts, S. J.;Freeman, D. M.;Gutheil, W. G.;Kelly, T. A.;Kennedy, C. A.;Krolikowski, D. A.;Leonard, S. F.;Pargellis, C. A.;Tong, L.; Adams, J. Studies on Proline Boronic Acid Dipeptide Inhibitors of Dipeptidyl Peptidase-Iv - Identification of a Cyclic Species Containing a B-N Bond *J. Am. Chem. Soc.* **1994**, *116*, 10860-10869.

20. Kettner, C.;Mersinger, L.; Knabb, R. The selective inhibition of thrombin by peptides of boroarginine *J. Biol. Chem.* **1990**, *265*, 18289-18297.

21. Leitner, A.;Amon, S.;Rizzi, A.; Lindner, W. Use of the arginine-specific butanedione/phenylboronic acid tag for analysis of peptides and protein digests using matrix-assisted laser desorption/ionization mass spectrometry *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1321-1330.

22. Lin, N.;Yan, J.;Huang, Z.;Altier, C.;Li, M.;Carrasco, N.;Suyemoto, M.;Johnston, L.;Wang, S.;Wang, Q.;Fang, H.;Caton-Williams, J.; Wang, B. Design and synthesis of boronic-acid-labeled thymidine triphosphate for incorporation into DNA *Nucleic Acids Res.* **2007**, *35*, 1222-1229.

23. LeBeau, A. M.;Singh, P.;Isaacs, J. T.; Denmeade, S. R. Potent and selective peptidyl boronic acid inhibitors of the serine protease prostate-specific antigen *Chem. Biol.* **2008**, *15*, 665-674.

24. Sasubilli, R.; Gutheil, W. G. General inverse solid-phase synthesis method for C-terminally modified peptide mimetics *J. Comb. Chem.* **2004**, *6*, 911-915.

25. Manku, S.; Hall, D. Synthesis, Decoding, and Preliminary Screening of a Bead-Supported Split-Pool Library of Triboronic Acid Receptors for Complex Oligosaccharides *Aust. J. Chem.* **2007**, *60*, 824-828.

26. Duggan, P.; Offerman, D. The Preparation of Solid-Supported Peptide Boronic Acids Derived from 4-Borono-$_L$-phenylalanine and their Affinity for Alizarin *Aust. J. Chem.* **2007**, *60*, 829-834.

27. Singhawangcha, S.;Hu, L.-E. C.;Poole, C. F.; Zlatkis, A. Pinacol as a reagent for the characterization of aromatic boronic acids *J. High. Resolut. Chromatogr. Commun.* **1978**, *1*, 304-306.

28. Rose, M. E.;Longstaff, C.; Dean, P. D. G. Gas chromatography of aromatic boronic acids: on-column derivatization *J. Chromatogr.* **1982**, *249*, 174-179.

29. Bryson, D. I.;Zhang, W.;Ray, W. K.; Santos, W. L. Screening of a branched peptide library with HIV-1 TAR RNA *Mol. Biosyst.* **2009**, *5*, 1070-1073.

30. Using CHCA as a matrix tends to be unpredicatable. For example, depending on the peptide in question 0.1M or 1.0 M pinacol may provide the correct molecular ion peak.

31. Gray, C. W., Jr.; Houston, T. A. Boronic acid receptors for alpha-hydroxycarboxylates: high affinity of Shinkai's glucose receptor for tartrate *J. Org. Chem.* **2002**, *67*, 5426-5428.

32. Kaiser, E.;Colescott, R. L.;Bossinger, C. D.; Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides *Anal. Biochem.* **1970**, *34*, 595-598.

33. Malan, C.; Morin, C. A Concise Preparation of 4-Borono-$_L$-phenylalanine ($_L$-BPA) from $_L$-Phenylalanine *J. Org. Chem.* **1998**, *63*, 8019–8020.

34. Details are given in Table 1.

Chapter 5 Deconvolution of a branched boronic acid peptide library

Contributions

This chapter represents a modified version of a manuscript to be submitted for publication, which describes boronic acid modified peptides as RNA binding ligands. Contributions from the authors are as follows: Jason Crumpton (author of this dissertation) performed the synthesis of the modified Fmoc-L-borono-phenylalanine amino acid (F_{BPA}) as well as the N-ε-(4-boronobenzoyl)-L-lysine (K_{BBA}) used for peptide synthesis. The author also acquired all MALDI spectra, coordinated and collaborated with technicians during the COPAS sorting experiment, and made intellectual contributions to the design of the 4.4.4 peptide library as well as significant contributions to the writing and editing of the manuscript. The author also made intellectual contributions to the construction of novel software for the *de novo* sequencing of unknown libraries of branched boronic acid peptides. Jon Bernard made intellectual contributions to and wrote the code for the *de novo* software package. David Bryson made intellectual contributions to the design of the 4.4.4 peptide library and performed all K_d analysis, both by EMSA and Dot-Blot assays. The synthesis and purification of the peptides used for this study was performed by Wenyu Zhang who also made contributions towards the design of the 4.4.4 peptide library. Dr. Webster Santos made contributions to the writing and editing of the manuscript as well as significant contributions to experimental design.

5.1 Introduction

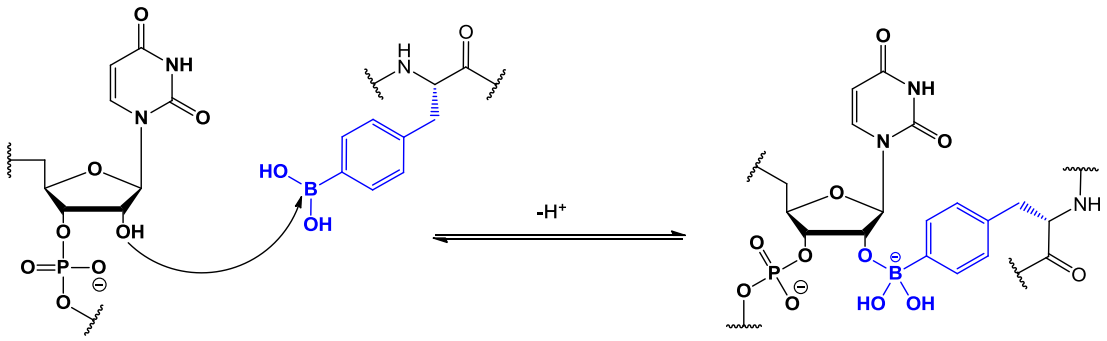
5.1.1 HIV-1 and Rev/RRE

The successful nuclear export of unspliced HIV-1 has been shown to be highly dependent on the Rev binding protein.[1] Specifically, HIV-1 proliferation is fundamentally tied to the interaction of the Rev protein with RRE. Thus, interfering with this interaction has the potential to prevent viral particle assembly. Specifically, targeting of the nuclear export factor CRM1, [2-4] the Rev-RRE interaction, [5-7] and the Rev protein itself [8,9] have emerged as effective approaches for developing anti-viral therapeutics. Indeed, potent binders such as **DB340** [10] have been found to bind with high affinity to the RRE. Unfortunately, high affinity is, more often than not, compromised by low selectivity, which impedes further therapeutic development.

5.1.2 Novel RNA binding modes via boronate complexation

We are interested in elucidating RNA binders that display high selectivity as well as affinity. To pursue this goal, we have decided to expand upon previous work regarding multivalent peptide binding motifs[11] by incorporating boronic acid moieties into peptidic structures. Boron is well-suited for incorporation into peptidic structures as amino acid analogues such as boronophenyl alanine (F_{BPA})[12,13] have been extensively explored in the literature. More importantly, boron is interesting because it can potentially form reversible Lewis acid/base interactions with the 2' hydroxyl of RNA through its vacant p-orbital (Figure 5.1). In fact, exploiting this RNA/boron interaction has already been demonstrated with the development of leucyl tRNA synthetase inhibitors such as the benzoxaborole AN2690.[14-16] Unfortunately, there is no established method for the development of RNA ligands that achieve high affinity while maintaining specificity

towards the sequence of interest. In order to evaluate whether boronic acids could potentially promote specific RNA binding, it is necessary to explore vast combinatorial libraries. Herein, we report the first example of the successful deconvolution of a one bead one compound (OBOC) peptide library containing boronic acid.



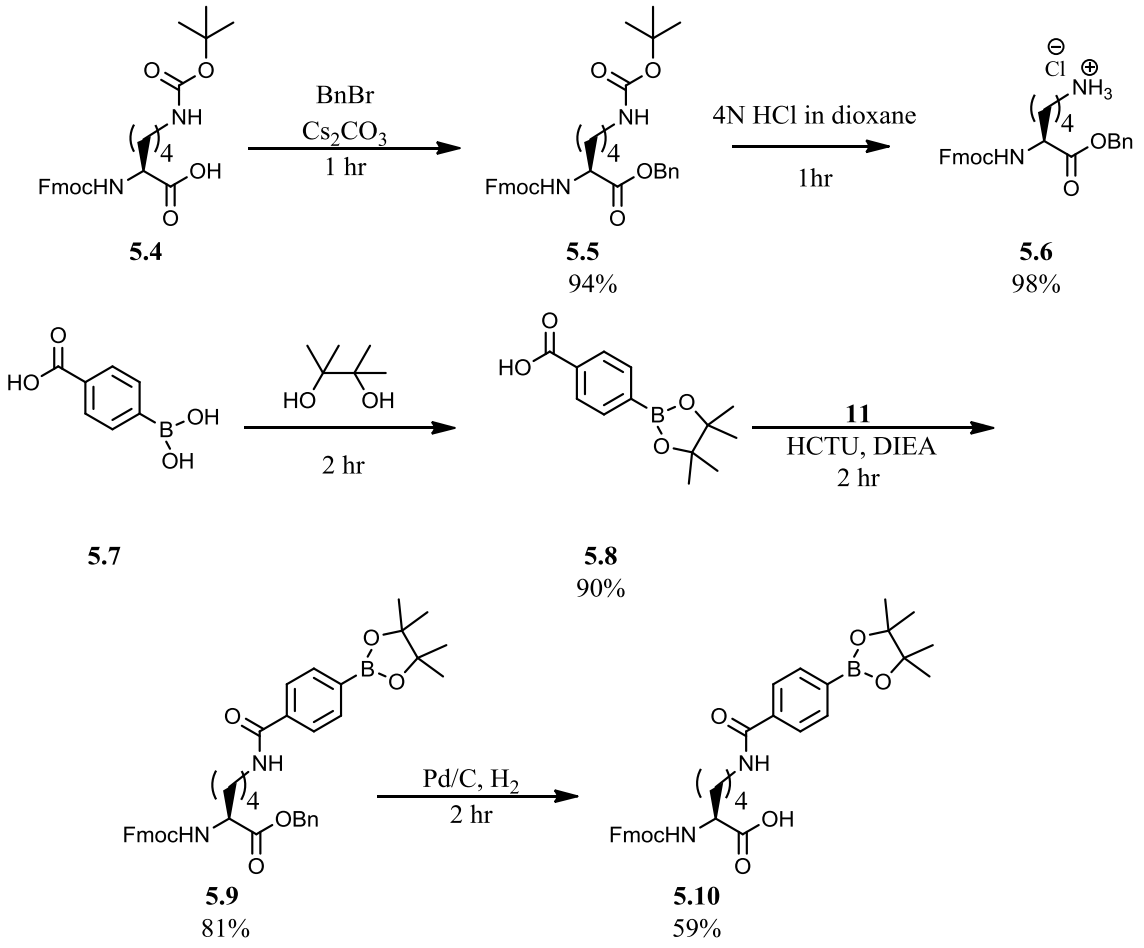
Scheme 5.1 Formation of reversible covalent bond between RNA and F_{BPA} residue

5.2 Results

5.2.1 MALDI analysis of 4.4.4 peptides

We have previously reported a facile method for efficiently detecting boronic acid peptides via matrix assisted laser desorption ionization mass spectrometry (MALDI-MS).[17] It is well-known that free boronic acids are notoriously difficult to detect due to dehydration/cyclization reactions. Fortuitously, we found that the MALDI matrix 2,5-dihydroxy benzoic acid (DHB) was capable of simultaneously acting as both matrix and boronic acid derivatization agent, allowing effective peptide boronic acid detection and eliminating the necessity for pinacol functionalization. Although this protocol was optimized for branched peptides displaying F_{BPA} residues, we found that this technique was robust and capable of handling a wide variety of boronic acids. We have since confirmed the compatibility of this detection protocol with the novel amino acid analogue

N-ε-(4-boronobenzoyl)-L-lysine (K_{BBA}) residue in addition to F_{BPA} residues. This residue was synthesized according to **Scheme 5.2**.



Scheme 5.2 Synthesis Scheme for K_{BBA}. 40% overall yield

Not only are the boronopeptide parent ions detected with high abundance and S/N ratios, but the resulting MS-MS spectra are also high quality with fragmentation corresponding to virtually all possible amide bond fragmentations (**Figure 5.1**). We have also discovered that the detection protocol is capable of successfully functionalizing and detecting peptides composed of mixed (F_{BPA} or K_{BBA}) boronoamino acid residues (Figure 5.2). The MS-MS spectra of these mixed residue samples are detected reproducibly with excellent fragmentation.

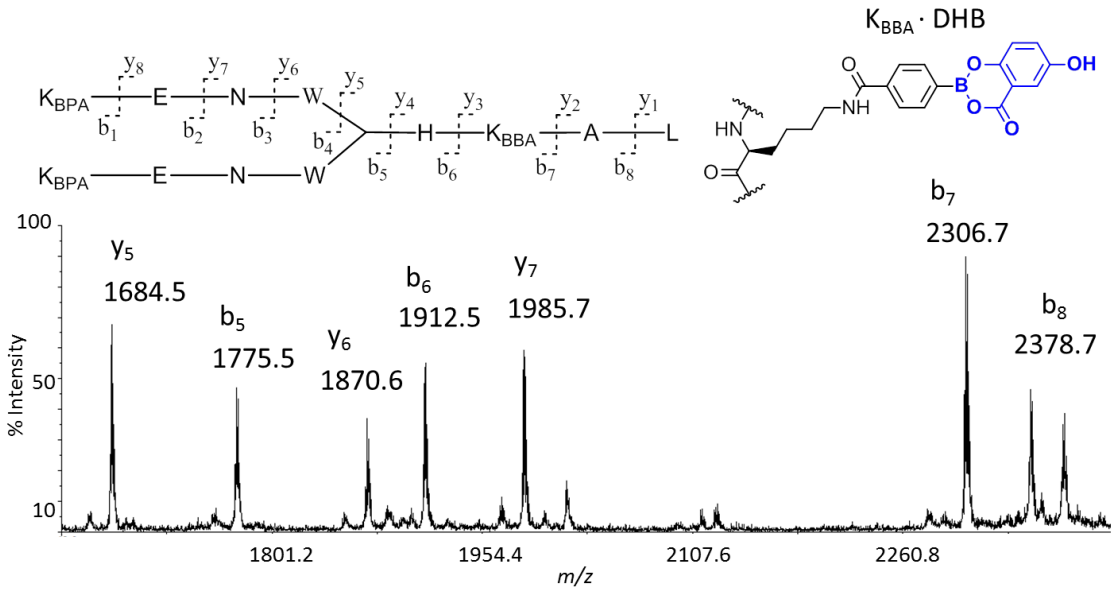


Figure 5.1 MS-MS of 4.4.4 peptide with K_{BBA} residues

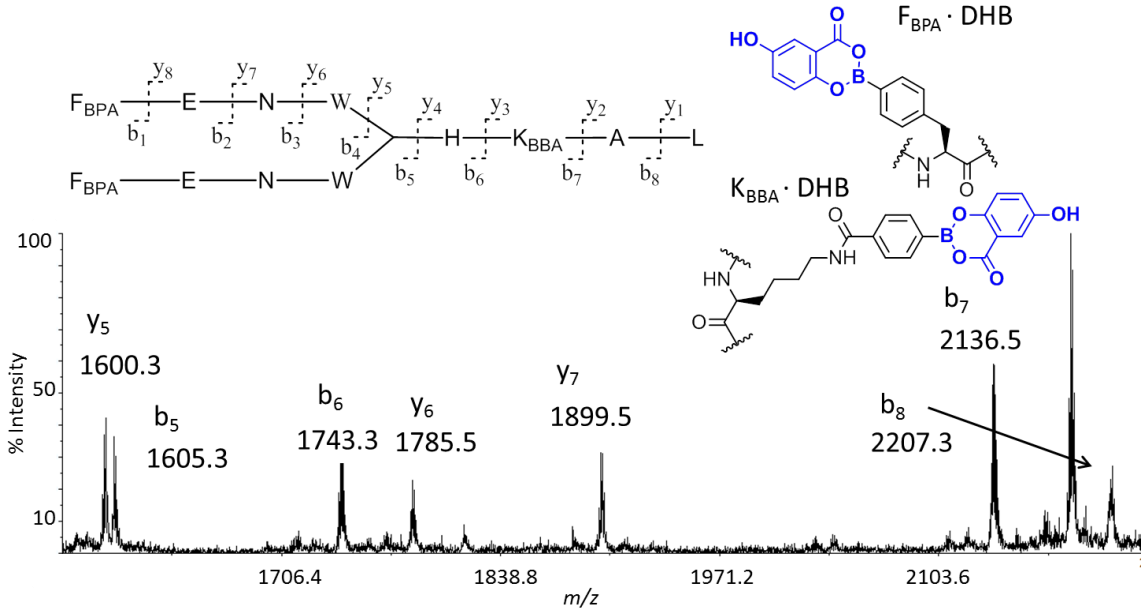


Figure 5.2 MS-MS of 4.4.4 peptide with mixed boronic acids F_{BPA} and K_{BBA}

5.2.2 Limitations of DHB detection protocol

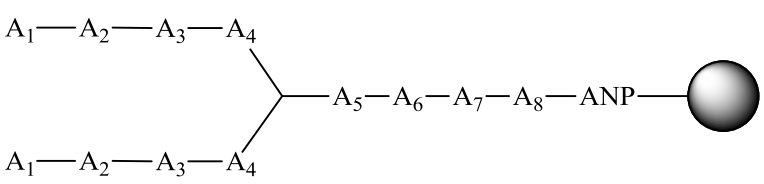
The only instance where the DHB method failed to detect both parent and MS-MS peaks occurs when serine and threonine residues are located n+1 and n-1 to the K_{BBA}.

Interestingly, the presence of alcoholic residues does not impede detection as long as this (n+1)/(n-1) configuration is avoided. Although the full extent of this limitation has not been fully explored, we expect that any combination of natural alcoholic residues flanking the K_{BBA} residue would yield similar results. Fortunately, we have established that parent ions of these troubled sequences are easily detected using the pinacol protection protocol previously outlined.[17]

5.2.3 Design of the 4.4.4 Library

We also previously reported the synthesis of a trivalent branched peptide library with 4 variable amino acids at each position. Each branch contained 3 amino acids, and was thus referred to as a 3.3.3 peptide.[11] This 4,096 (4^6) membered library was screened for affinity towards HIV-1 TAR, and yielded multiple low μM binders.[18] For the second permutation of the branched peptide library, we made several changes to the overall design keeping several goals in mind.

The first goal was to increase the diversity of the branched peptide library. The easiest method for accomplishing this goal was to simply increase the size of the library to a 4.4.4 variant, which would yield 65,536 (4^8) unique branched peptides and increase our previous library's sampling size by an order of magnitude (Figure 5.3).



4.4.4. Library (8 variable positions)

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
Glu	Asn	Asp	Gln	Asn	Asp	Gln	Glu
Thr	Ala	Ser	Leu	Ile	Val	Pro	Ser
Tyr	Phe	His	Trp	Tyr	Trp	His	Phe

K_{BBA} →

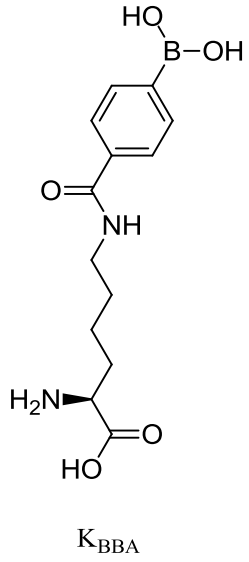


Figure 5.3 Possible amino acids for 4.4.4 library

The second goal was the removal of any potential basic residues (Lys/Arg) from the final peptide structure to minimize non-specific electrostatic interactions with RNA substrates. This goal posed several issues that included the potential for reduced aqueous solubility of the resulting peptides. However, we hypothesized that the inclusion of boronic acid moieties within the peptide provided an elegant solution to simultaneously address solubility as well as selectivity. Boronic acids are known to establish equilibrium in aqueous media between the neutral boronic acid and the charged boronate. Thus, we suspected that this charged boronate would impart increased water solubility over traditional non-basic peptide sequences. Because the pKa of phenyl boronic acid is 8.8,[19] we utilized only the K_{BBA}, which theoretically possesses a depressed pKa and would therefore exist predominantly as the boronate at lower pH levels. Our motivation to include boron was also driven by potential participation in Lewis acid/base binding modes. This type of reversible covalent binding (Scheme 5.1) would theoretically impart increased selectivity of the peptide towards the RNA substrate. Indeed, there is precedent

in the literature for boron being utilized to target the 2'-hydroxyl of tRNA.[15] As shown in Figure 5.4, the K_{BBA} residue was included at positions A_1-A_8. We included various aromatic residues at positions A_1-A_8 which offered potential π-stacking interactions. Carboxylic acid (Asp/Glu), carboxamide (Asn/Gln), and alcoholic (Thr/Ser) residues were included as potential hydrogen bond donor/acceptor possibilities. Finally, hydrophobic residues were included for potential Van der Waals interactions with RNA grooves.[20]

5.2.4 Sorting of one bead one compound library

The 4.4.4 library was synthesized by standard solid phase peptide synthesis (SPPS) utilizing split and pool techniques to generate a 'one bead one compound' combinatorial library. Every potential peptide was synthesized in triplicate. Therefore, to completely screen the 4.4.4 library, ~200,000 beads needed to be screened and evaluated for fluorescence. The first generation peptide library (~12,000 beads) required 4 weeks to screen manually using fluorescence microscopy. A graphic comparison of the number of beads in the two libraries makes obvious the implication that a complete manual sort would have been prohibitively laborious (Figure 5.4).

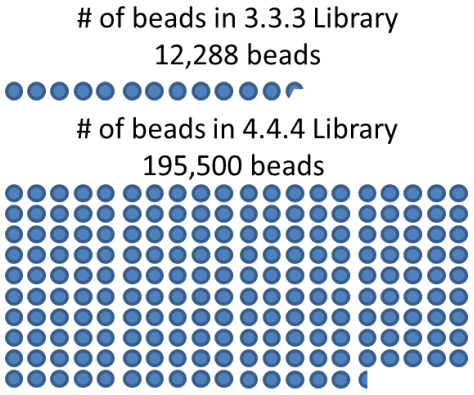


Figure 5.4 Graphic comparison of 3.3.3. to 4.4.4 library

Therefore, we employed the Complex Object Parametric Analyzer and Sorter (COPAS) instrument by Union Biometrica, which is able to rapidly evaluate and collect samples with fluorescence above designated background. With COPAS, we initially screened ~100,000 beads within 2 hours. The resulting scatter plot yielded ~1,500 beads (1.5 %) with intensities above the average background fluorescence (Figure 5.5). We reasoned that beads with higher fluorescence were indicative of higher quantities of fluorescent RRE RNA and were thus indicative of high affinity binders. These collected beads were then evaluated manually using fluorescence microscopy, which yielded 6 beads that were designated significantly above background fluorescence (Figure 5.6). These beads were photocleaved using methods reported in the literature[11] and analyzed by MALDI-MS using DHB as matrix.

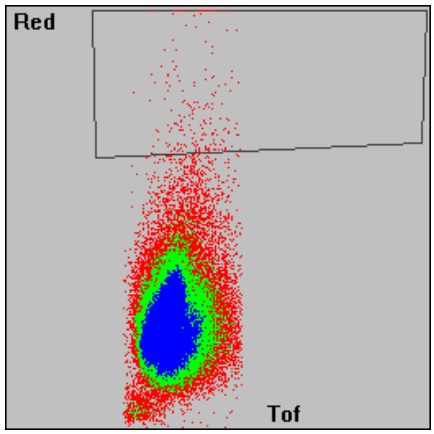


Figure 5.5 Scatter plot of COPAS sort of 4.4.4 library.

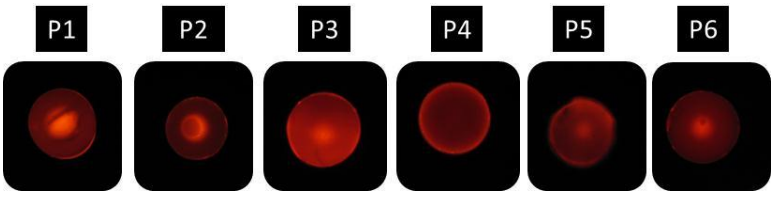


Figure 5.6 Beads designated "hits" by fluorescence

5.2.5 *De Novo* Sequencing

When performing *de novo* analysis of protein digests, online analysis protocols such as Protein Prospector[21] or proprietary software packages are available to assist in sequencing. Unfortunately, there are no available open source *de novo* sequencing protocols for branched peptides. Therefore, to successfully sequence the first generation 3.3.3 branched peptide library, an effective but tedious manual search of all 4,096 potential peptides and their corresponding fragments was performed using spreadsheet searching. This approach was deemed impractical for the much larger second generation 4.4.4 library. Therefore, in-house software was developed specifically to analyze branched peptides of various lengths and amino acid composition. Using input from Figure 5.3, this software identifies all peptide sequences corresponding to the DHB parent ion identified by MALDI as well as their theoretical MS-MS fragments.[22]

Using this method, sequences of the 6 "hit" peptides were deconvoluted in as little as 10 minutes of data analysis per peptide, shown in Table 5.1 along with the corresponding logo produced using WebLogo Version 2.8.2 (Figure 5.7).[23] Interestingly, asparagine was clearly preferred for all 6 peptide sequences at position 6. Also, there was a 1:1 distribution of either a leucine or glutamine residue at position 4. The K_{BBA} residue was evenly distributed between positions 1-3 and 7-9 with no clear preference for any particular position. Another interesting feature of these peptides is that very few aromatic residues were observed with no clear consensus for any particular position within the peptide sequence.

Table 5.1 Deconvoluted peptide sequences. B = K_{BBA}

Peptide	Sequence	K_d (μM)
P1	(**B**AHL)$_2$ NDPE	NA
P2	(EA**B**Q)$_2$ NW**B**E	NA
P3	(EFHL)$_2$ N**B**H**B**	NA
P4	(TNSQ)$_2$ N**B**QSY	> 100
P5	(Y**B**SL)$_2$ NWP**B**	NA
P6	(Y**BB**Q)$_2$ NDQE	NA

Figure 5.7 Frequency plot of deconvoluted sequences.

5.2.6 Determination of binding affinity

With the sequences of our hit peptides in hand, we proceeded to resynthesize these compounds using traditional solid phase peptide synthesis (SPPS) techniques. A fluorescein isothiocyanate (FITC) fluorophore was covalently attached to each 4.4.4 peptide to provide a tag for quantitation as well as a fluorescent reporter group for future cell uptake studies. Unfortunately, we observed complete insolubility in aqueous as well as DMSO/aqueous media combinations for all FITC-labeled peptides. We next investigated solubility of the branched boronopeptides that lacked the FITC label. Although we initially observed that all peptides displayed moderate to good solubility in

aqueous solvents, peptides began to crash out of solution within several hours in the form of white films. These peptides were found to be insoluble in up to 30% DMSO solutions as well as methanol solutions. This result was surprising as we had previously observed no such solubility issues with the boron containing peptides prepared for preliminary MALDI analysis. Presently, it seems possible that the general insolubility of these peptides is due to either the lack of basic amino acid residues or extended inter- or intra-molecular hydrogen bonded macrostructures between boronic acid residues.

Interestingly, the only peptide that was soluble in water was peptide **P4**, which contained only one boronic acid moiety and multiple threonine and serine residues. Although peptide **P4** was amenable to K_d determination experiments, the lack of a chromophore prevented quantitation by UV-vis spectrometry. Accordingly, we appended the sequence with an additional tyrosine on the carboxyamide terminus of the peptide to allow for quantitation. After performing K_d analysis using both dot blot and electro-mobility shift assays (EMSA), it was determined that the tyrosine labeled peptide **P4** had a binding constant of >100 μM (Figure 5.8). Although a disappointing result, several key lessons can be drawn from our experiences with these borono branched peptides. One valuable lesson learned from this study is that the K_{BBA} boron moiety is not sufficient to promote efficient solvation of non-basic peptides. Even with peptides **P2**, **P5**, and **P6** which displayed 3 and 4 boronic acid moieties respectively, the resulting peptides displayed poor solubility in aqueous solvent. Indeed, it now seems quite probable that high density functionalization with boronic acids may promote the formation of the observed insoluble films. Unfortunately, since we were only able to acquire one data point correlating peptide sequence to affinity for RRE, it is impossible at this time to

evaluate whether or not boronic acid peptide/RNA binding motifs is worthy of further pursuit.

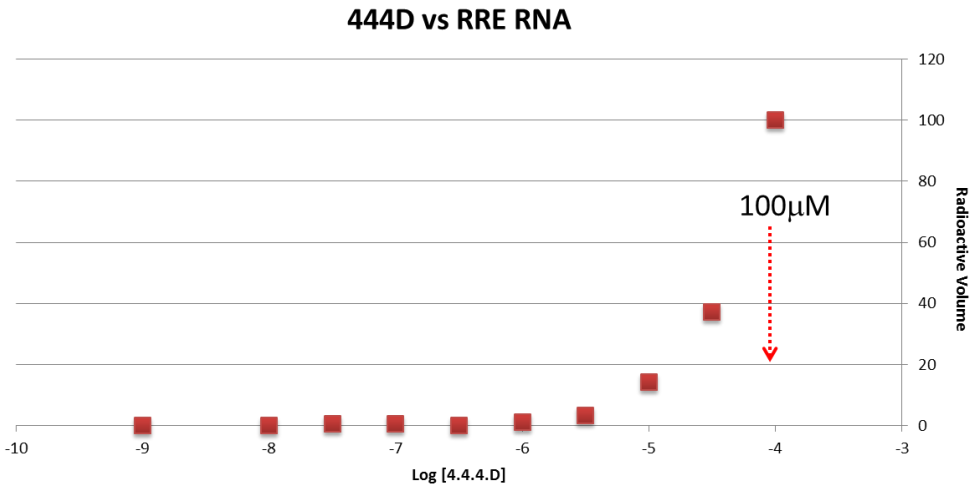


Figure 5.8 K_d plot of peptide P4

5.3 Conclusions

We have developed a simple, rapid, and highly effective protocol for the analysis of branched peptides as well as branched boronic acid peptides. Also, our previously developed protocols for MALDI analysis of boronic acid peptides were compatible with the analysis of mixed boronic acid peptides. We explored possible Lewis acid/base binding motifs between RNA and boron by assaying a peptide combinatorial library of 65,536 unique members. COPAS analysis was utilized to isolate the top 1.5% most fluorescent beads, which were then manually assayed by fluorescence microscopy to yield 6 designated hit peptides. To deconvolute the sequences of these peptides, we developed an in-house software package for analyzing MS-MS fragments of branched peptides. This analysis protocol is currently available for complementary use. The deconvoluted peptide sequences were resynthesized and analyzed for their affinity

towards RRE. Unfortunately, only one of the 6 hits was found to be soluble in aqueous solvents and hence was the only peptide assayed for binding affinity to RRE. All other peptides formed insoluble films that were not soluble in up to 30% DMSO. Presently, it is not known whether these films are the result of the non-basic nature of the peptides or extended inter- or intra-molecular hydrogen bonded macrostructures between boronic acid residues. Further studies to explore these phenomena are underway.

5.4 Experimental

5.4.1 RRE incubation conditions

All beads were incubated with 0.3 mg/mL tRNA, 10nM RRE, and 1 mg/mL BSA simultaneously for 5 hours in HEPES-1 buffer (10mM HEPES, 1mM KCl, 1mM $MgCl_2$, and 20mM NaCl).

5.4.2 Photocleavage and MALDI Analysis

The 4.4.4 peptides were removed from the resin by photocleavage of the ANP photolinker. Beads selected for photocleavage were taken up individually into clear non-stick 0.5 mL microfuge tubes with 20 μL of 1:1 MeOH: H_2O. The microfuge tubes containing resin were placed into a foil-lined container and were irradiated for 1 hour with light at 365 nm in using a 4 W handheld UV lamp. The resin was immediately removed from the supernatant after photocleavage. MS and MS-MS data was obtained from MALDI-TOF analysis. The molecule 2,5-dihydroxybenzoic acid (DHB) was utilized as MALDI matrix for all samples to yield the DHB boronic esters of analyzed peptides.

5.4.3 Preparation of ^{32}P-Labeled RNA

HIV-1 RRE double stranded DNA was obtained *via* PCR using HotStar Taq Polymerase (Qiagen) with synthetic template: 5'-GGCTGGTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGCC and two synthetic primers: 5'-GGCTGGCCTGTAC and 5'-ATGTAATACGACTCACTATAGG (Integrated DNA Technologies). HIV-1 RRE RNA was prepared by *in vitro* transcription from PCR DNA product using the Ribomax T7 Express System (Promega) according to the manufacturers protocol. The DNA template was degraded by addition of DNase after transcription was complete, and the transcript was purified by 10% PAGE containing 7.5 M urea. RRE RNA was eluted from the gel overnight at 4 °C in 300 mM sodium acetate, 10 mM Tris•HCl, pH 7.4, and 10 mM EDTA. Eluent was desalted using a Sep-Pak C-18 (Waters) and lyophilized before treating with calf intestinal phosphatase (New England BioLabs). Dephosphorylated RNA was recovered by phenol extraction followed by ethanol precipitation.

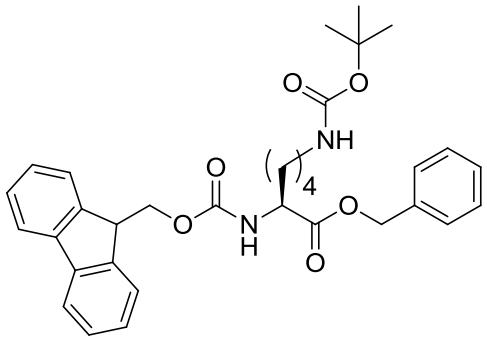
HIV-1 RRE RNA was 5'-end labeled by treating 100 pmol of dephosphorylated RNA with 20 nmol of [γ-^{32}P] ATP (111 TBq mol^{-1}) (PerkinElmer) and 20 units of T4 polynucleotide kinase in 70 mM Tris•HCl, pH 7.6, 10 mM $MgCl_2$, and 5 mM dithiothreitol (New England BioLabs). The RNA was labeled at 37 °C for 50 minutes followed by a 10 minute incubation at 65 °C and was finally recovered by ethanol precipitation. Sample integrity was determined by 12% PAGE containing 7.5 M urea. Labeled RNA stocks were stored at -20 °C as 500 nM stocks in nuclease free water (Qiagen).

5.4.4 Dot Blot Assays

Assays were performed by refolding 800 pM ^{32}P-labeled HIV-1 RRE RNA in 2X phosphate buffer (20 mM potassium phosphate, pH 7, 200 mM KCl, 2 mM $MgCl_2$, and 40 mM NaCl) at 95 °C for 3 min and cooling slowly to room temperature over 20 min. Peptides were diluted to 2X concentration from 200 μM stocks in nuclease-free water. Twenty-five microliter aliquots of refolded RNA were added to 25 μL of peptide to achieve the desired concentration of each component. The mixtures incubated at room temperature for four hours prior to filtering each sample through pre-equilibrated Whatman 0.45 μm pore size Protran nitrocellulose membranes using a Whatman Minifold I 96 well Dot Blot system. Each filtration was immediately followed by two consecutive washes with 50 μL of 1X phosphate buffer. Peptide binding was visualized by autoradiography using a storage phosphor screen and Typhoon Trio (GE Healthcare). Densitometry measurements were made in ImageQuant TL software (Amersham Biosciences).

5.4.5 Synthesis of K_{BBA}

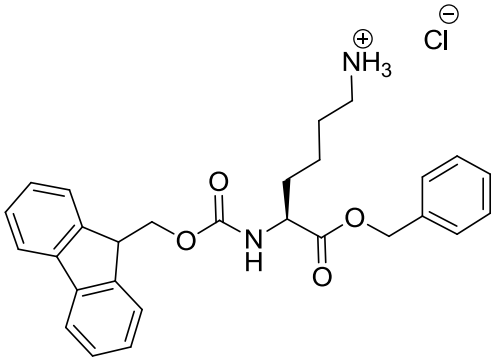
(*S*)-Benzyl 2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-((tert-butoxycarbonyl)amino)hexanoate **(5.5)**



To a round bottom flask was added (*S*)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-((tert-butoxycarbonyl)amino)hexanoic acid **5.4** (9.00 g, 19.2 mmol) followed by 20 mL of dry DMF. To the mixture was added cesium carbonate (7.51 g, 23.1 mmol), which yielded a cloudy solution. Benzyl bromide (2.97 mL, 25.0 mmol) was

added dropwise. After 4 hours, 100 mL of LiBr was added then extracted with 4 x 50 mL of diethyl ether. The combined organic layers were washed with 1 x 50 mL LiBr and 1 x 50 mL brine. The organic layer was dried with sodium sulfate and concentrated under reduced pressure. The crude product was purified on silica gel and concentrated to yield the title compound (10 g, 94%) as a viscous semi-solid. TLC: 3:1 / Hexanes:EtOAc, Rf 0.3. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.68 (d, *J* = 7.5 Hz, 2H), 7.52 (d, *J* = 7.4 Hz, 2H), 7.32 (t, *J* = 7.5 Hz, 2H), 7.29 – 7.19 (m, 7H), 5.35 (d, *J* = 6.3 Hz, 1H), 5.13 (d, *J* = 12.2 Hz, 1H), 5.07 (d, *J* = 12.2 Hz, 1H), 4.44 (broad s, 1H), 4.38 – 4.26 (m, 3H), 4.13 (t, *J* = 7.0 Hz, 1H), 2.99 (broad s, 2H), 1.88 – 1.72 (m, 1H), 1.66 – 1.60 (m, 1H), 1.36 (s, 13H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 172.5, 156.2, 156.2, 143.9, 141.4, 135.4, 128.7, 128.6, 128.5, 127.8, 127.2, 125.2, 120.1, 79.2, 67.3, 67.1, 53.9, 47.3, 40.1, 32.1, 29.7, 28.6, 22.5. HRMS (ESI+) *m/z* calcd for $C_{33}H_{38}N_2O_6$ $[M+H]^+$ 559.2803, found 559.2839

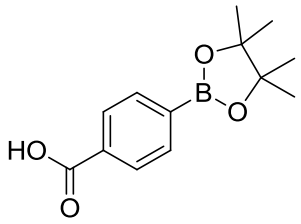
(*S*)-5-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-6-(benzyloxy)-6-oxohexan-1-aminium chloride **(5.6)**



Compound **5.5** (10.0 g, 19.9 mmol) was suspended in 20 mL of 4 N HCl in dioxane. The cloudy mixture quickly became completely translucent. After 1 hour, TLC indicated complete consumption of the starting materials using the solvent system for **5.5**. The dioxane was removed under reduced pressure, and the resulting oil was washed with ether (2 x 30 mL). After pouring off the ethereal supernatant and removal of residual ether under

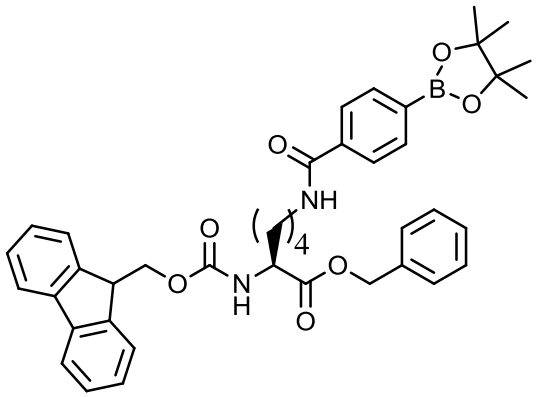
reduced pressure, the product was obtained as a fine white powder (8.7 g, 98%), mp 79.2 - 81.6 ºC. ^{1}H NMR (500 MHz, MeOD) δ 7.81 (d, *J* = 7.6 Hz, 2H), 7.66 (t, *J* = 6.6 Hz, 2H), 7.40 (t, *J* = 7.5 Hz, 2H), 7.37 – 7.28 (m, 7H), 5.21 (d, *J* = 12.3 Hz, 1H), 5.15 (d, *J* = 12.3 Hz, 1H), 4.42 (dd, *J* = 10.6, 7.0 Hz, 1H), 4.31 (dd, *J* = 10.6, 7.0 Hz, 1H), 4.24 (dd, *J* = 9.4, 5.0 Hz, 1H), 4.19 (t, *J* = 7.0 Hz, 1H), 2.92 – 2.84 (m, 2H), 1.94 – 1.84 (m, 1H), 1.80 – 1.60 (m, 3H), 1.53 – 1.38 (m, 2H).^{13}C NMR (126 MHz, MeOD) δ 172.4, 157.4, 143.8, 141.3, 135.9, 128.3, 128.1, 128.0, 127.5, 126.9, 124.9, 119.7, 66.8, 66.6, 54.0, 47.1, 39.2, 30.6, 26.7, 22.6. HRMS (ESI+) *m/z* calcd for $C_{28}H_{30}N_2O_4$ $[M+H]^+$ 459.2206, found 459.2295.

4-(4,4,5,5-Tetramethyl-1,3,2-dioxaborolan-2-yl)benzoic acid **(5.8)**



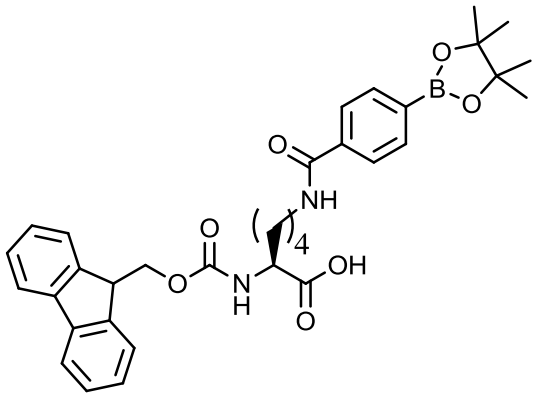
4-boronobenzoic acid (**5.7**) (10.0 g, 60.3 mmol) was dissolved in 125 mL of ether (not completely soluble), and pinacol diol (9.26 g, 78.0 mmol) was added to the mixture. The initially cloudy reaction mixture became translucent after ~ 2 hours. The reaction was allowed to proceed for a total of 4 hours. The organic layer was washed with 3 x 75 mL of water and 1 x 75 mL of brine. After removal of the ether under reduced pressure, the title compound was obtained as a white solid (13 g, 90%). No further purification was performed, mp 220 - 223 ºC. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.11 (d, *J* = 8.2 Hz, 2H), 7.91 (d, *J* = 8.2 Hz, 2H), 1.36 (s, 12H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 172.5, 134.9, 131.6, 129.3, 84.4, 25.0. ^{11}B NMR (160 MHz, $CDCl_3$) δ 30.6. HRMS (ESI+) *m/z* calcd for $C_{13}H_{17}BO_4$ $[M+H]^+$ 249.1236, found 249.1309**.**

(*S*)-Benzyl 2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzamido)hexanoate **(5.9)**



To a round bottom flask was added 10.3 g of **5.6** (20.9 mmol), which was then dissolved in 60 mL of DMF. Compound **5.8** was dissolved in 40 mL of DMF, and 7.41 g of HCTU (17.9 mmol) was added. To this mixture, 7.27 ml (41.8 mmol) of DIEA was added, mixed for ~ 1 minute, and then quickly added to the lysine solution. After vigorous stirring for 4 hours, the reaction mixture was poured into 200 mL of LiBr and extracted with 3 x 100 mL of diethyl ether. The organic layer was washed with 1 x 100 mL of brine and dried with Na_2SO_4. This solution was concentrated under reduced pressure and purified on silica to yield the title compound (8.4 g, 81%) as a white foamy solid, mp 85.3 - 88.2 ^{o}C. TLC: 1:1 / Hexanes:EtOAc, Rf 0.5. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.83 (d, *J* = 7.9 Hz, 2H), 7.74 (d, *J* = 7.3 Hz, 4H), 7.57 (d, *J* = 7.4 Hz, 2H), 7.38 (t, *J* = 7.4 Hz, 2H), 7.36 – 7.25 (m, 7H), 6.42 (broad s, 1H), 5.58 (d, *J* = 7.7 Hz, 1H), 5.20 (d, *J* = 12.2 Hz, 1H), 5.14 (d, *J* = 12.2 Hz, 1H), 4.52 – 4.25 (m, 3H), 4.18 (t, *J* = 7.0 Hz, 1H), 3.49 – 3.27 (m, 2H), 1.96 – 1.83 (m, 1H), 1.80 – 1.69 (m, 1H), 1.68 – 1.49 (m, 2H), 1.34 (s, 14H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 172.4, 167.9, 156.3, 143.8, 141.4, 136.8, 135.4, 135.0, 128.7, 128.6, 128.4, 127.8, 127.2, 126.2, 125.2, 120.1, 84.2, 67.3, 67.1, 53.8, 47.2, 39.7, 32.3, 29.0, 25.0, 22.6. ^{11}B NMR (160 MHz, $CDCl_3$) δ 30.2. HRMS (ESI+) *m/z* calcd for $C_{41}H_{45}BN_2O_7$ $[M+H]^+$ 689.3362, found 689.3435

(*S*)-2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-6-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzamido)hexanoic acid **(5.10)**



10% Pd on carbon (0.901 g, 0.846 mmol) was added to a solution **5.9** *(*5.83 g, 8.46 mmol) in EtOH (35 mL). A hydrogen balloon apparatus was attached to the reaction flask, and the reaction was allowed to proceed for 2 hours, after which TLC, using the solvent system for **5.9**, indicated complete consumption of **5.9**. The product was purified on silica gel to yield the title compound as a foamy solid (3.0 g, 59%), mp 66.1 - 70.0 °C. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.81 (d, *J* = 8.0 Hz, 2H), 7.75 (d, *J* = 7.9 Hz, 2H), 7.69 (d, *J* = 7.6 Hz, 2H), 7.53 (t, *J* = 7.0 Hz, 2H), 7.33 (t, *J* = 7.4 Hz, 2H), 7.23 (t, *J* = 7.4 Hz, 2H), 6.80 (broad s, 1H), 5.88 (d, *J* = 8.0 Hz, 1H), 4.46 – 4.34 (m, 1H), 4.34 – 4.26 (m, 2H), 4.14 (t, *J* = 7.1 Hz, 1H), 3.39 (broad s, 2H), 1.96 – 1.82 (m, 1H), 1.82 – 1.67 (m, 1H), 1.67 – 1.36 (m, 4H), 1.32 (s, 12H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 175.1, 168.4, 156.5, 143.8, 141.3, 136.5, 135.0, 127.8, 127.2, 126.3, 125.3, 120.0, 84.3, 67.2, 53.7, 47.2, 39.9, 32.0, 29.0, 25.0, 22.5.^{11}B NMR (160 MHz, $CDCl_3$) δ 30.1. HRMS (ESI+) *m/z* calcd for $C_{34}H_{39}BN_2O_7$ $[M+H]^+$ 599.2914, found 599.2969.

5.5 Coding for de novo sequencing program

5.5.1 Pepseq Code

```
#!/usr/bin/python
# -*- coding: utf-8 -*-
import os
import sys
# Add project root directory (enable symlink, and trunk execution).
PROJECT_ROOT_DIRECTORY = os.path.abspath(
    os.path.dirname(os.path.dirname(os.path.realpath(sys.argv[0]))))
if (os.path.exists(os.path.join(PROJECT_ROOT_DIRECTORY, 'pepseq'))
    and PROJECT_ROOT_DIRECTORY not in sys.path):
    sys.path.insert(0, PROJECT_ROOT_DIRECTORY)
    os.putenv('PYTHONPATH', PROJECT_ROOT_DIRECTORY)
import logging
import optparse
import pepseq
import itertools
import ConfigParser
import pepseq.amino_acid
import gettext
from gettext import gettext as _
gettext.textdomain('pepseq')
LEVELS = (logging.ERROR,
```

```
        logging.WARNING,
        logging.INFO,
        logging.DEBUG)
def parse_args():
    usage = _("pepseq [options] <inputfile>")
    parser = optparse.OptionParser(version="pepseq %s" % pepseq.__version__,
                                   usage=usage)
    parser.add_option('-d', '--debug', dest='debug_mode', action='store_true',
      help=_('Print the maximum debugging info (implies -vv)'))
    parser.add_option('-v', '--verbose', dest='logging_level', action='count',
      help=_('set error_level output to warning, info, and then debug'))
    parser.set_defaults(logging_level=0)
    (options, args) = parser.parse_args()
    if options.debug_mode:
      loglevel = 3
    else:
      try:
        loglevel = LEVELS[options.logging_level]
      except IndexError:
        loglevel = 3
    logging.basicConfig(level=loglevel,
        format='%(asctime)s %(levelname)s %(message)s')
    return parser.parse_args()
```

```
def main():
    (options, args) = parse_args()
    config = ConfigParser.ConfigParser()
    try:
        config.readfp(open(args[0]))
    except IndexError:
        try:
            config.readfp(open("sample-input"))
        except IOError:
            config.readfp(open("../sample-input"))
    bucket1 = []
    for char in config.get('DEFAULT', 'bucket1').split(' '):
        bucket1.append(pepseq.amino_acid.AminoAcid(char))
    bucket2 = []
    for char in config.get('DEFAULT', 'bucket2').split(' '):
        bucket2.append(pepseq.amino_acid.AminoAcid(char))

    bucket3 = []
    for char in config.get('DEFAULT', 'bucket3').split(' '):
        bucket3.append(pepseq.amino_acid.AminoAcid(char))
    bucket4 = []
    for char in config.get('DEFAULT', 'bucket4').split(' '):
        bucket4.append(pepseq.amino_acid.AminoAcid(char))
```

```
bucket5 = []
for char in config.get('DEFAULT', 'bucket5').split(' '):
    bucket5.append(pepseq.amino_acid.AminoAcid(char))
    bucket5.append(pepseq.amino_acid.AminoAcid('N'))
bucket6 = []
for char in config.get('DEFAULT', 'bucket6').split(' '):
    bucket6.append(pepseq.amino_acid.AminoAcid(char))
bucket7 = []
for char in config.get('DEFAULT', 'bucket7').split(' '):
    bucket7.append(pepseq.amino_acid.AminoAcid(char))
bucket8 = []
for char in config.get('DEFAULT', 'bucket8').split(' '):
    bucket8.append(pepseq.amino_acid.AminoAcid(char))
bucket9 = []
for char in config.get('DEFAULT', 'bucket9').split(' '):
    bucket9.append(pepseq.amino_acid.AminoAcid(char))
possibilities = (bucket1, bucket2, bucket3,
                 bucket4, bucket5, bucket6,
                 bucket7, bucket8, bucket9)
peptides = []
sequences = tuple(itertools.product(*possibilities))
for sequence in sequences:
    if sequence[4].letter == 'K':
```

```
            peptide = pepseq.amino_acid.Peptide(sequence)
            peptide.branch_before(4)
            peptides.append(peptide)
    peptides.sort()
    output = []
    for peptide in peptides:
        if peptide.weight > float(config.get('DEFAULT', 'low')):
            if peptide.weight < float(config.get('DEFAULT', 'high')):
                output.append(peptide)
    print 'Number of matches: %d' % len(output)
    for peptide in output:
        print ("%.2f  %s   " % (peptide.weight, peptide) + '  '.join(peptide.subweights))
if __name__ == "__main__":
    main()
```

5.5.2 Amino Acid Definitions and Sorting Protocols

```
# -*- coding: utf-8 -*-

class AminoAcid(object):

    """ Amino acids are molecules containing an amine group, a carboxylic acid

        group and a side chain that varies between different amino acids. """

    aminoAcids = {'G': ('Glycine',              'Gly', 57.00),

                  'A': ('Alanine',              'Ala', 71.00),

                  'S': ('Serine',               'Ser', 87.00),

                  'T': ('Threonine',            'Thr', 101.00),

                  'C': ('Cysteine',             'Cys', 103.00),

                  'V': ('Valine',               'Val', 99.10),

                  'L': ('Leucine',              'Leu', 113.10),

                  'I': ('Isoleucine',           'Ile', 113.10),

                  'M': ('Methionine',           'Met', 131.00),

                  'P': ('Proline',              'Pro', 97.10),

                  'F': ('Phenylalanine',        'Phe', 147.10),

                  'Y': ('Tyrosine',             'Tyr', 163.10),

                  'W': ('Tryptophan',           'Trp', 186.10),

                  'D': ('Aspartic Acid',        'Asp', 115.00),

                  'E': ('Glutamic Acid',        'Glu', 129.00),

                  'N': ('Asparagine',           'Asn', 114.00),

                  'Q': ('Glutamine',            'Gln', 128.10),

                  'H': ('Histidine',            'His', 137.10),
```

```
            'K': ('Lysine',                 'Lys', 128.10),
            'R': ('Arginine',               'Arg', 156.10),
            'B': ('Borono Phenylalanine',   'Bpa', 394.10),
    }
    def __init__(self, letter):
        self.letter = letter.upper()
        self.name = AminoAcid.aminoAcids[self.letter][0]
        self.triple = AminoAcid.aminoAcids[self.letter][1]
        self.weight = AminoAcid.aminoAcids[self.letter][2]
    def __str__(self):
        return "%s (%s, %s) MW: %.2f" % (
              self.name, self.triple, self.letter, self.weight)
class Peptide(object):
    def __init__(self, amino_acids):
        self.amino_acids = amino_acids
        self.weight = 0
        self.subweights = []
        self.string = ""
        for amino_acid in amino_acids:
            self.weight += amino_acid.weight
            self.string += ('%s' % (amino_acid.letter))
    def branch_before(self, index):
        self.weight = 18.00
```

```
    for (counter, amino_acid) in enumerate(self.amino_acids):
        if counter < index:
            self.weight += 2 * amino_acid.weight
        else:
            self.weight += amino_acid.weight
    self.string = ('(' + self.string[0:index] + ')2 ' +
                   self.string[index + 1:])
    self._calculate_subweights()
def _calculate_subweights(self):
    self.subweights.append(str(self.weight - self.amino_acids[0].weight))
    self.subweights.append(str(float(self.subweights[0]) - self.amino_acids[1].weight))
    self.subweights.append(str(float(self.subweights[1]) - self.amino_acids[2].weight))
    self.subweights.append(str(float(self.subweights[2]) - self.amino_acids[3].weight))
    self.subweights.append(str(self.weight - 17.1 - self.amino_acids[-1].weight))
    self.subweights.append(str(float(self.subweights[-1]) - self.amino_acids[-2].weight))
    self.subweights.append(str(float(self.subweights[-1]) - self.amino_acids[-3].weight))
    self.subweights.append(str(float(self.subweights[-1]) - self.amino_acids[-4].weight))
def __cmp__(self, other):
    return cmp(self.weight, other.weight)
def __str__(self):
    return self.string
```

5.5.3 Test_Amino_Acid Code

```
# -*- coding: utf-8 -*-
from amino_acid import AminoAcid
import unittest
class AminoAcidTestCase(unittest.TestCase):
    def test_glycine(self):
        amino_acid = AminoAcid('g')
        self.assertEqual(amino_acid.name, "Glycine")
        self.assertEqual(amino_acid.triple, "Gly")
        self.assertEqual(amino_acid.letter, "G")
        self.assertEqual(amino_acid.weight, 57.05)
    def test_alanine(self):
        amino_acid = AminoAcid('a')
        self.assertEqual(amino_acid.name, "Alanine")
        self.assertEqual(amino_acid.triple, "Ala")
        self.assertEqual(amino_acid.letter, "A")
        self.assertEqual(amino_acid.weight, 71.09)
    def test_serine(self):
        amino_acid = AminoAcid('s')
        self.assertEqual(amino_acid.name, "Serine")
        self.assertEqual(amino_acid.triple, "Ser")
        self.assertEqual(amino_acid.letter, "S")
```

```
    self.assertEqual(amino_acid.weight, 87.08)
def test_threonine(self):
    amino_acid = AminoAcid('t')
    self.assertEqual(amino_acid.name, "Threonine")
    self.assertEqual(amino_acid.triple, "Thr")
    self.assertEqual(amino_acid.letter, "T")
    self.assertEqual(amino_acid.weight, 101.11)


def test_cysteine(self):
    amino_acid = AminoAcid('c')
    self.assertEqual(amino_acid.name, "Cysteine")
    self.assertEqual(amino_acid.triple, "Cys")
    self.assertEqual(amino_acid.letter, "C")
    self.assertEqual(amino_acid.weight, 103.15)
def test_valine(self):
    amino_acid = AminoAcid('v')
    self.assertEqual(amino_acid.name, "Valine")
    self.assertEqual(amino_acid.triple, "Val")
    self.assertEqual(amino_acid.letter, "V")
    self.assertEqual(amino_acid.weight, 99.14)
def test_leucine(self):
    amino_acid = AminoAcid('l')
    self.assertEqual(amino_acid.name, "Leucine")
```

```
        self.assertEqual(amino_acid.triple, "Leu")
        self.assertEqual(amino_acid.letter, "L")
        self.assertEqual(amino_acid.weight, 113.16)
    def test_isoleucine(self):
        amino_acid = AminoAcid('i')
        self.assertEqual(amino_acid.name, "Isoleucine")
        self.assertEqual(amino_acid.triple, "Ile")
        self.assertEqual(amino_acid.letter, "I")
        self.assertEqual(amino_acid.weight, 113.16)
    def test_methionine(self):
        amino_acid = AminoAcid('m')
        self.assertEqual(amino_acid.name, "Methionine")
        self.assertEqual(amino_acid.triple, "Met")
        self.assertEqual(amino_acid.letter, "M")
        self.assertEqual(amino_acid.weight, 131.19)

    def test_proline(self):
        amino_acid = AminoAcid('p')
        self.assertEqual(amino_acid.name, "Proline")
        self.assertEqual(amino_acid.triple, "Pro")
        self.assertEqual(amino_acid.letter, "P")
        self.assertEqual(amino_acid.weight, 97.12)
    def test_phenylalanine(self):
```

```
        amino_acid = AminoAcid('f')
        self.assertEqual(amino_acid.name, "Phenylalanine")
        self.assertEqual(amino_acid.triple, "Phe")
        self.assertEqual(amino_acid.letter, "F")
        self.assertEqual(amino_acid.weight, 147.18)
    def test_tyrosine(self):
        amino_acid = AminoAcid('y')
        self.assertEqual(amino_acid.name, "Tyrosine")
        self.assertEqual(amino_acid.triple, "Tyr")
        self.assertEqual(amino_acid.letter, "Y")
        self.assertEqual(amino_acid.weight, 163.18)
    def test_tryptophan(self):
        amino_acid = AminoAcid('w')
        self.assertEqual(amino_acid.name, "Tryptophan")
        self.assertEqual(amino_acid.triple, "Trp")
        self.assertEqual(amino_acid.letter, "W")
        self.assertEqual(amino_acid.weight, 186.21)
    def test_aspartic_acid(self):
        amino_acid = AminoAcid('d')
        self.assertEqual(amino_acid.name, "Aspartic Acid")
        self.assertEqual(amino_acid.triple, "Asp")
        self.assertEqual(amino_acid.letter, "D")
        self.assertEqual(amino_acid.weight, 115.09)
```

```
def test_glutamic_acid(self):
    amino_acid = AminoAcid('e')
    self.assertEqual(amino_acid.name, "Glutamic Acid")
    self.assertEqual(amino_acid.triple, "Glu")
    self.assertEqual(amino_acid.letter, "E")
    self.assertEqual(amino_acid.weight, 129.12)
def test_asparagine(self):
    amino_acid = AminoAcid('n')
    self.assertEqual(amino_acid.name, "Asparagine")
    self.assertEqual(amino_acid.triple, "Asn")
    self.assertEqual(amino_acid.letter, "N")
    self.assertEqual(amino_acid.weight, 114.11)
def test_glutamine(self):
    amino_acid = AminoAcid('q')
    self.assertEqual(amino_acid.name, "Glutamine")
    self.assertEqual(amino_acid.triple, "Gln")
    self.assertEqual(amino_acid.letter, "Q")
    self.assertEqual(amino_acid.weight, 128.14)
def test_histidine(self):
    amino_acid = AminoAcid('h')
    self.assertEqual(amino_acid.name, "Histidine")
    self.assertEqual(amino_acid.triple, "His")
```

```
        self.assertEqual(amino_acid.letter, "H")
        self.assertEqual(amino_acid.weight, 137.14)
    def test_lysine(self):
        amino_acid = AminoAcid('k')
        self.assertEqual(amino_acid.name, "Lysine")
        self.assertEqual(amino_acid.triple, "Lys")
        self.assertEqual(amino_acid.letter, "K")
        self.assertEqual(amino_acid.weight, 128.17)

    def test_arginine(self):
        amino_acid = AminoAcid('r')
        self.assertEqual(amino_acid.name, "Arginine")
        self.assertEqual(amino_acid.triple, "Arg")
        self.assertEqual(amino_acid.letter, "R")
        self.assertEqual(amino_acid.weight, 156.19)
    def test_borono_phenylalanine(self):
        amino_acid = AminoAcid('b')
        self.assertEqual(amino_acid.name, "Borono Phenylalanine")
        self.assertEqual(amino_acid.triple, "Bpa")
        self.assertEqual(amino_acid.letter, "B")
        self.assertEqual(amino_acid.weight, 309.08)
def suite():
    suite = unittest.TestLoader().loadTestsFromTestCase(AminoAcidTestCase)
```

```
    return suite

if __name__ == '__main__':

    unittest.TextTestRunner().run(suite())
```

5.6 References

1. Suhasini, M.; Reddy, T. R. Cellular proteins and HIV-1 Rev function *Curr. HIV Res.* **2009**, *7*, 91-100.

2. Hamy, F.;Felder, E.;Lipson, K.; Klimkait, T. Merged screening for human immunodeficiency virus Tat and Rev inhibitors *J. Biomol. Screen.* **2001**, *6*, 179-187.

3. De Clercq, E. New approaches toward anti-HIV chemotherapy *J. Med. Chem.* **2005**, *48*, 1297-1313.

4. Daelemans, D.;Afonina, E.;Nilsson, J.;Werner, G.;Kjems, J.;De Clercq, E.;Pavlakis, G. N.; Vandamme, A. M. A synthetic HIV-1 Rev inhibitor interfering with the CRM1-mediated nuclear export *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14440-14445.

5. Chapman, R. L.;Stanley, T. B.;Hazen, R.; Garvey, E. P. Small molecule modulators of HIV Rev/Rev response element interaction identified by random screening *Antiviral Res.* **2002**, *54*, 149-162.

6. Baba, M. Inhibitors of HIV-1 gene expression and transcription *Curr. Top. Med. Chem.* **2004**, *4*, 871-882.

7. DeJong, E. S.;Chang, C. E.;Gilson, M. K.; Marino, J. P. Proflavine acts as a Rev inhibitor by targeting the high-affinity Rev binding site of the Rev responsive element of HIV-1 *Biochemistry.* **2003**, *42*, 8035-8046.

8. Plavec, I.;Agarwal, M.;Ho, K. E.;Pineda, M.;Auten, J.;Baker, J.;Matsuzaki, H.;Escaich, S.;Bonyhadi, M.; Bohnlein, E. High transdominant RevM10 protein levels are required to inhibit HIV-1 replication in cell lines and primary T cells: implication for gene therapy of AIDS *Gene Ther.* **1997**, *4*, 128-139.

9. Chaloin, L.;Smagulova, F.;Hariton-Gazal, E.;Briant, L.;Loyter, A.; Devaux, C. Potent inhibition of HIV-1 replication by backbone cyclic peptides bearing the Rev arginine rich motif *J. Biomed. Sci.* **2007**, *14*, 565-584.

10. Li, K.;Davis, T. M.;Bailly, C.;Kumar, A.;Boykin, D. W.; Wilson, W. D. A heterocyclic inhibitor of the REV-RRE complex binds to RRE as a dimer *Biochemistry.* **2001**, *40*, 1150-1158.

11. Bryson, D. I.;Zhang, W.;Ray, W. K.; Santos, W. L. Screening of a branched peptide library with HIV-1 TAR RNA *Mol. Biosyst.* **2009**, *5*, 1070-1073.

12. Malan, C.; Morin, C. A Concise Preparation of 4-Borono-$_L$-phenylalanine ($_L$-BPA) from $_L$-Phenylalanine *J. Org. Chem.* **1998**, *63*, 8019–8020.

13. Duggan, P.; Offerman, D. The Preparation of Solid-Supported Peptide Boronic Acids Derived from 4-Borono-$_L$-phenylalanine and their Affinity for Alizarin *Aust. J. Chem.* **2007**, *60*, 829-834.

14. Baker, S. J.;Ding, C. Z.;Akama, T.;Zhang, Y. K.;Hernandez, V.; Xia, Y. Therapeutic potential of boron-containing compounds *Future Med. Chem.* **2009**, *1*, 1275-1288.

15. Rock, F. L.;Mao, W.;Yaremchuk, A.;Tukalo, M.;Crepin, T.;Zhou, H.;Zhang, Y. K.;Hernandez, V.;Akama, T.;Baker, S. J.;Plattner, J. J.;Shapiro, L.;Martinis, S. A.;Benkovic, S. J.;Cusack, S.; Alley, M. R. An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site *Science.* **2007**, *316*, 1759-1761.

16. Hui, X.;Baker, S. J.;Wester, R. C.;Barbadillo, S.;Cashmore, A. K.;Sanders, V.;Hold, K. M.;Akama, T.;Zhang, Y. K.;Plattner, J. J.; Maibach, H. I. In Vitro penetration of a novel oxaborole antifungal (AN2690) into the human nail plate *J. Pharm. Sci.* **2007**, *96*, 2622-2631.

17. Crumpton, J. B.;Zhang, W.; Santos, W. L. Facile analysis and sequencing of linear and branched boronic acids by MALDI mass spectrometry *Anal. Chem.* **2011**, *83*, 3548-3554.

18. Bryson, D. I.;Zhang, W.;McLendon, P. M.;Reineke, T. M.; Santos, W. L. Toward Targeting RNA Structure: Branched Peptides as Cell-Permeable Ligands to TAR RNA *ACS Chem. Biol.* **2011**.

19. Springsteen, G.; Wang, B. A detailed examination of boronic acid–diol complexation *Tetrahedron.* **2002**, *58*, 5291-5300.

20. Chen, L.; Frankel, A. D. A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 5077-5081.

21. http://prospector.ucsf.edu/prospector/mshome.htm.

22. This program was further revised and adapted for variable amino acid analogues and branched peptide length and is available for complimentary use with correspondence. .

23. http://weblogo.berkeley.edu/logo.cgi.

Chapter 6 Future Directions

6.1 Analysis of an additional 3.3.4 branched boronic acid library

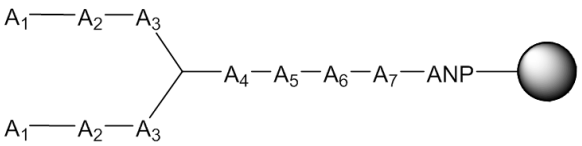
6.1.1 Contributions

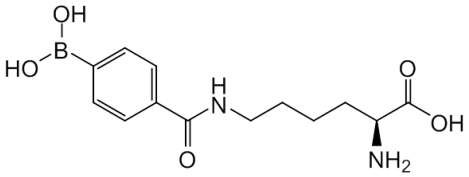
The following section outlines key points from a manuscript in preparation for publication. Contributions from the authors are as follows: Jason Crumpton, (author of this dissertation) performed the synthesis of the modified Fmoc-L-borono-phenylalanine amino acid (F_{BPA}) as well as the N-ε-(4-boronobenzoyl)-L-lysine (K_{BBA}) used for peptide synthesis, performed acquisition of all MALDI spectra, made intellectual contributions to the design of the 3.3.4 peptide library, made significant intellectual contributions to the construction of novel software for the *de novo* sequencing of unknown peptides, and performed the synthesis of selected peptides. David Bryson made intellectual contributions to the design of the 3.3.4 peptide library, performed all RNA footprint assays as well as footprint assay optimization, and made contributions towards experimental design. Wenyu Zhang performed the synthesis and purification of branched boronic acid (BB) peptides, made contributions towards experimental design, and performed all K_d experiments by Dot-Blot and EMSA assays. Jessica Wynn performed the synthesis of selected peptides, as well as all cell permeability and MTT toxicity assays. Dr. Webster Santos made contributions to the writing and editing of the manuscript as well as significant contributions to experimental design.

6.1.2 Modifications of the previous 4.4.4 BB library

This chapter is a summary of a manuscript currently in preparation. During the preparation of this dissertation, a second generation 3.3.4 BB peptide library was synthesized with several modifications to the first generation 4.4.4 library design

presented in Chapter 5 of this dissertation (**Figure 6.1**). The library was prepared such that there were three variable amino acid positions at both the N- and C-termini (A_1-A_3 and A_4-A_6, respectively), and each variable position was composed of six possible side chains



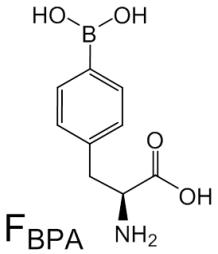


K_{BBA}

3.3.4. Library (6 variable positions)

A_1	A_2	A_3	A_4	A_5	A_6	A_7
Trp	Tyr	Phe	His	Tyr	Trp	Tyr
Ser	Asn	Thr	Gln	Thr	Ser	Tyr
Val	Pro	Ala	Pro	Leu	Ala	Tyr
Lys	Lys	Lys	Lys	Lys	Lys	Tyr
K_{BBA}	K_{BBA}	K_{BBA}	K_{BBA}	K_{BBA}	K_{BBA}	Tyr
F_{BPA}	F_{BPA}	F_{BPA}	F_{BPA}	F_{BPA}	F_{BPA}	Tyr

6^6=46,656 library members



F_{BPA}

Figure 6.1 Combinatorial (3.3.4) branched-peptide library featuring boronic acid side chains

Each of the six possible side chains at variable positions A_1-A_6 was chosen for its potential to interact with the RRE-IIB target RNA through a unique mode of binding relative to the other side chains we made available at the same location. In position A_1 for example, we selected amino acids with functional groups that can interact with the RNA through electrostatic attraction (Lys), hydrogen bonding (Ser), π–stacking (Trp), and reversible covalent bonding between boron and a Lewis base presented by the RNA target (K_{BBA}/F_{BPA}).

The rational for including unnatural boronic acid side chains was again three-fold. First, we hypothesized that their inclusion would increase the selectivity of our compounds for RNA over DNA due to the presence of a 2'-hydroxyl that is absent in DNA, which may function as a potential electron donor to the empty p-orbital of boron. complexity of these compounds. This could improve the chances of discovering ligands that are selective for a single RNA target. Third, we envisioned that the presence of boron would increase overall affinity to the RNA target due to the potential formation of reversible covalent bonds between the RNA and BB ligand. We chose to incorporate two different boron-containing side chains at each variable position in the library to examine if RRE-IIB had a preference in the side chain length or electrophility of boron presented by F_{BPA} and K_{BBA}.

In order to ensure that each unique sequence in the library would contain at least one chromophore, Tyr was included at the C-terminus (position A_7) to allow facile determination of peptide concentrations for future biophysical assays. Thereafter, we generated three copies of a 3.3.4 branched boronic acid peptide library composed of 46,656 possible amino acid sequences that were linked to the bead by a photocleavable linker (3-amino-3-(2-nitrophenyl)propionic acid, ANP) at the C-terminus. The library was incubated with fluorescently tagged RRE-IIB RNA and putative hit peptides were isolated and their sequences were deconvoluted by *de novo* sequencing.

Table 6.1 Selected hit peptides

Peptide	Sequence	K_d(μM)
6.1	$(VPA)_2$*$\mathbf{F_{BPA}}$LAY	NB
6.2	$(KNT)_2$*NK$\mathbf{K_{BPL}}$Y	NB
6.3	$(VPA)_2$*N$\mathbf{F_{BPA}}$AY	NB
6.4	($\mathbf{K_{BPL}}$YK)$_2$*HKKY	86.5 ± 10.0
6.5	(K$\mathbf{K_{BPL}}$K)$_2$*KLKY	58.4 ± 4.0
6.6	$(WYK)_2$*PTWY	28.5 ± 4.4
6.7	(K$\mathbf{K_{BPL}}$F)$_2$*KKWY	27.2 ± 6.9
6.8	(KK$\mathbf{K_{BPL}}$)$_2$*$\mathbf{F_{BPA}}$TSY	26.8 ± 4.4
6.9	($\mathbf{K_{BPL}}$K$\mathbf{F_{BPA}}$)$_2$*$\mathbf{K_{BPL}}$KKY	3.3 ± 1.2
6.10	$(WKK)_2$*$\mathbf{K_{BPL}}$YWY	1.4 ± 0.4
6.11	($\mathbf{F_{BPA}}$Y$\mathbf{F_{BPA}}$)$_2$*NKSY	8.7 ± 2.3
6.12	$(FYF)_2$*NKSY	NB

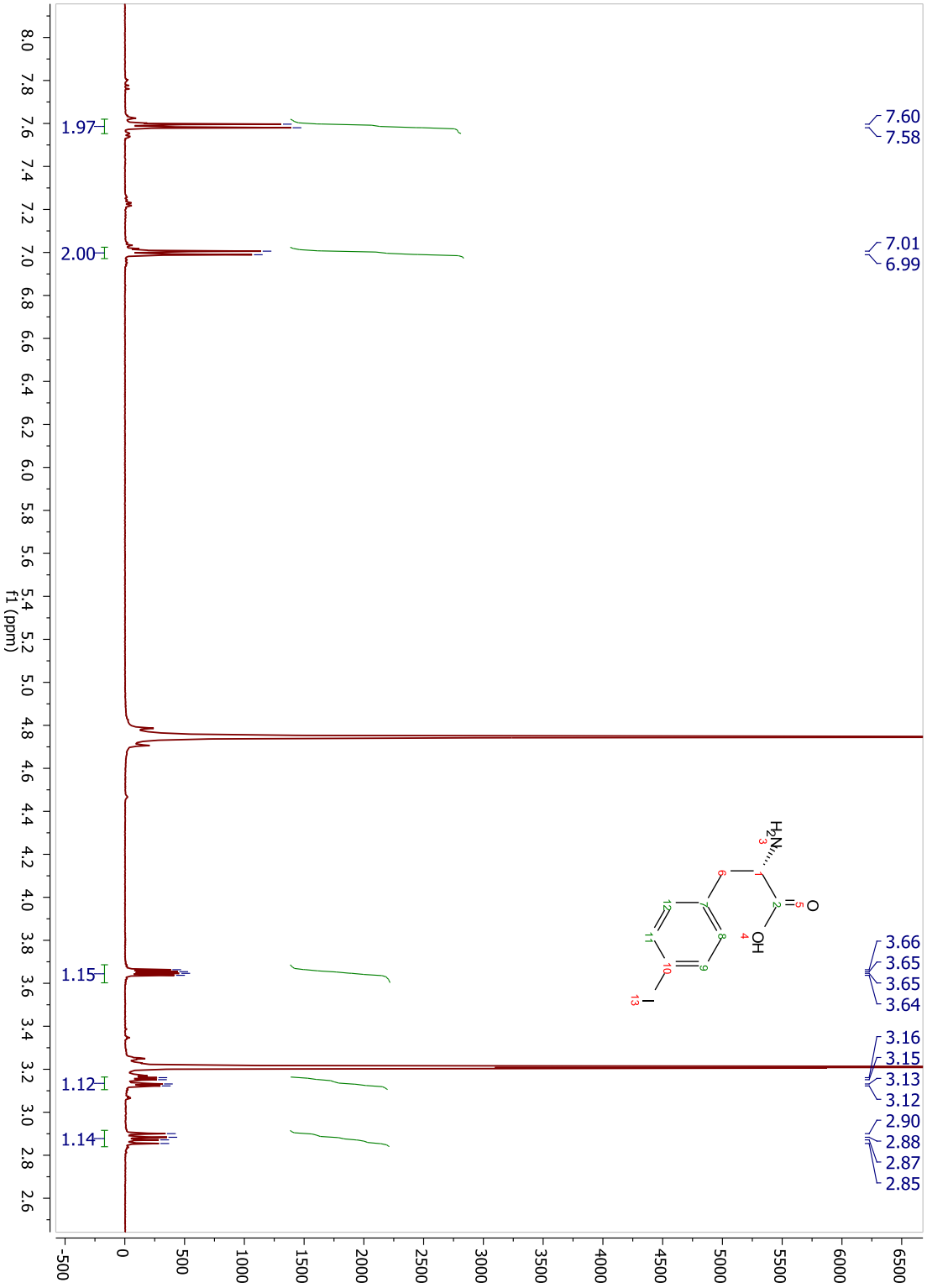
NB = no binding

As shown in Table 6.1, all peptides, excluding **6.6**, were found to contain varying degrees of K_{BBA} or F_{BPA} residues. Peptides **6.1** – **6.3** displayed no binding to RRE-IIB whereas peptides **6.4** – **6.11** displayed K_d values ranging from ~ 1 – 100 μM. Interestingly, when peptide **6.11** was mutated by replacing F_{BPA} residues with

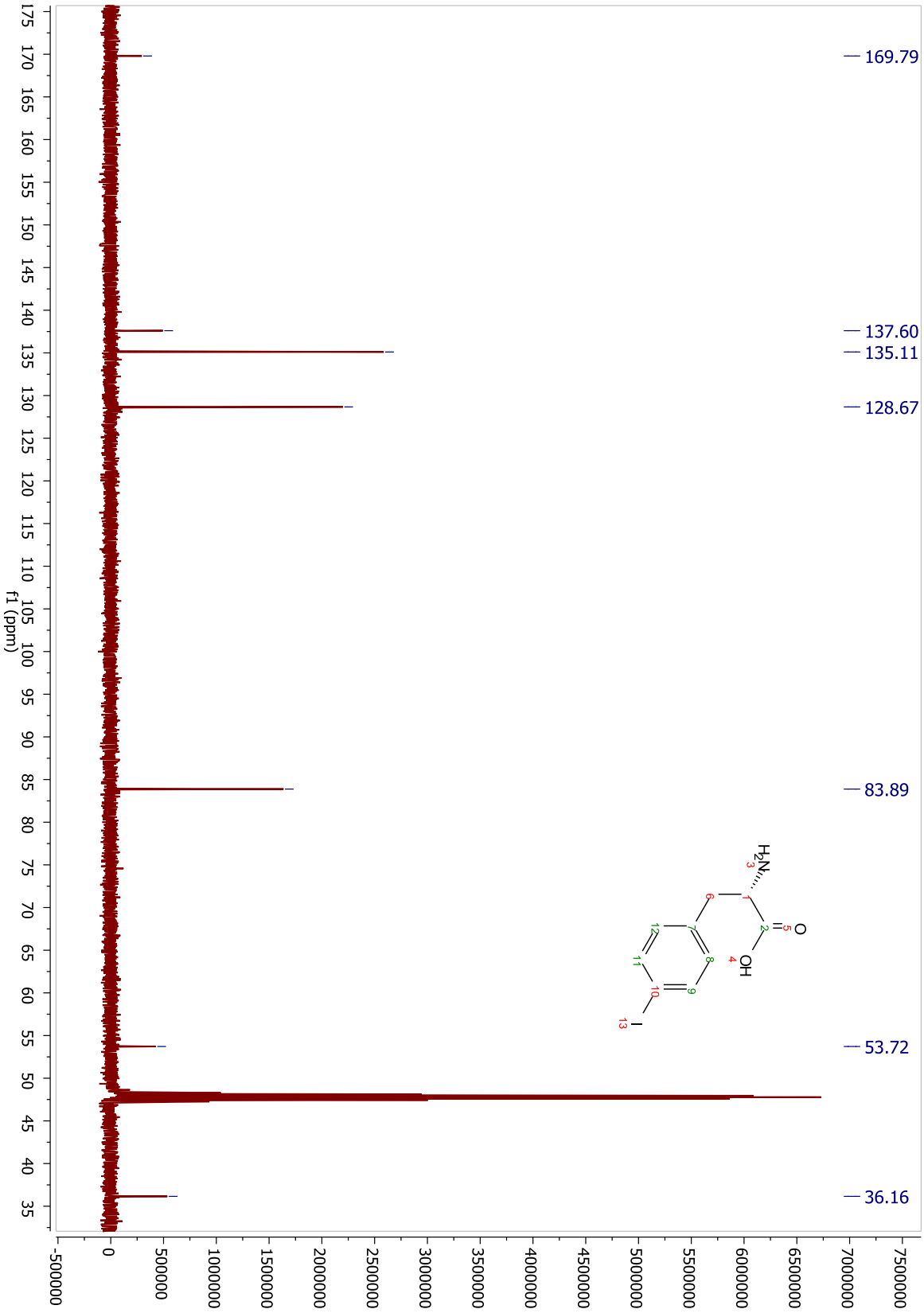
phenylalanine (peptide **6.12**), binding to RRE-IIB was completely abolished. Binding affinity studies of hit peptides to RRE-IIB as well as cell permeability, MTT toxicity, and RNA footprint assays were performed and further analyses of the results from these studies are currently underway. Complete details regarding experimental design, binding studies, and results will be explored in a forthcoming manuscript.

Appendix A: Supplementary Spectra for Chapter 4

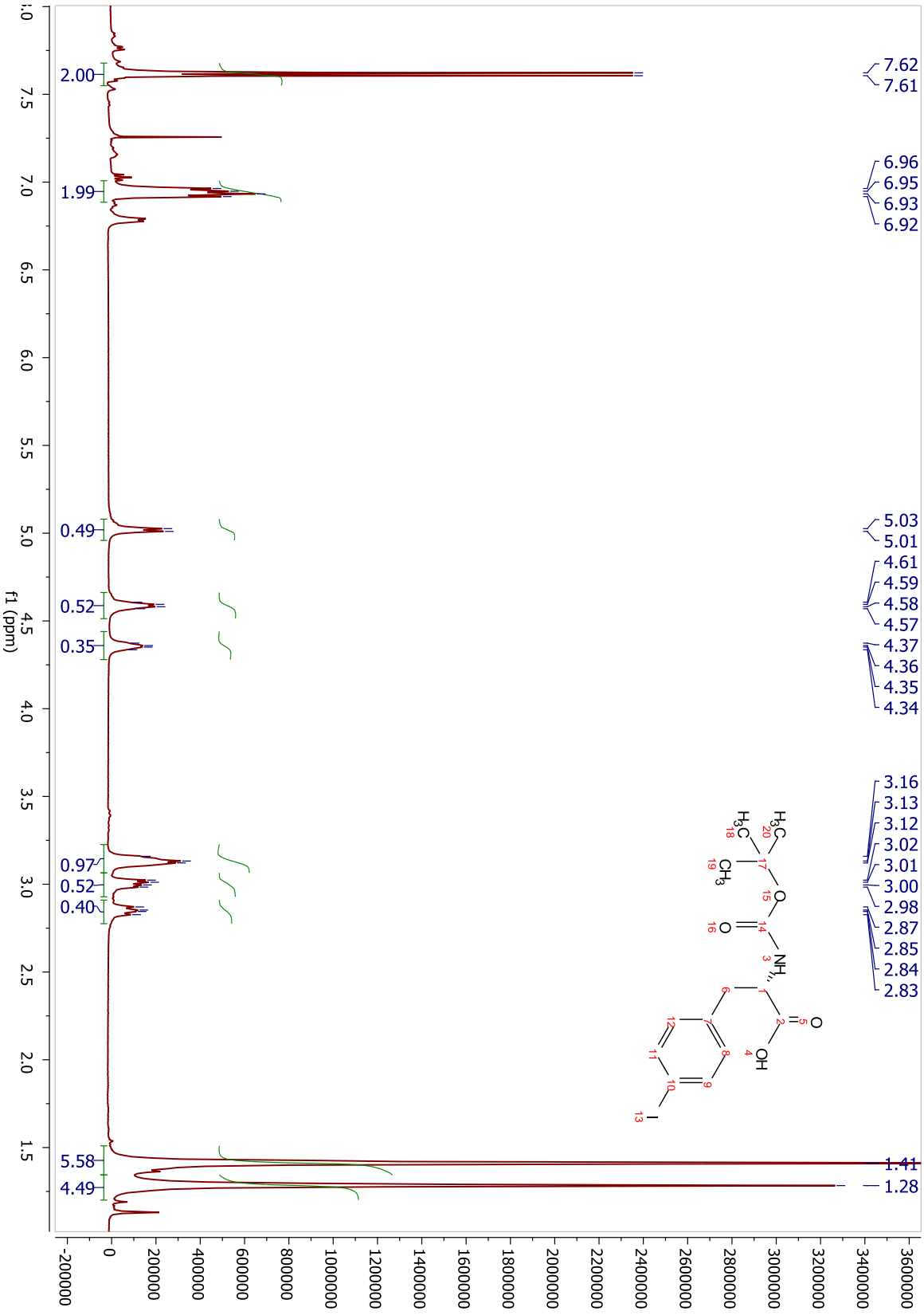
Compound 4.15



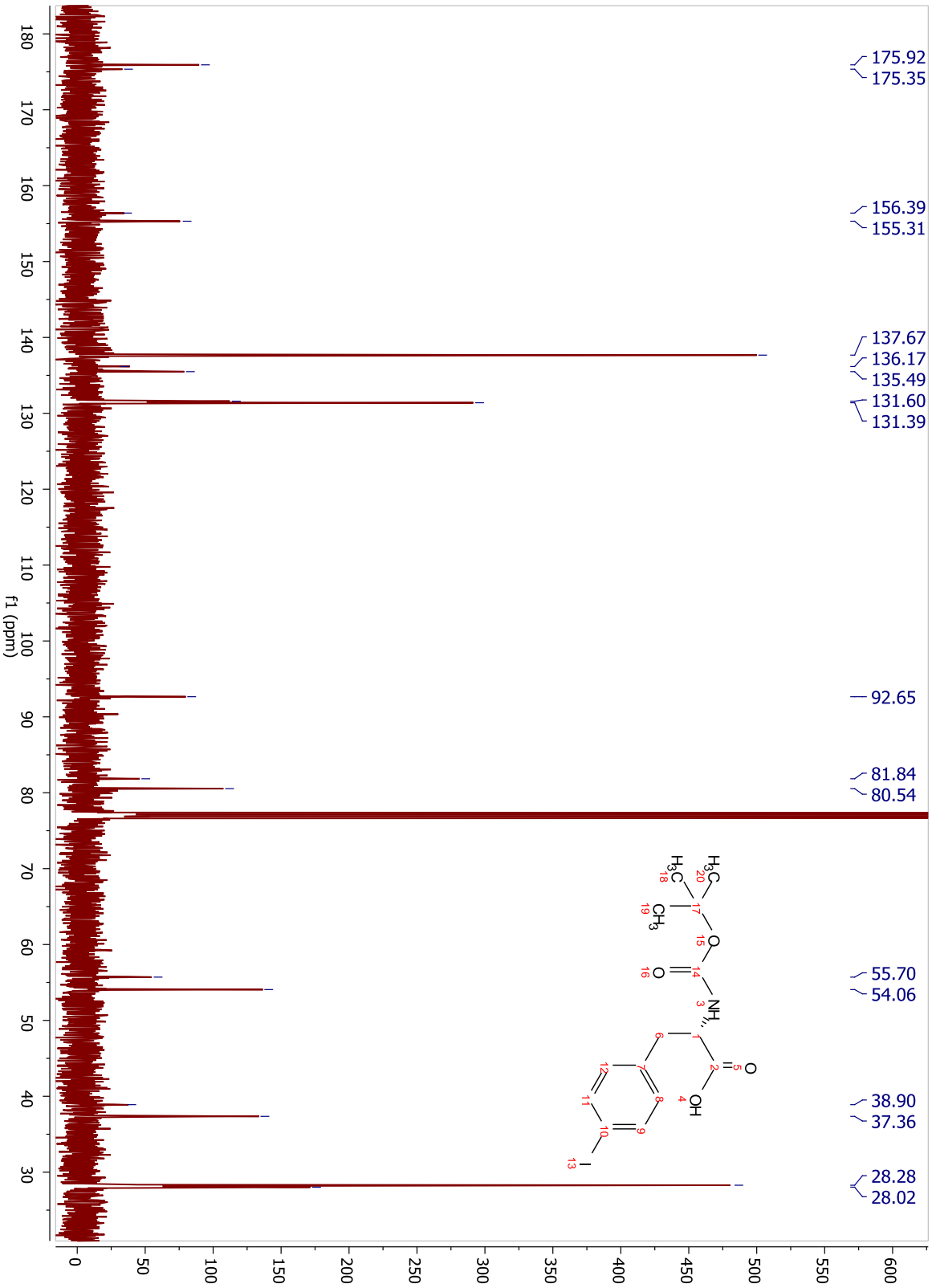
Compound 4.15



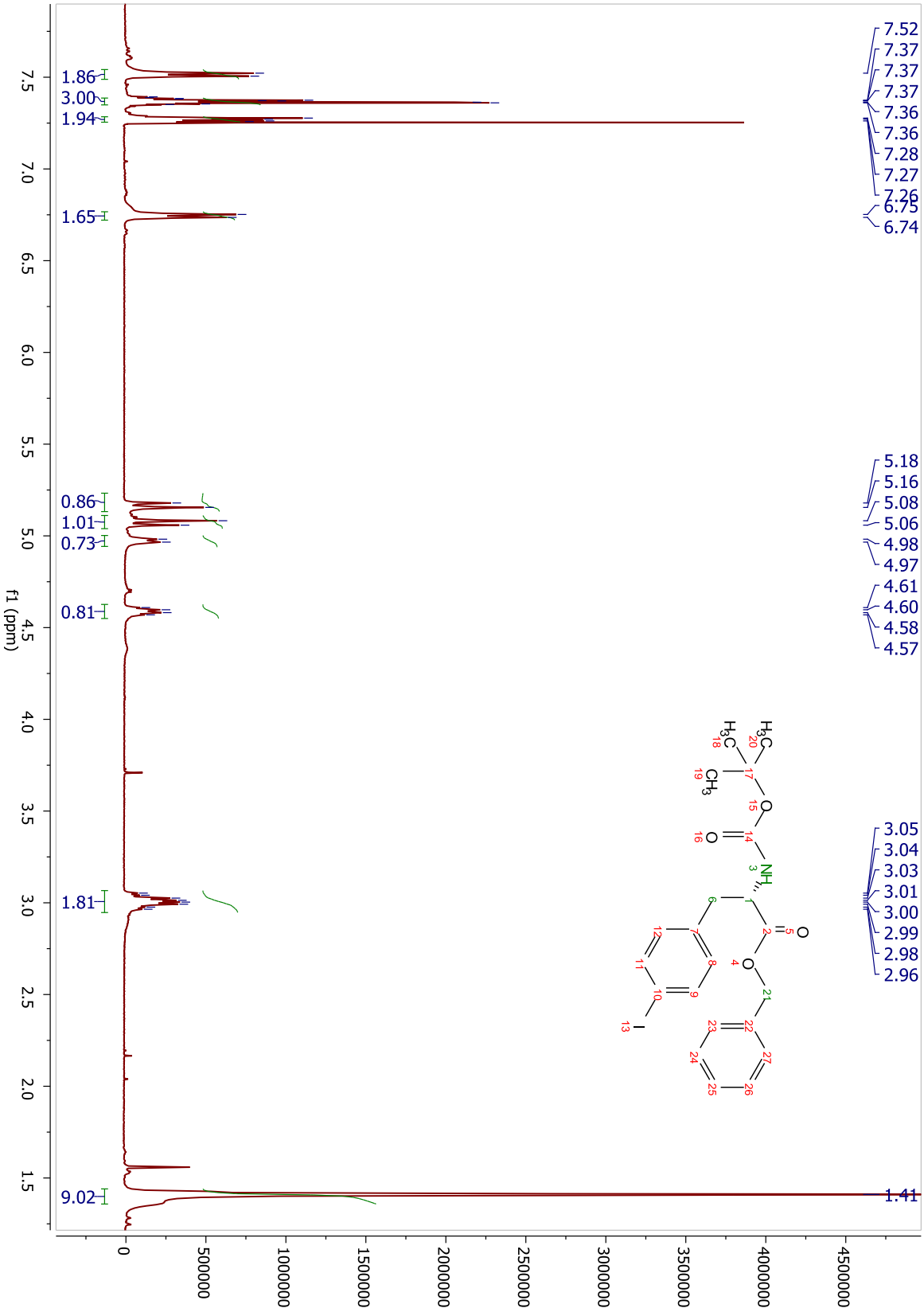
Compound 4.16



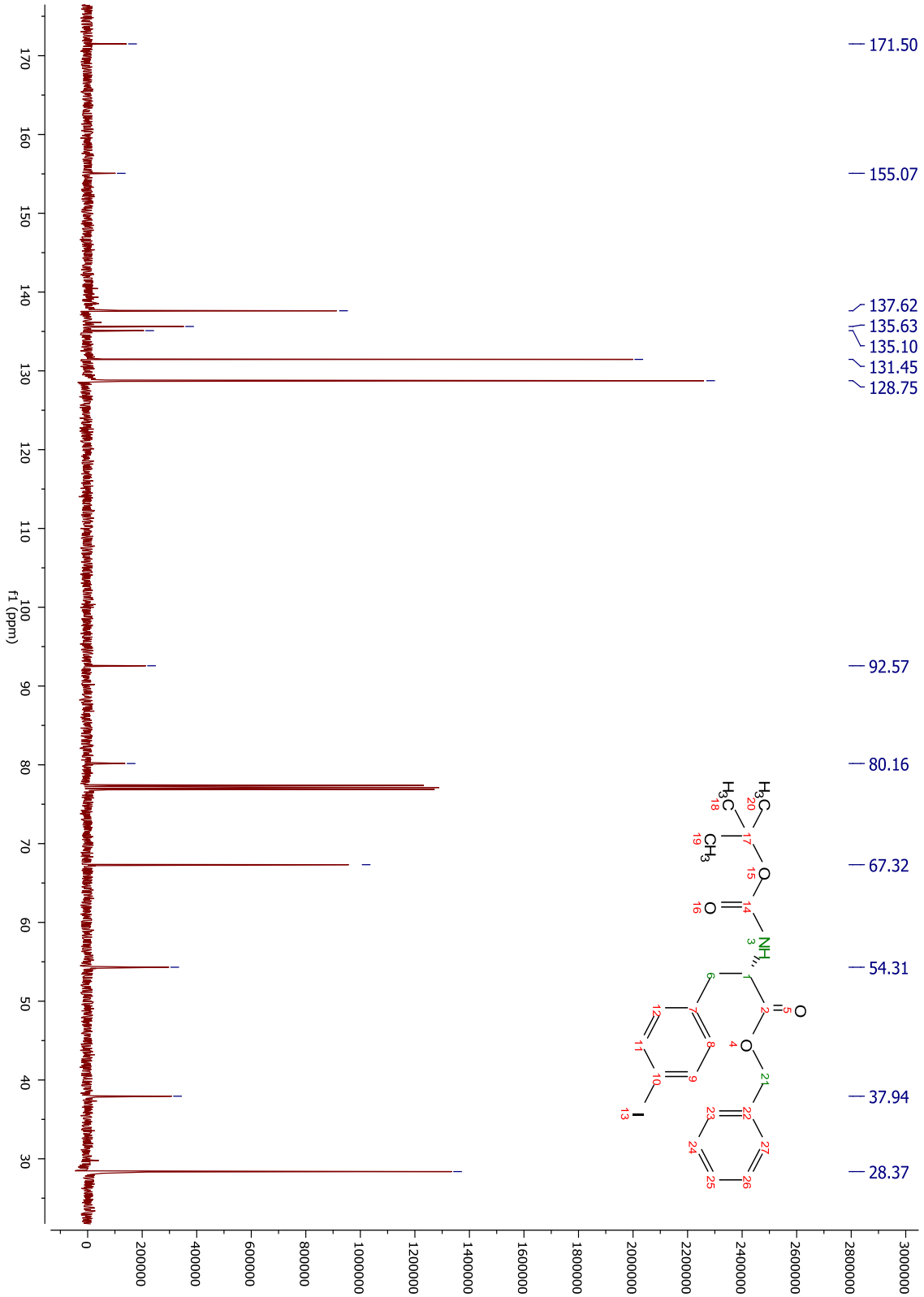
Compound 4.16



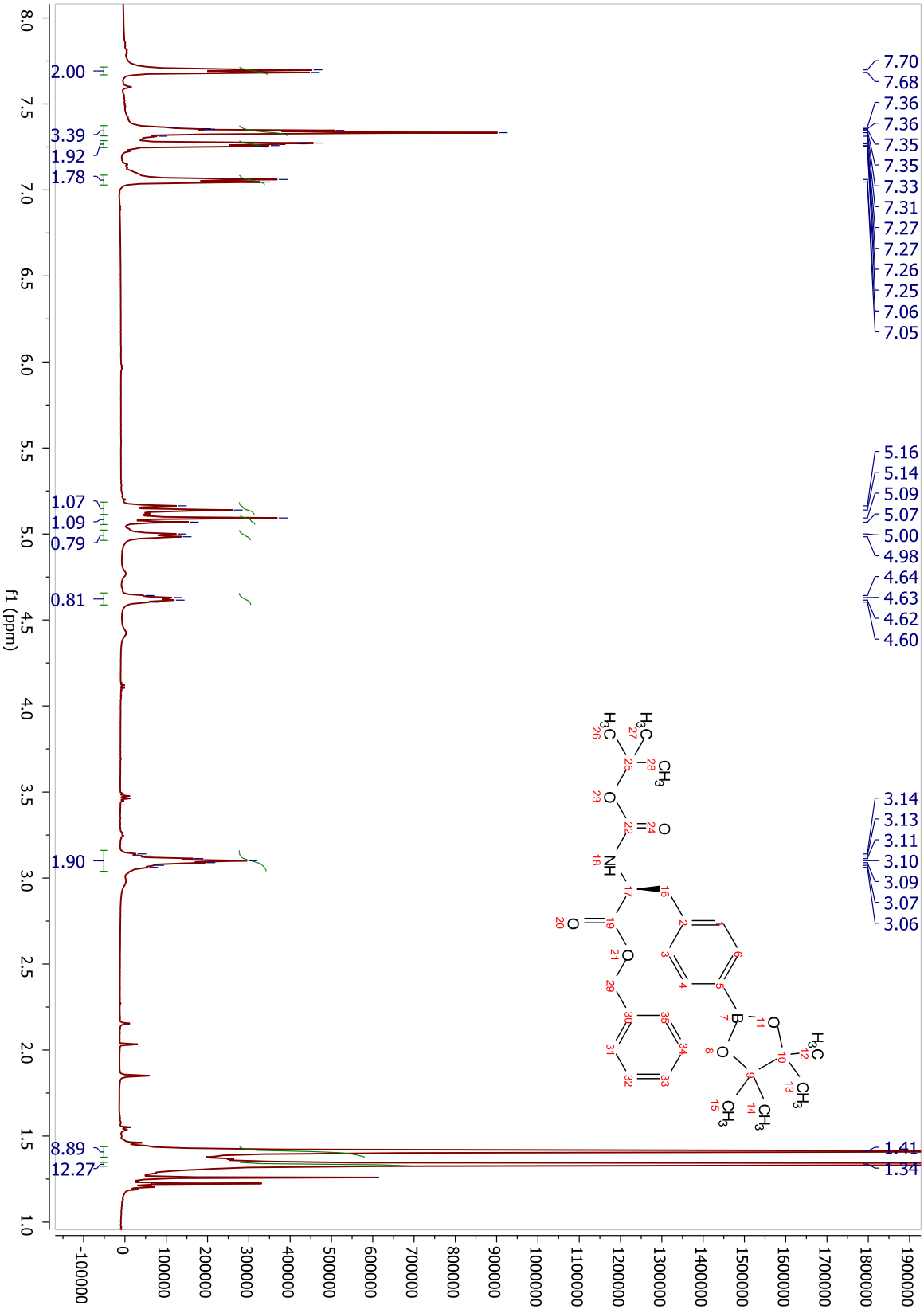
Compound 4.17



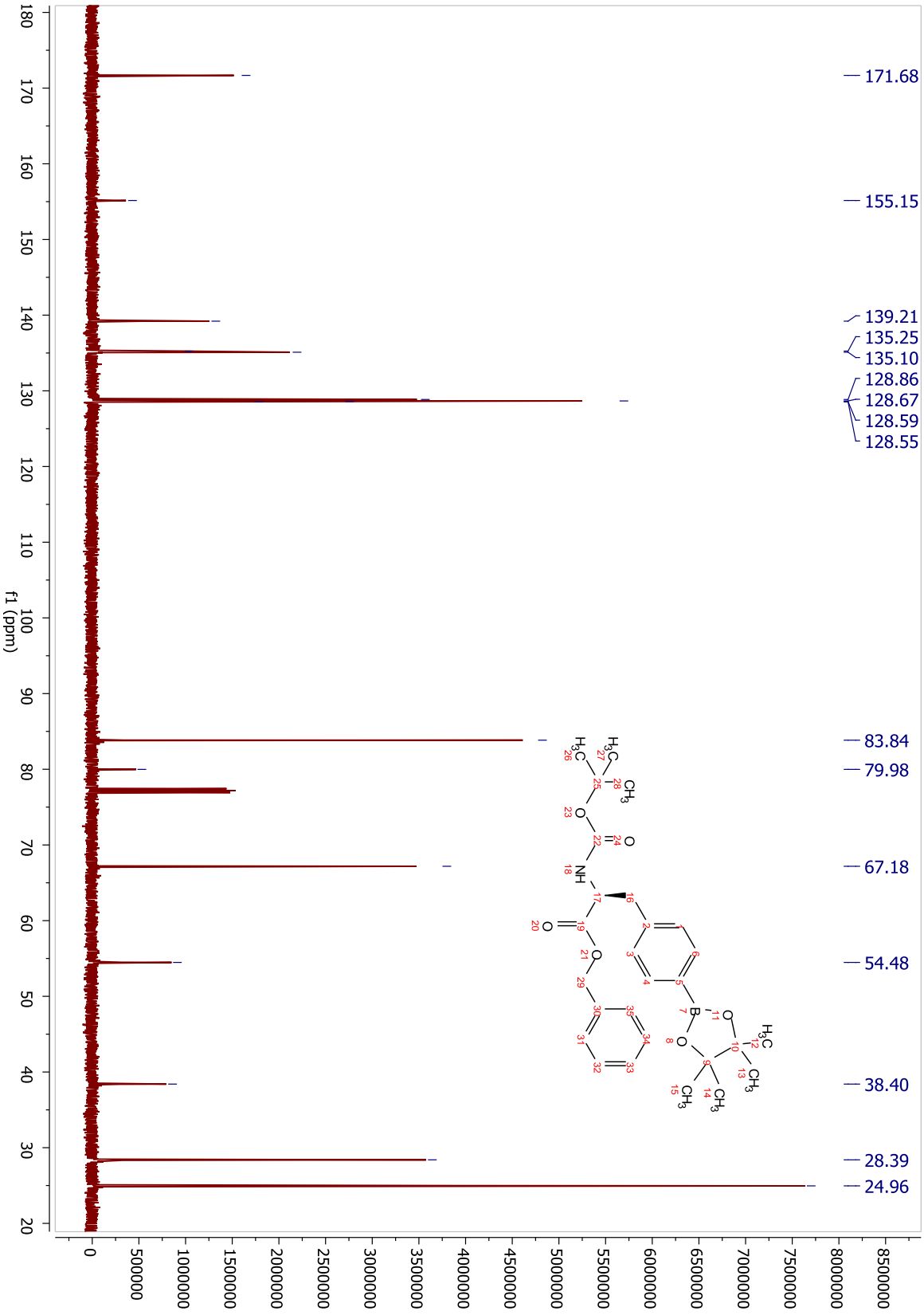
Compound 4.17



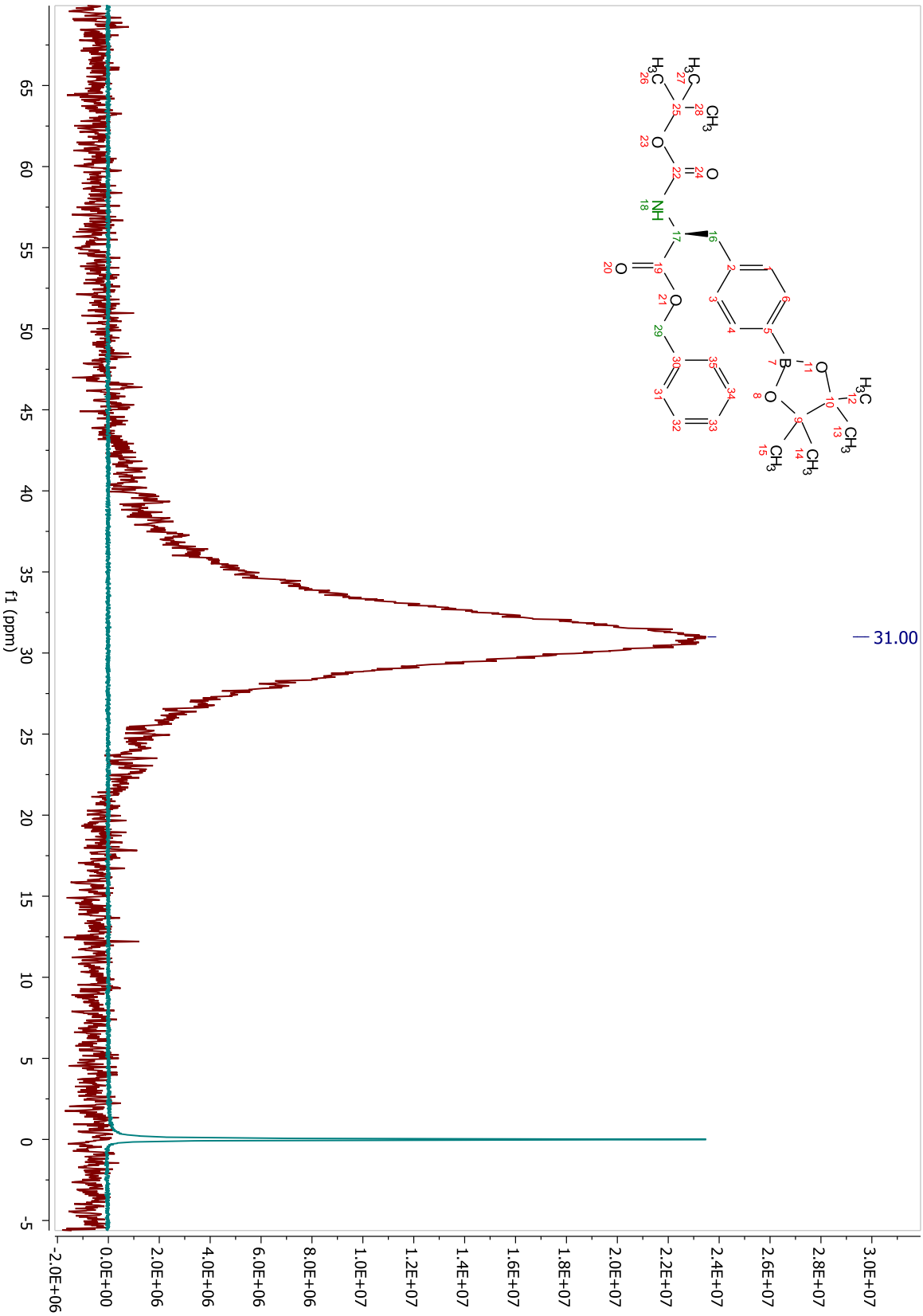
Compound 4.18



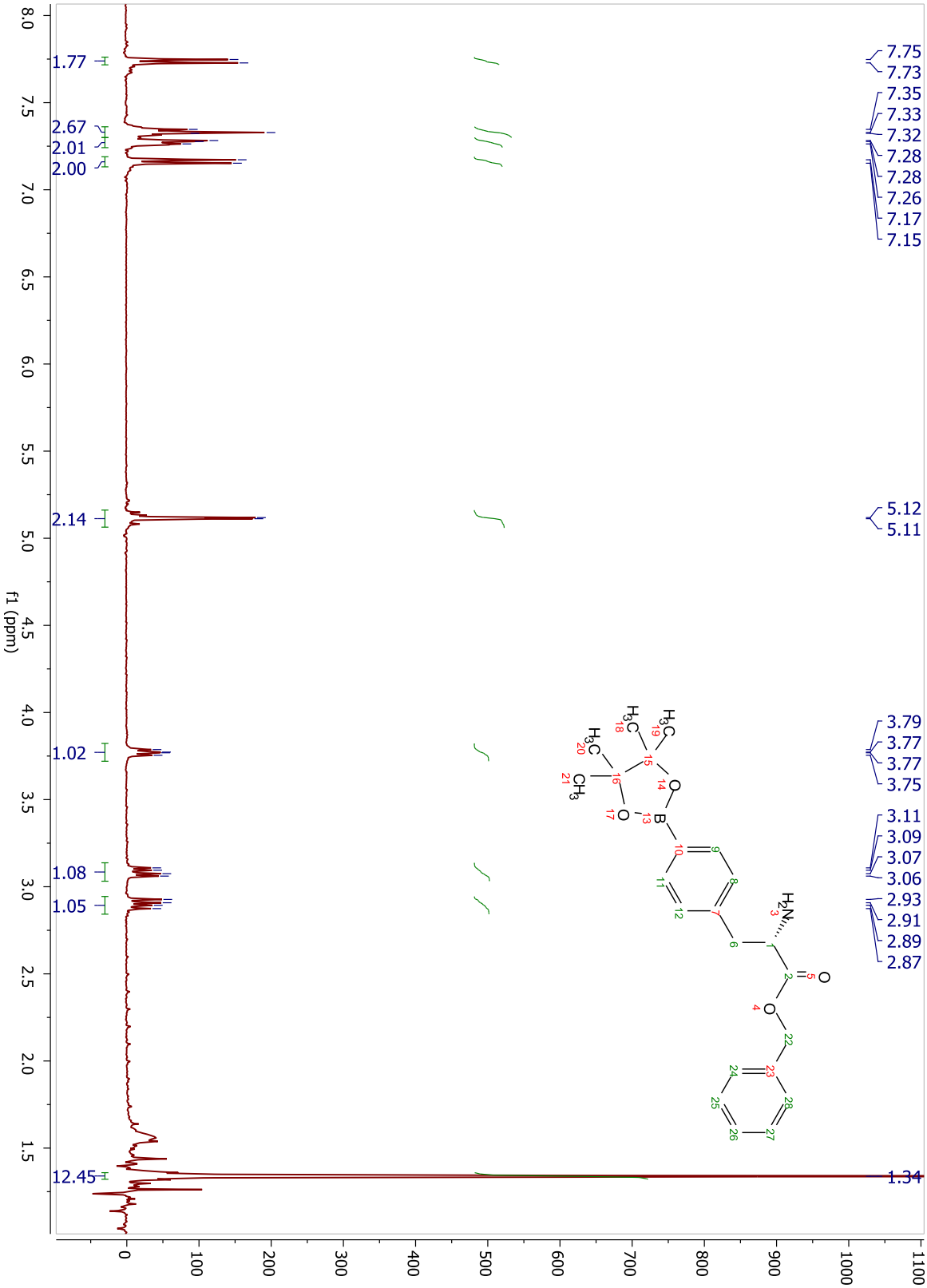
Compound 4.18



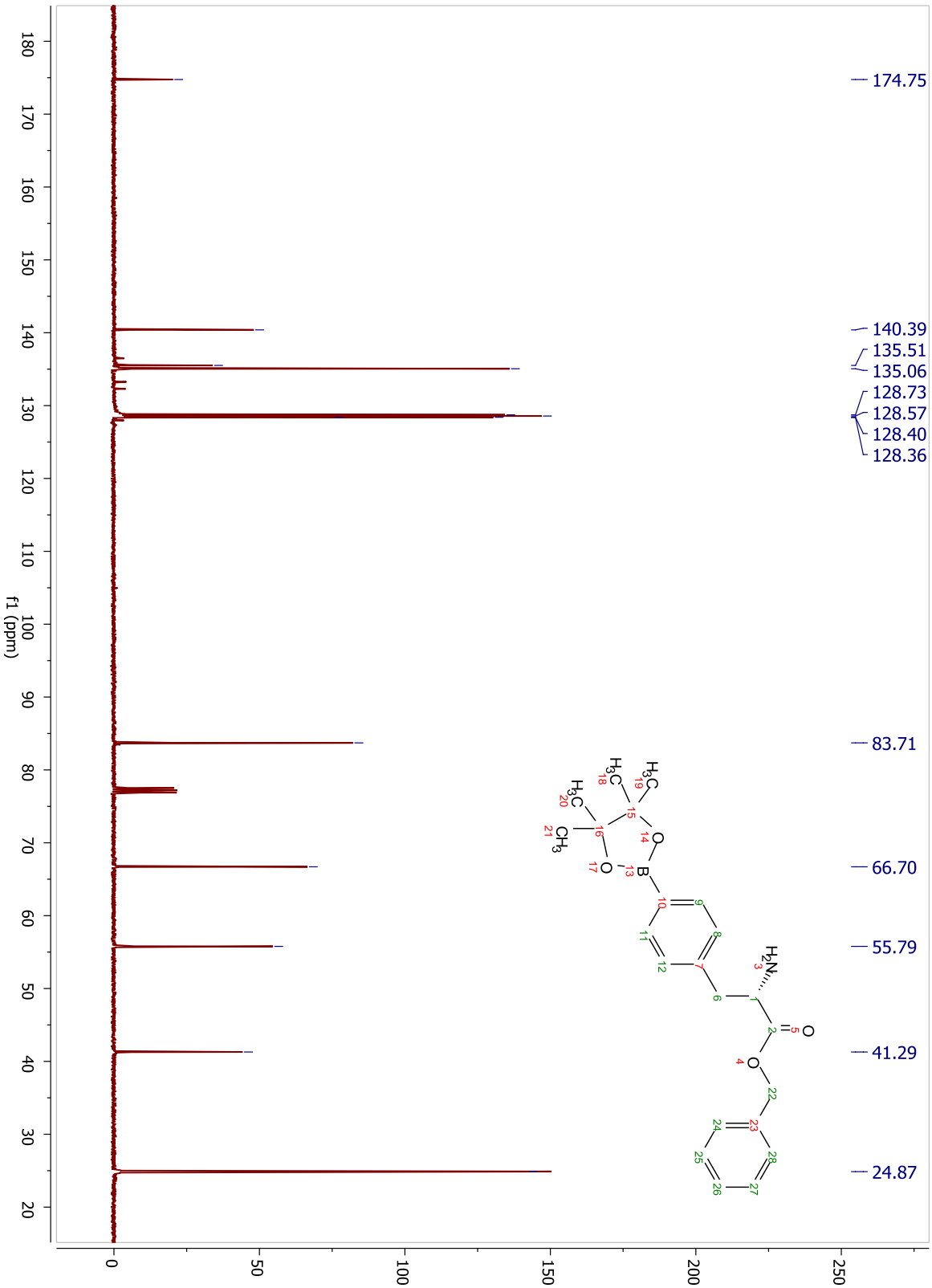
Compound 4.18



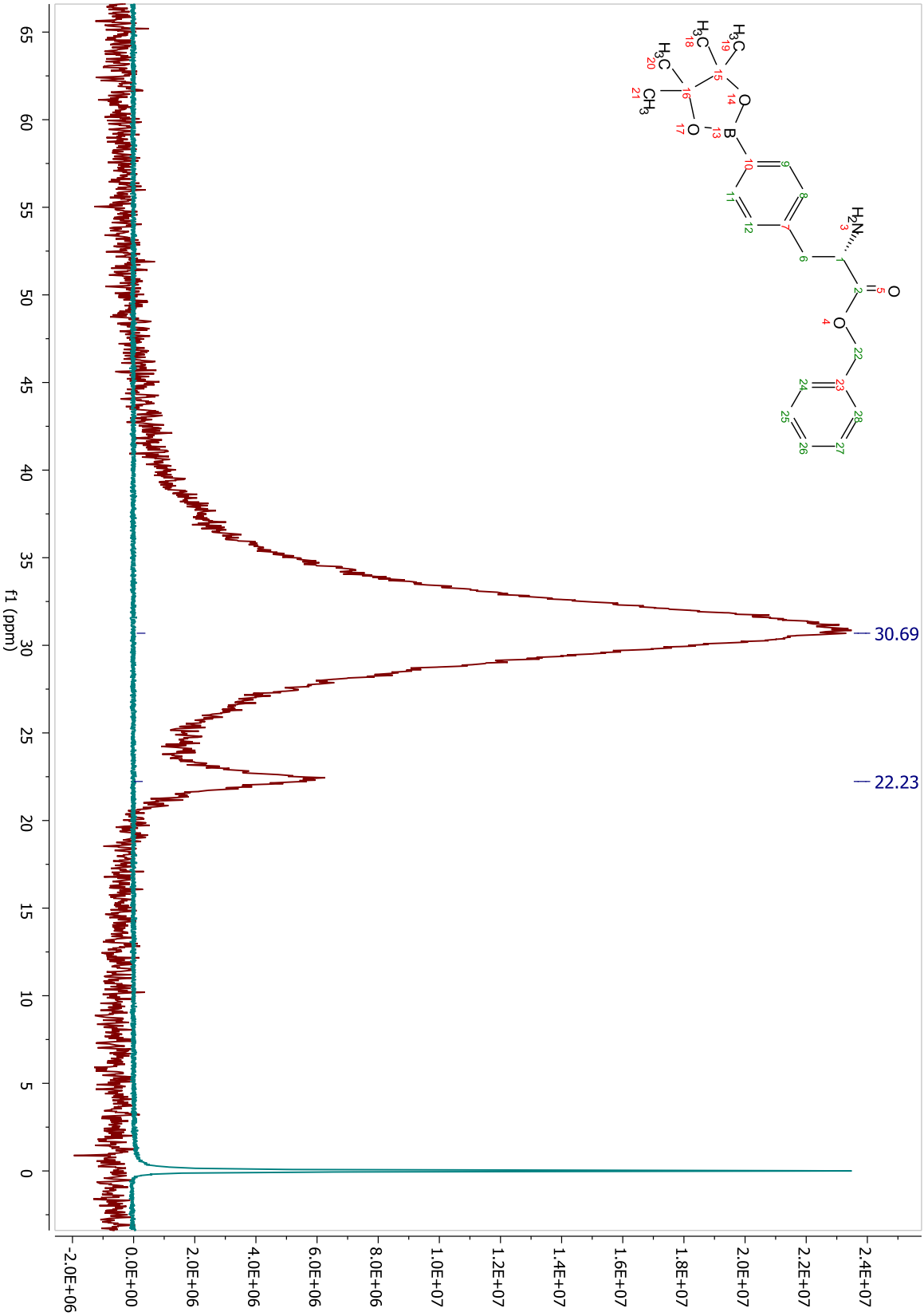
Compound 4.19



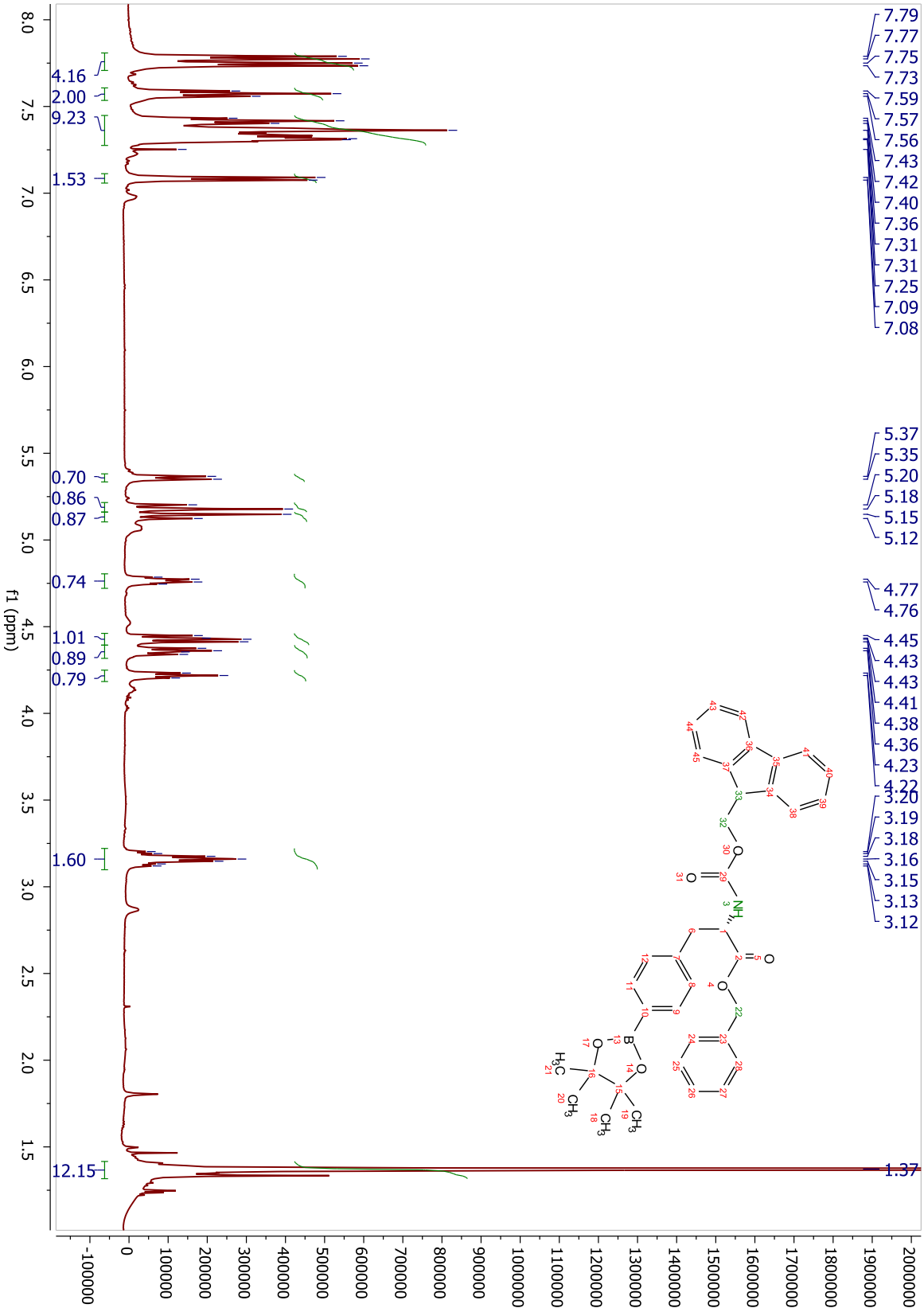
Compound 4.19



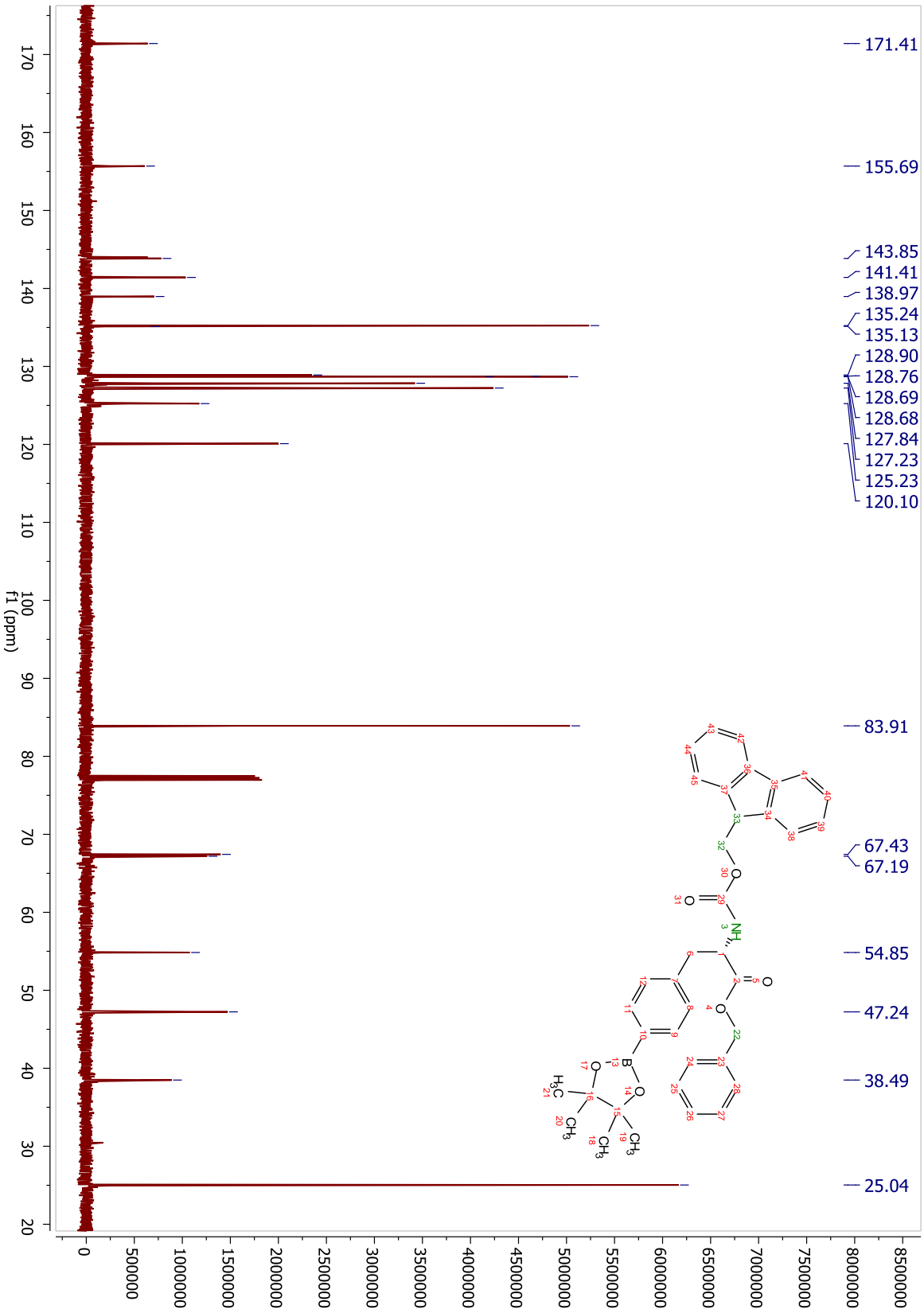
Compound 4.19



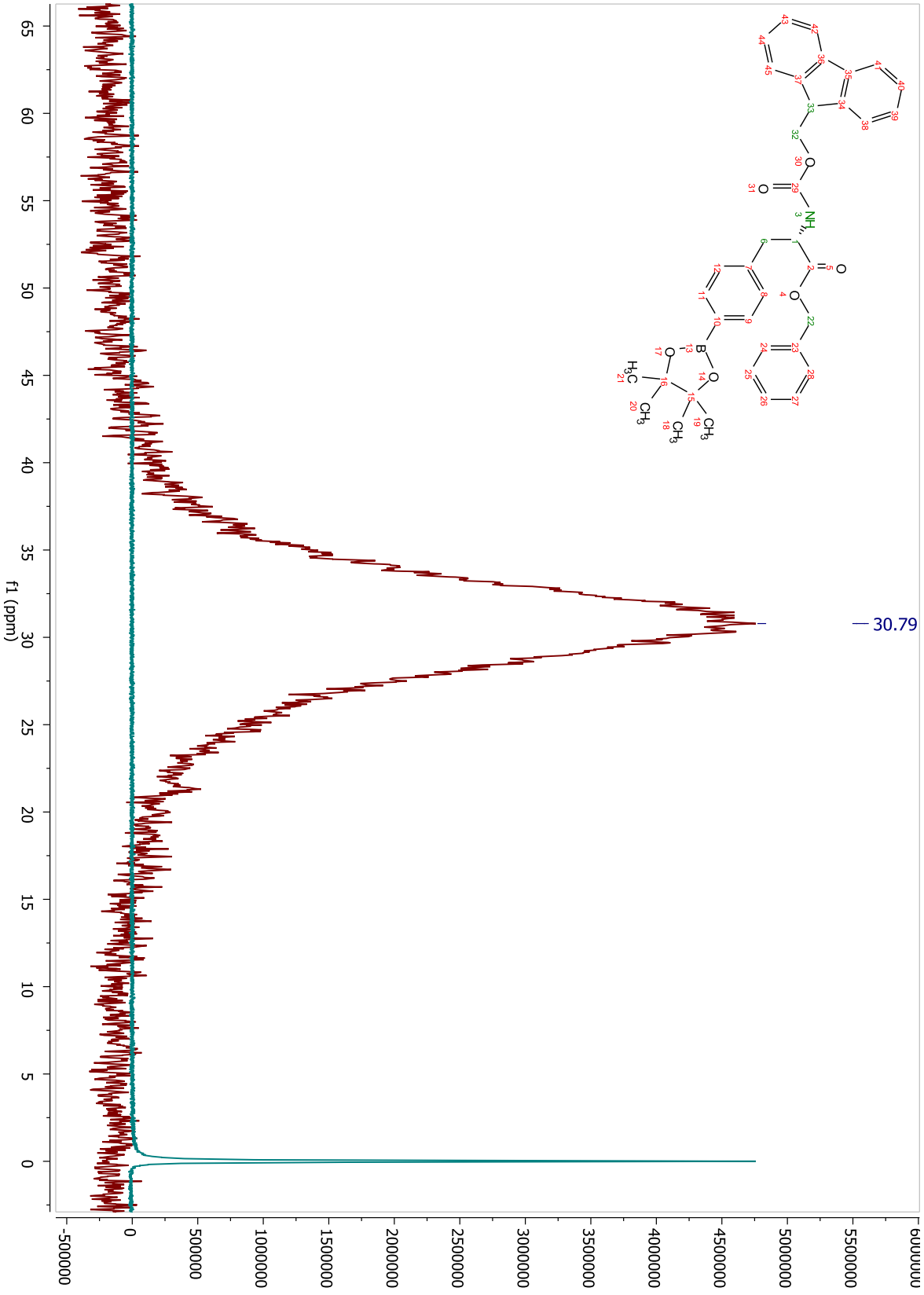
Compound 4.20



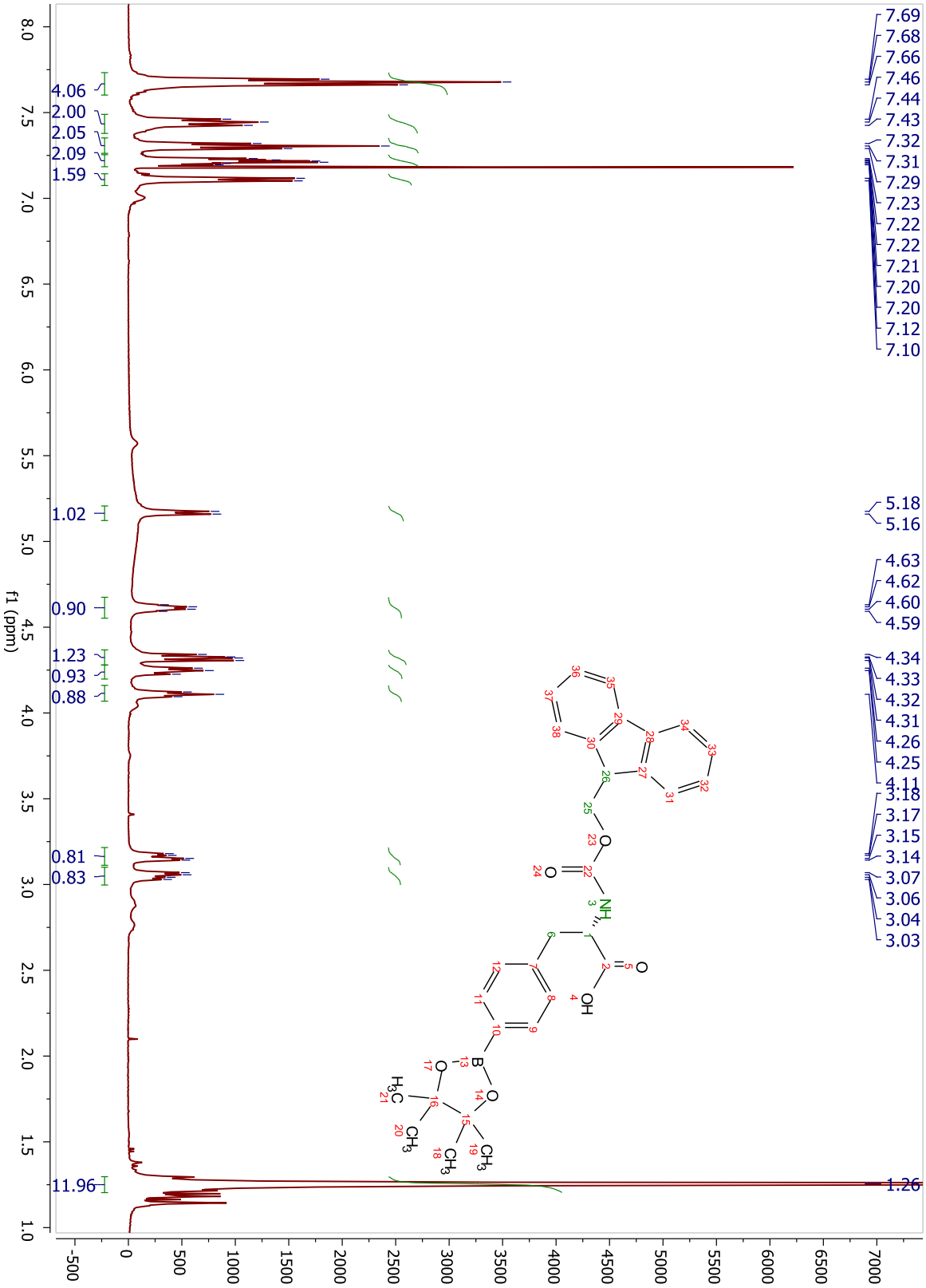
Compound 4.20



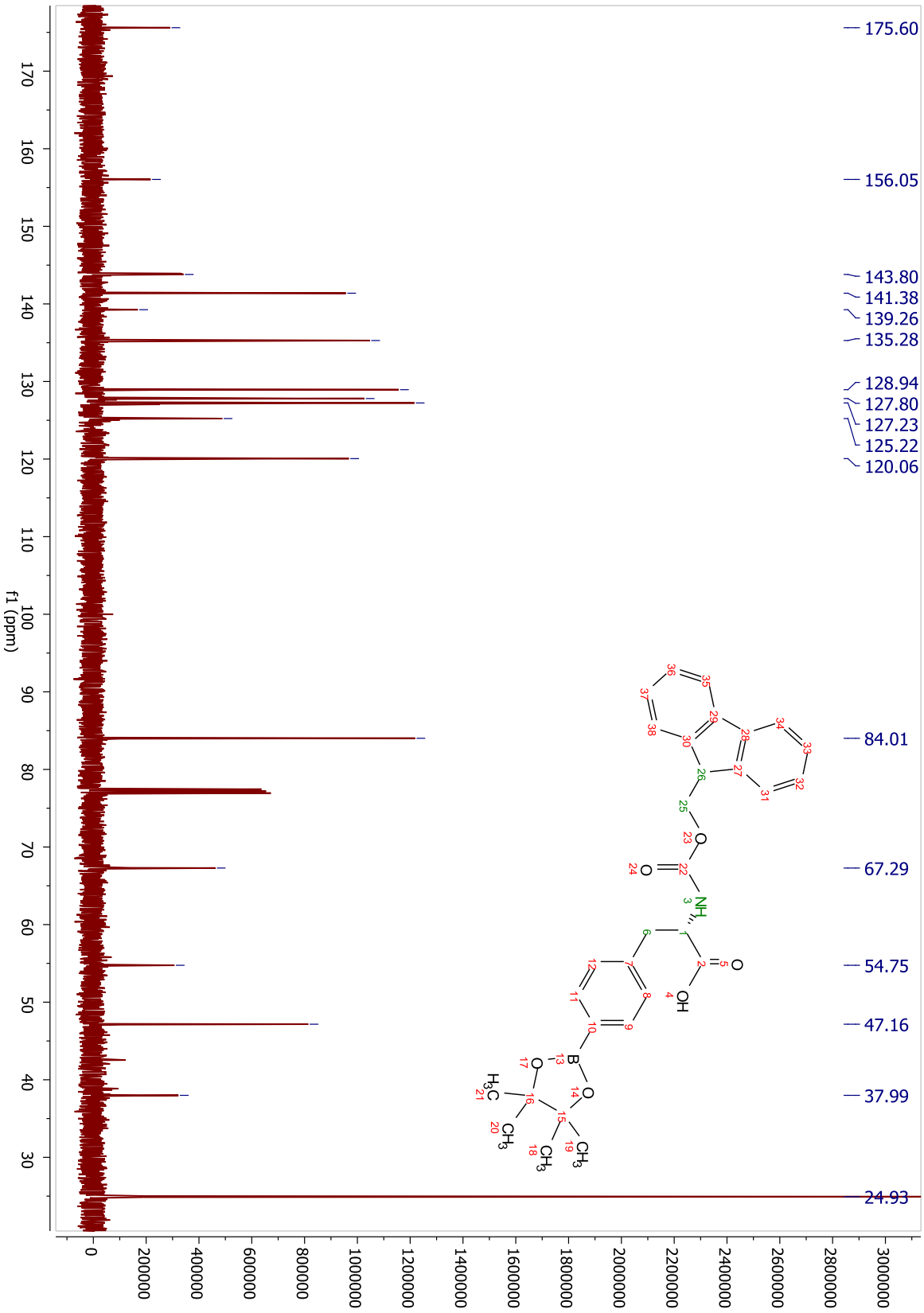
Compound 4.20



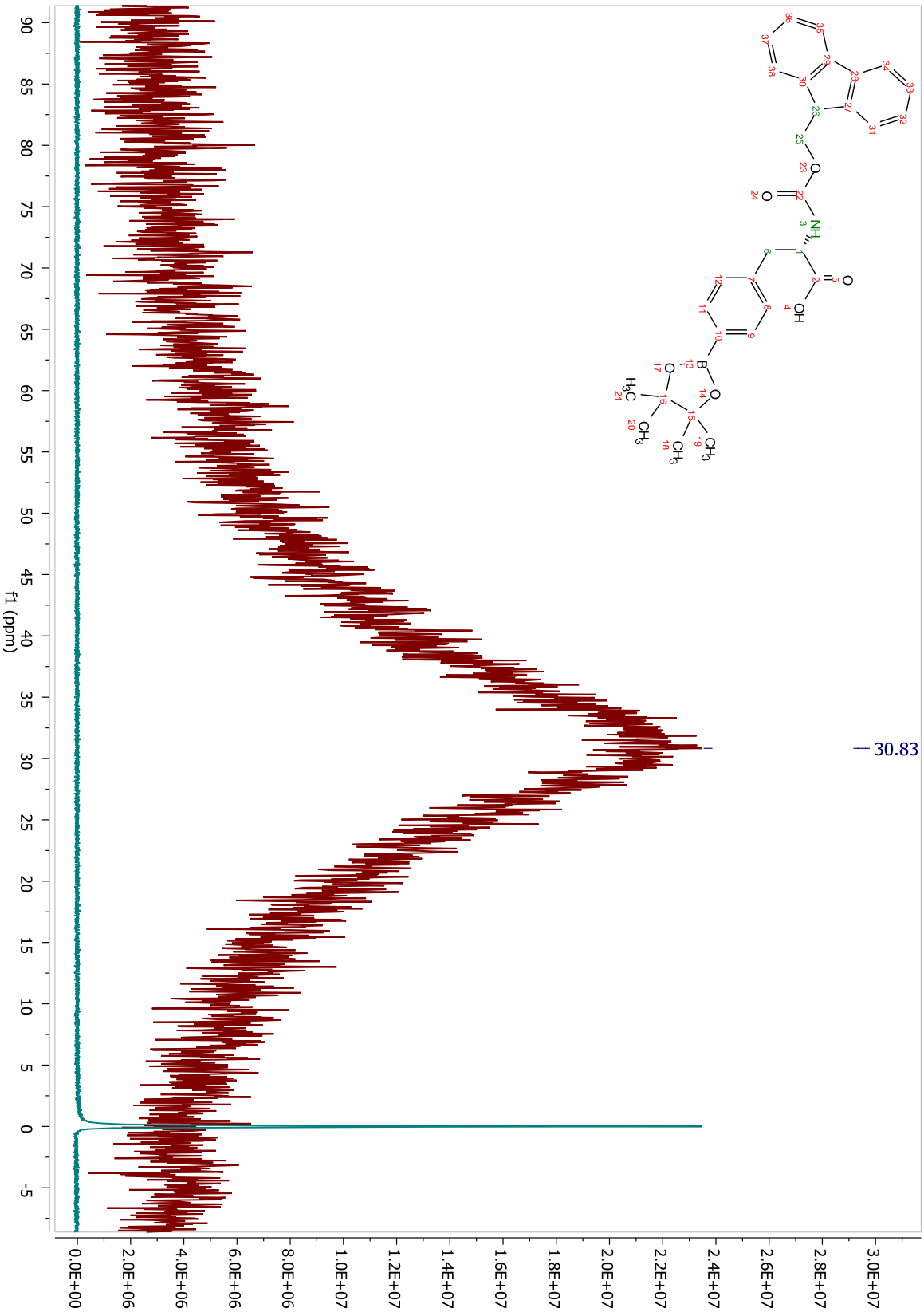
Compound 4.21



Compound 4.21

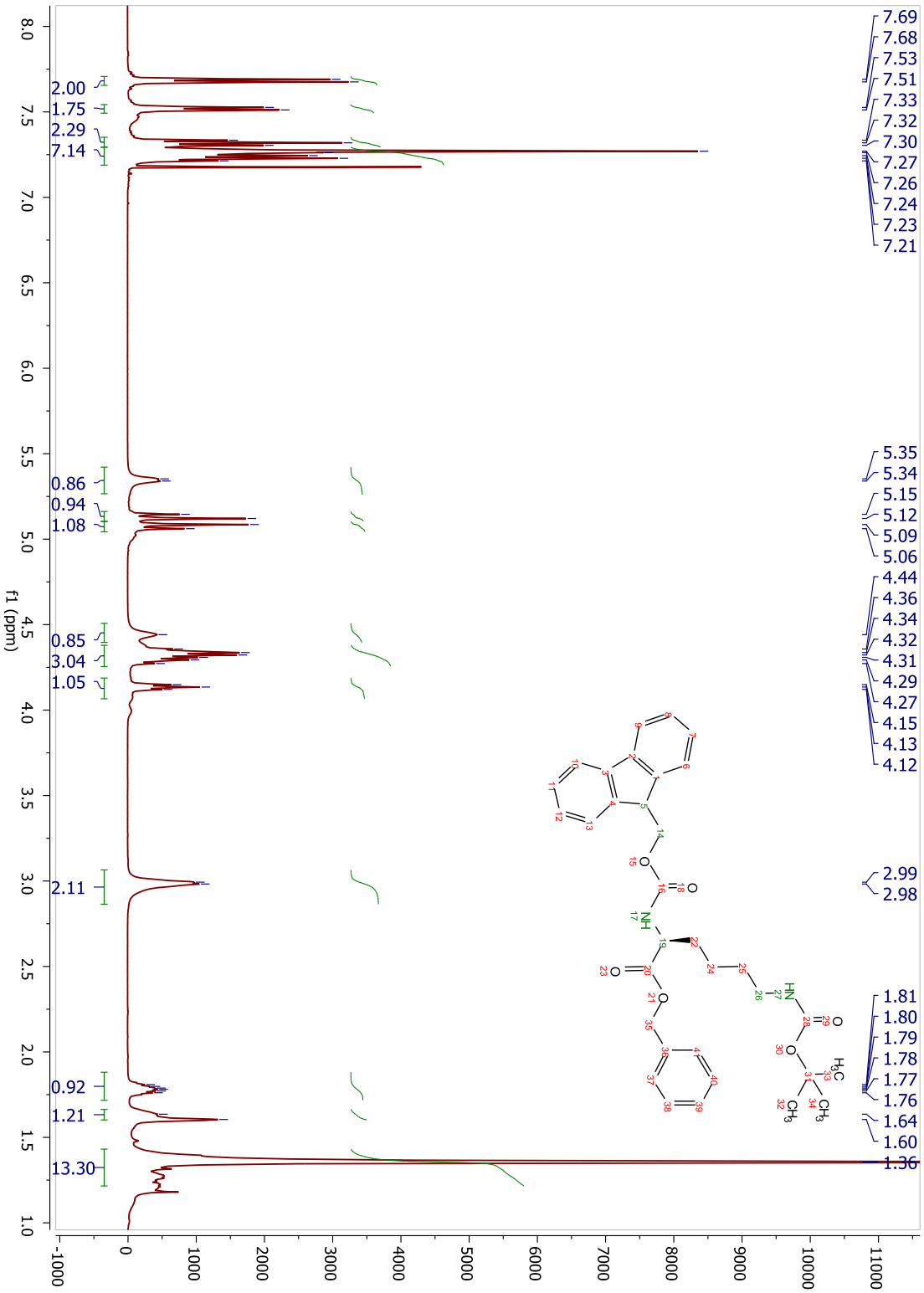


Compound 4.21

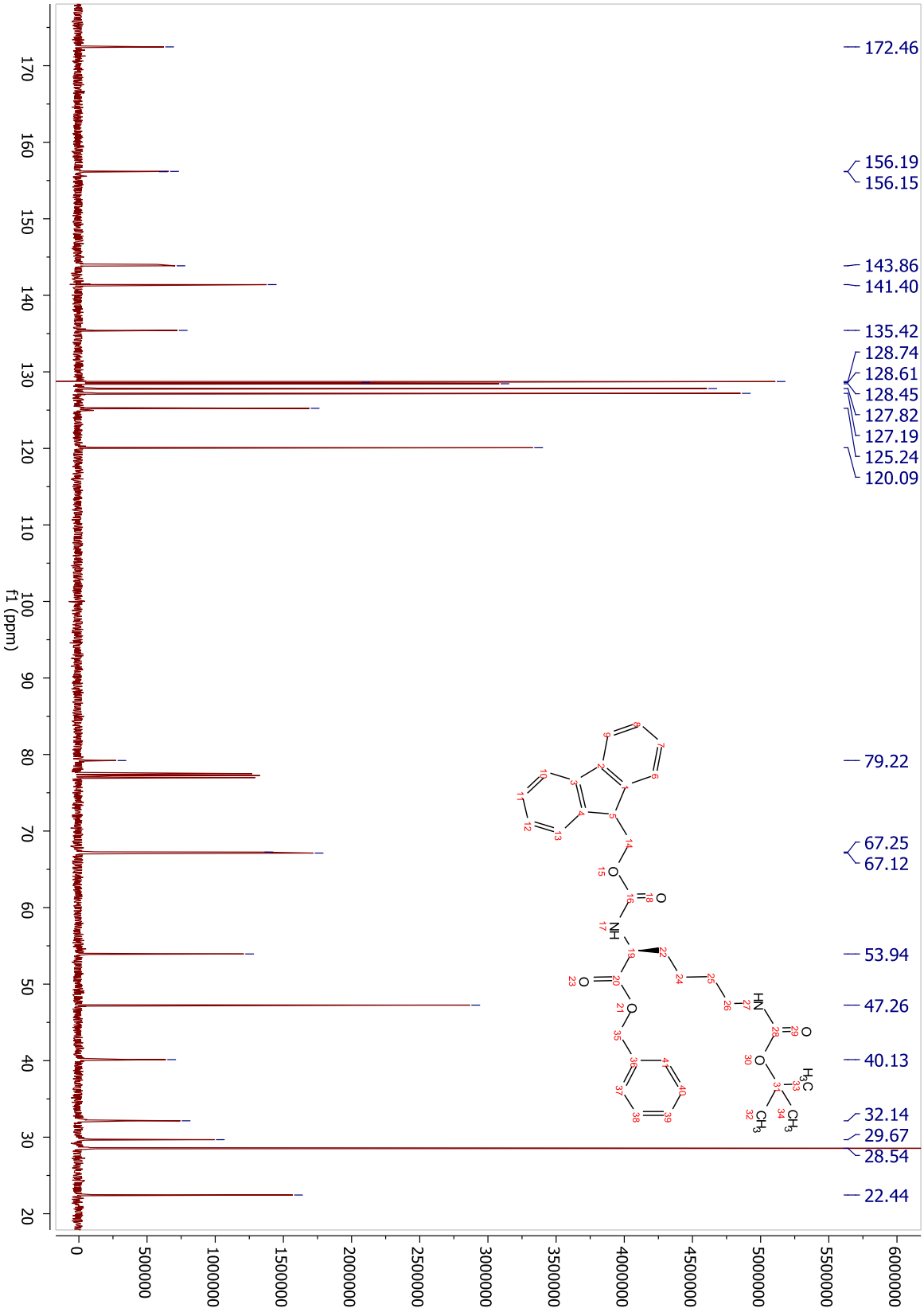


Appendix B: Supplementary Spectra for Chapter 5

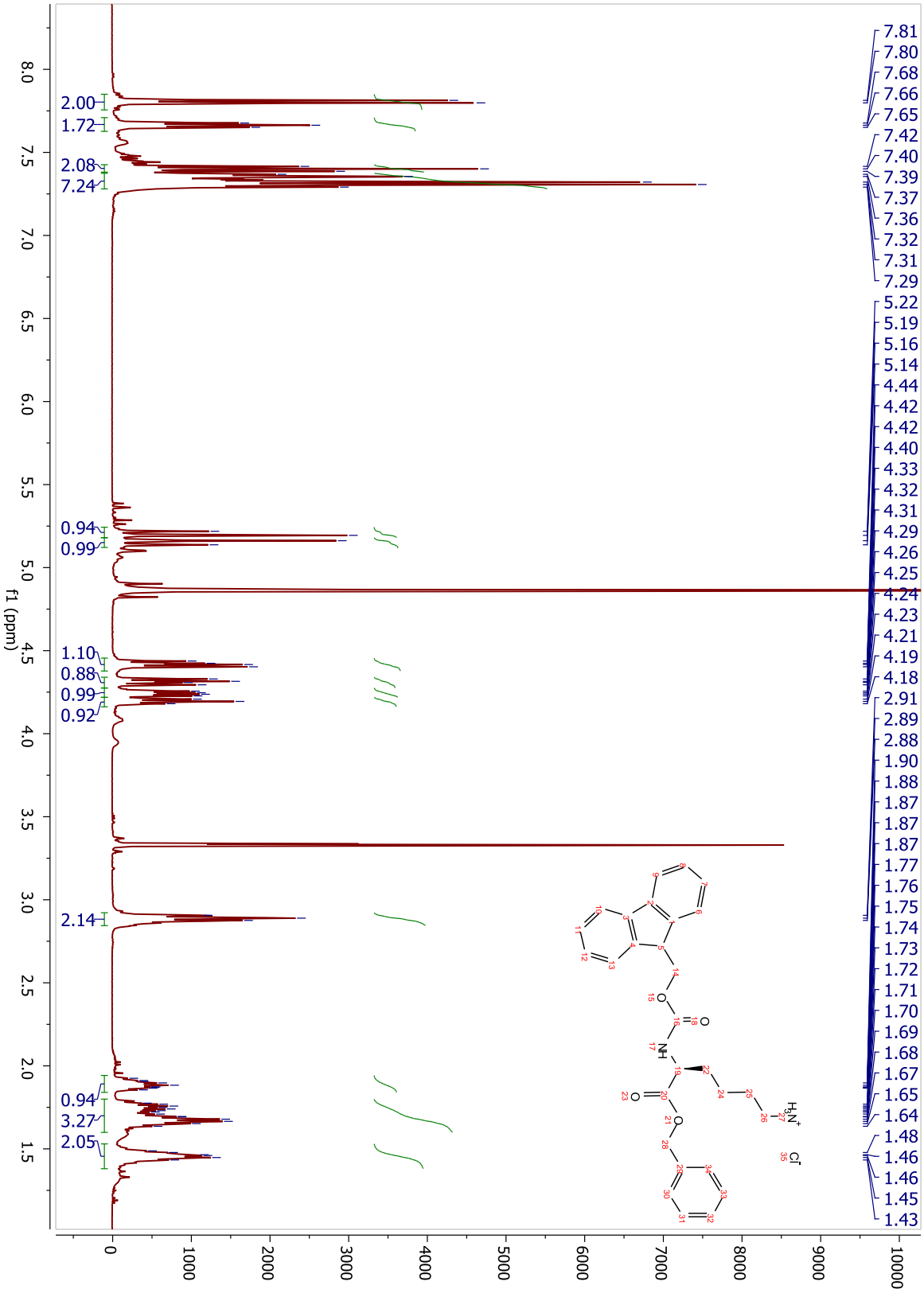
Compound 5.5



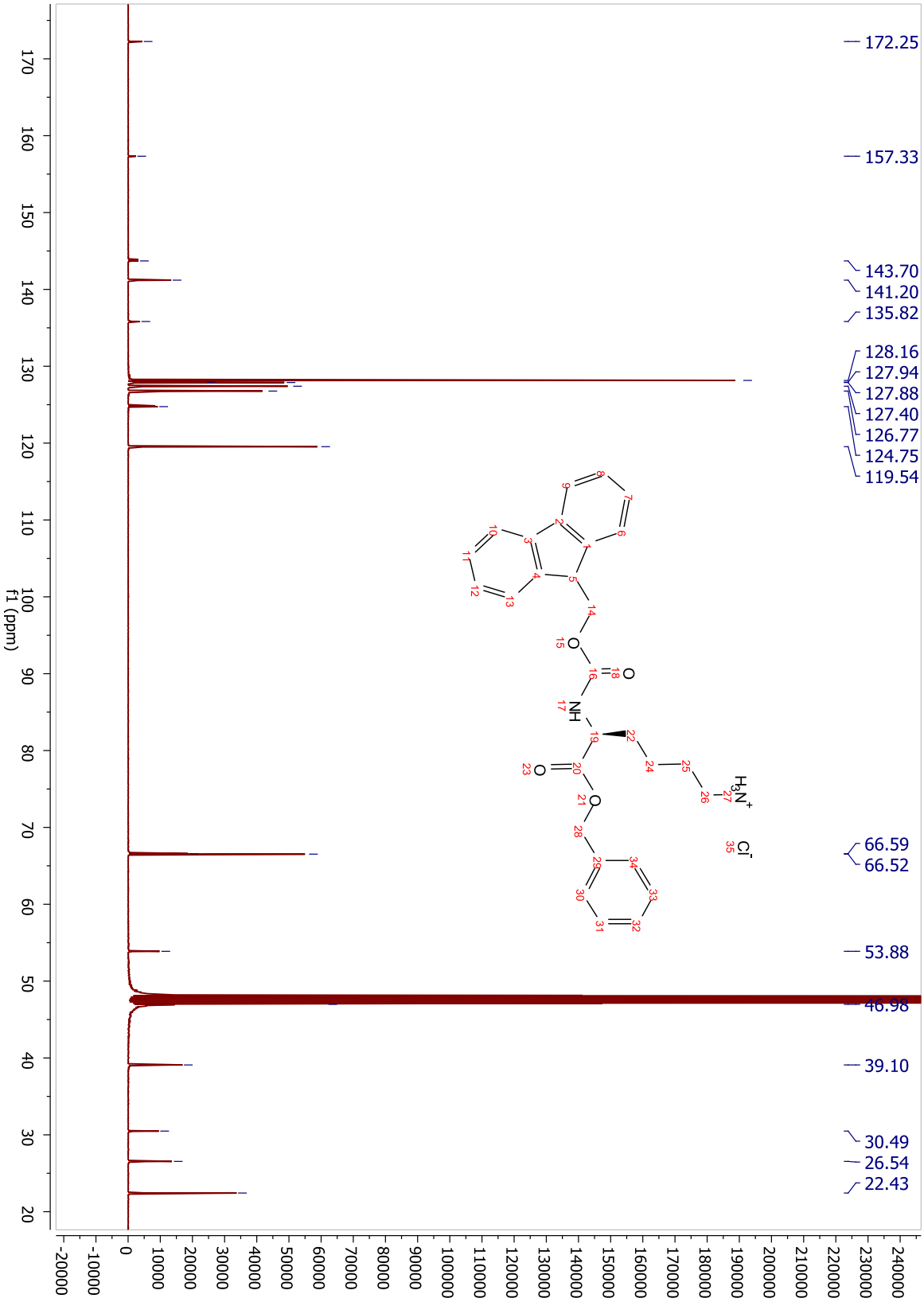
Compound 5.5


172.46
156.19
156.15
143.86
141.40
135.42
128.74
128.61
128.45
127.82
127.19
125.24
120.09
79.22
67.25
67.12
53.94
47.26
40.13
32.14
29.67
28.54
22.44
f1 (ppm)

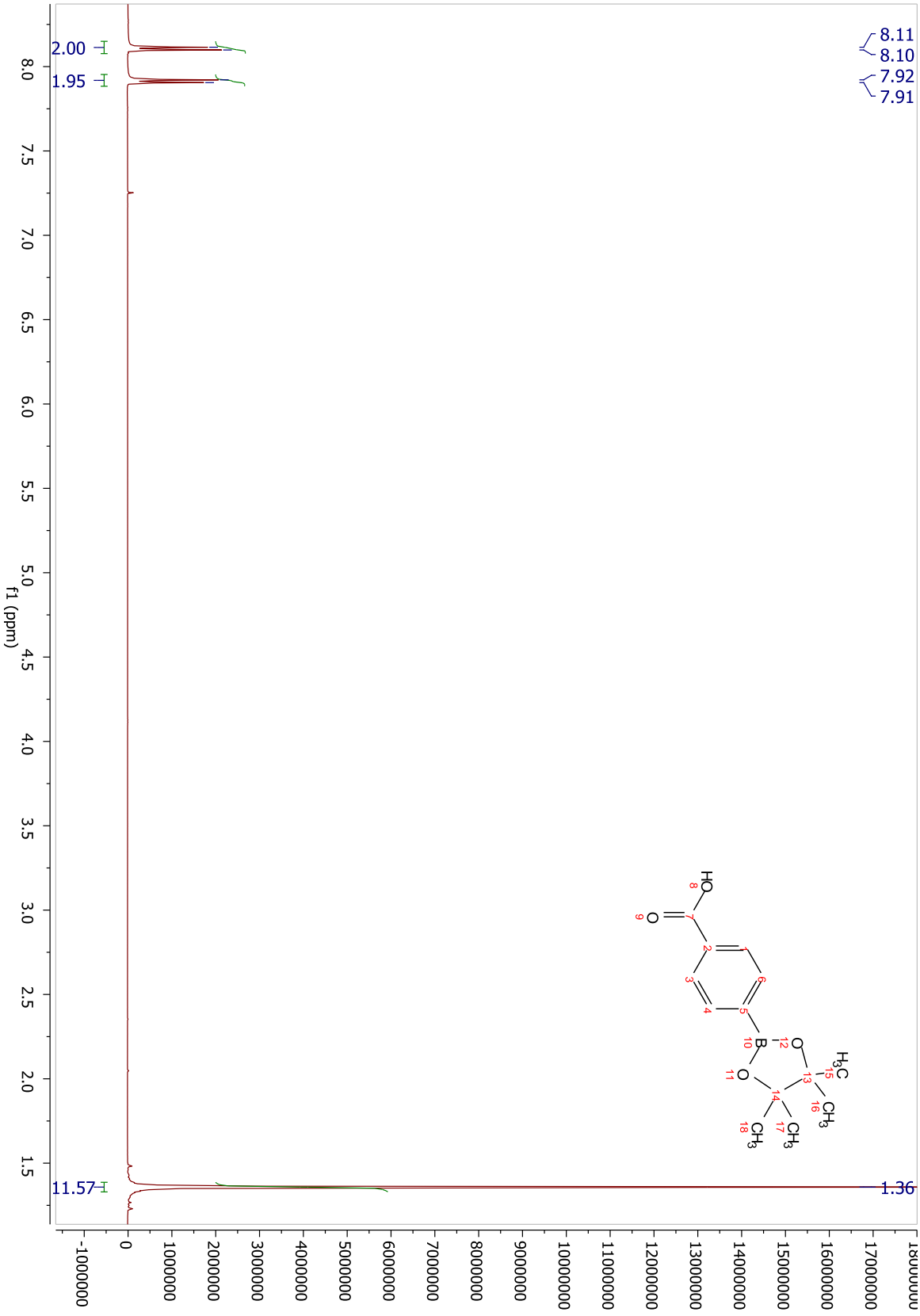
Compound 5.6


f1 (ppm)
2.00
1.72
2.08
7.24
0.94
0.99
1.10
0.88
0.99
0.92
2.14
0.94
3.27
2.05
7.81
7.80
7.68
7.66
7.65
7.42
7.40
7.39
7.37
7.36
7.32
7.31
7.29
5.22
5.19
5.16
5.14
4.44
4.42
4.42
4.40
4.33
4.32
4.31
4.29
4.26
4.25
4.24
4.23
4.21
4.19
4.18
2.91
2.89
2.88
1.90
1.88
1.87
1.87
1.87
1.77
1.76
1.75
1.74
1.73
1.72
1.71
1.70
1.69
1.68
1.67
1.65
1.64
1.48
1.46
1.46
1.45
1.43

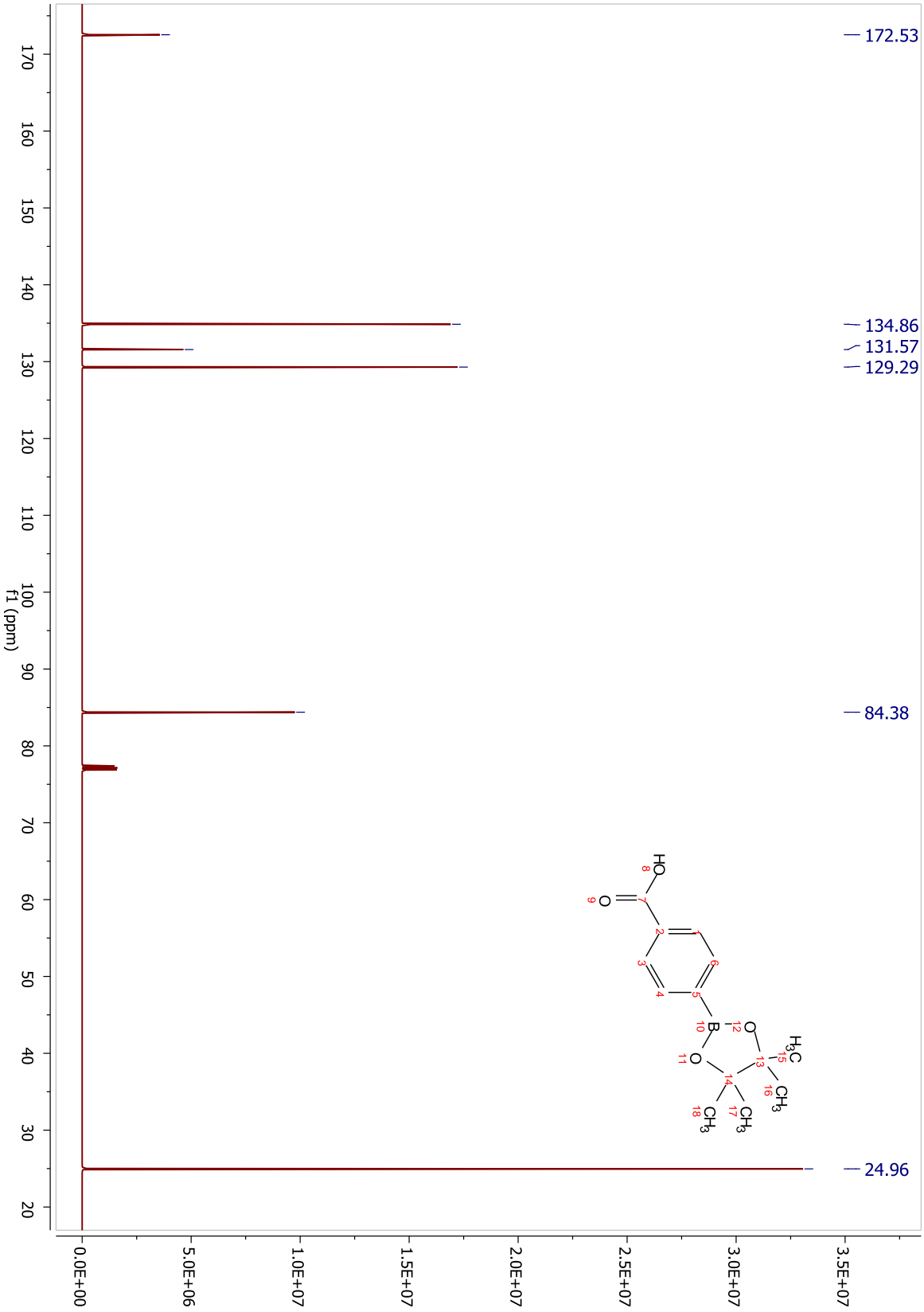
Compound 5.6


172.25
157.33
143.70
141.20
135.82
128.16
127.94
127.88
127.40
126.77
124.75
119.54
66.59
66.52
53.88
46.98
39.10
30.49
26.54
22.43
f1 (ppm)

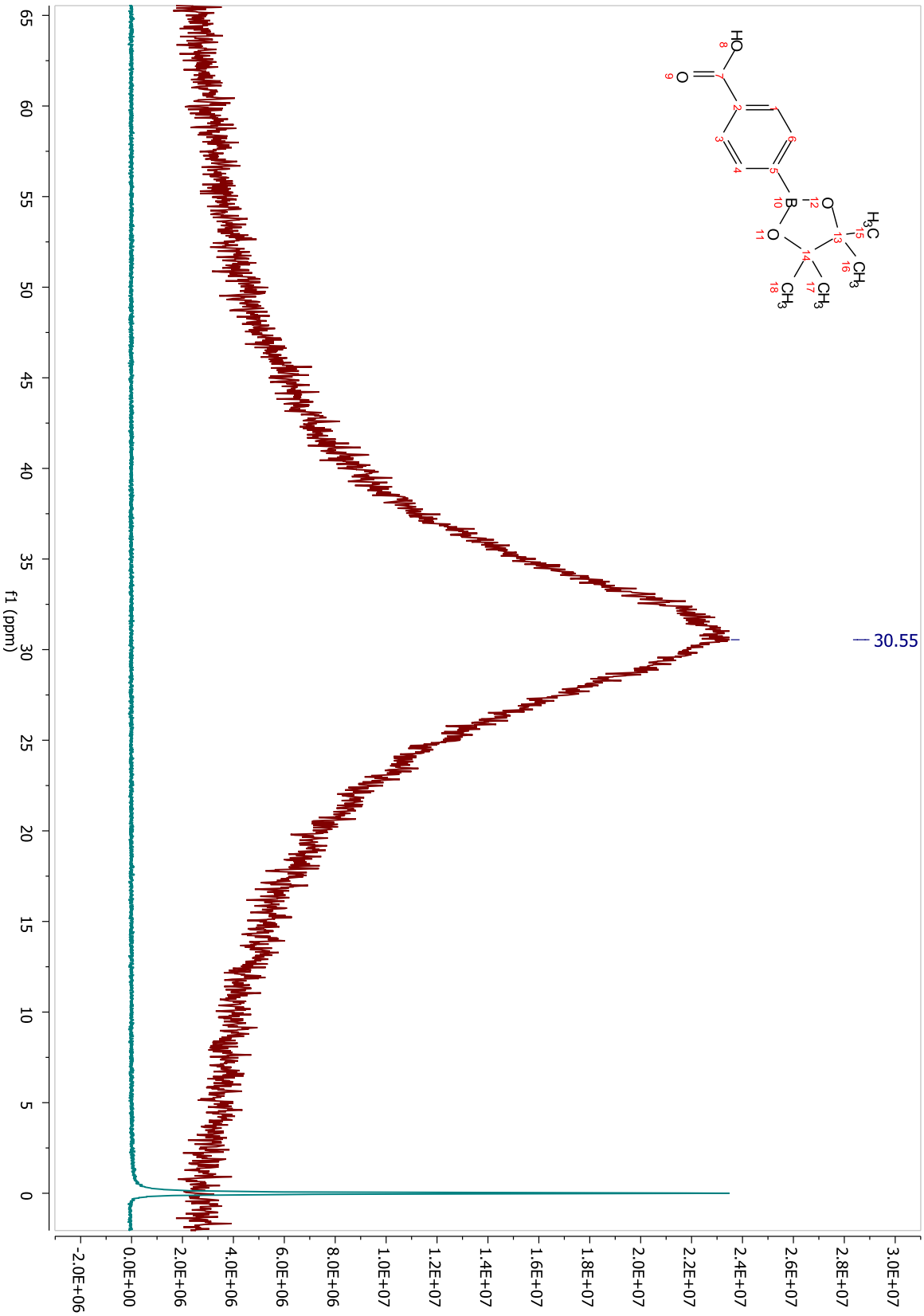
Compound 5.8



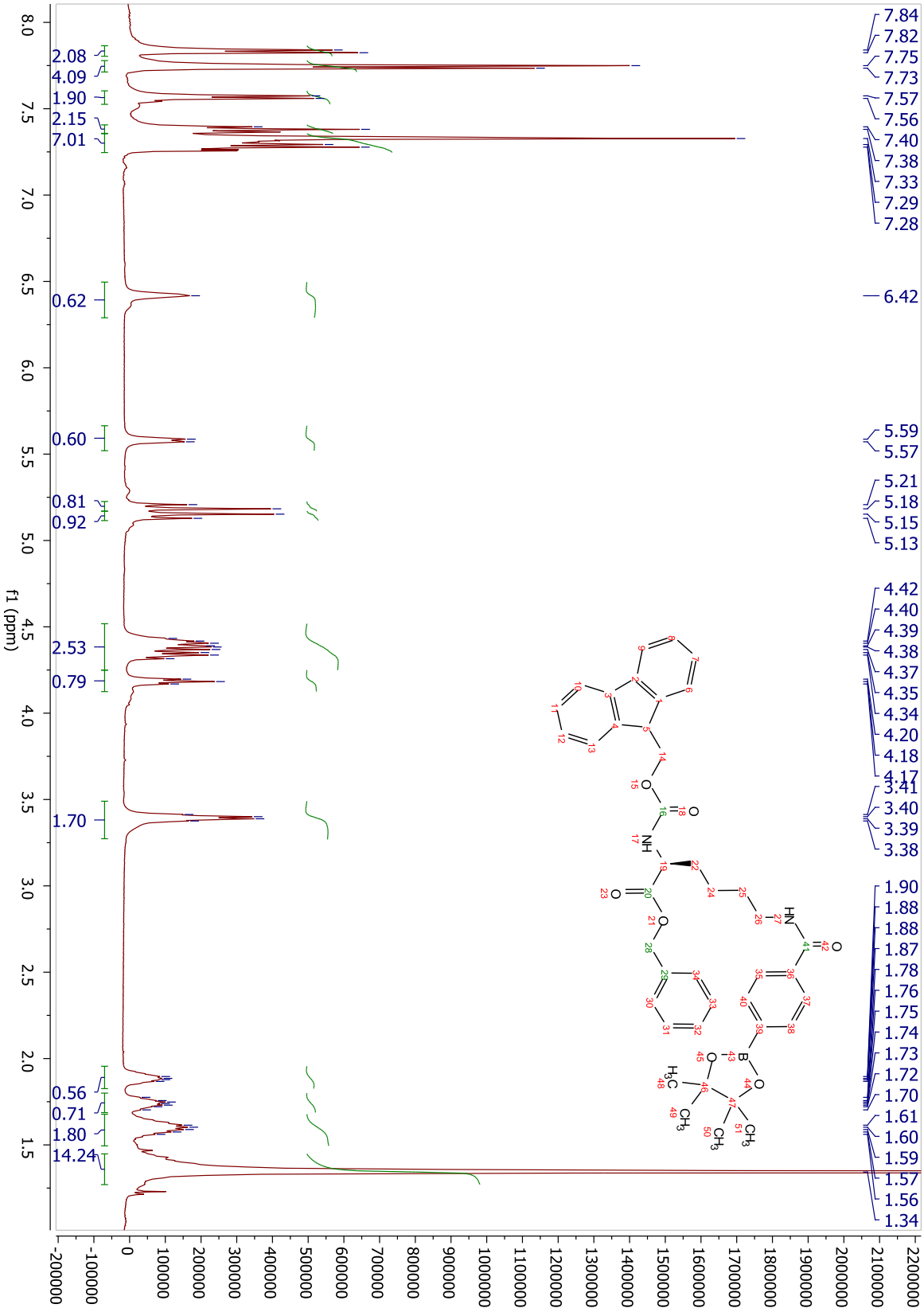
Compound 5.8


172.53
134.86
131.57
129.29
84.38
24.96
f1 (ppm)
HO
O
B
O
O
H3C
CH3
CH3
CH3

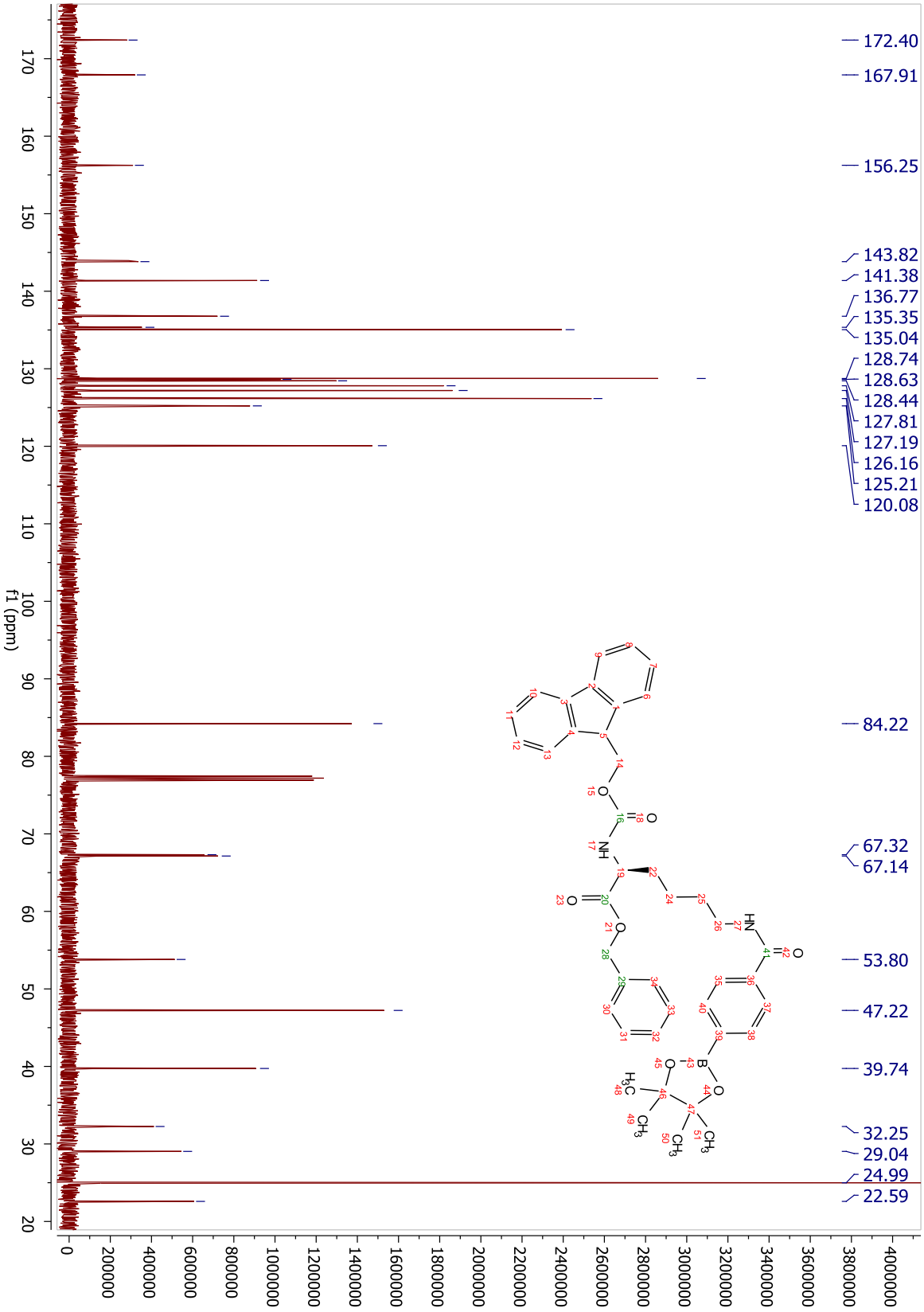
Compound 5.8



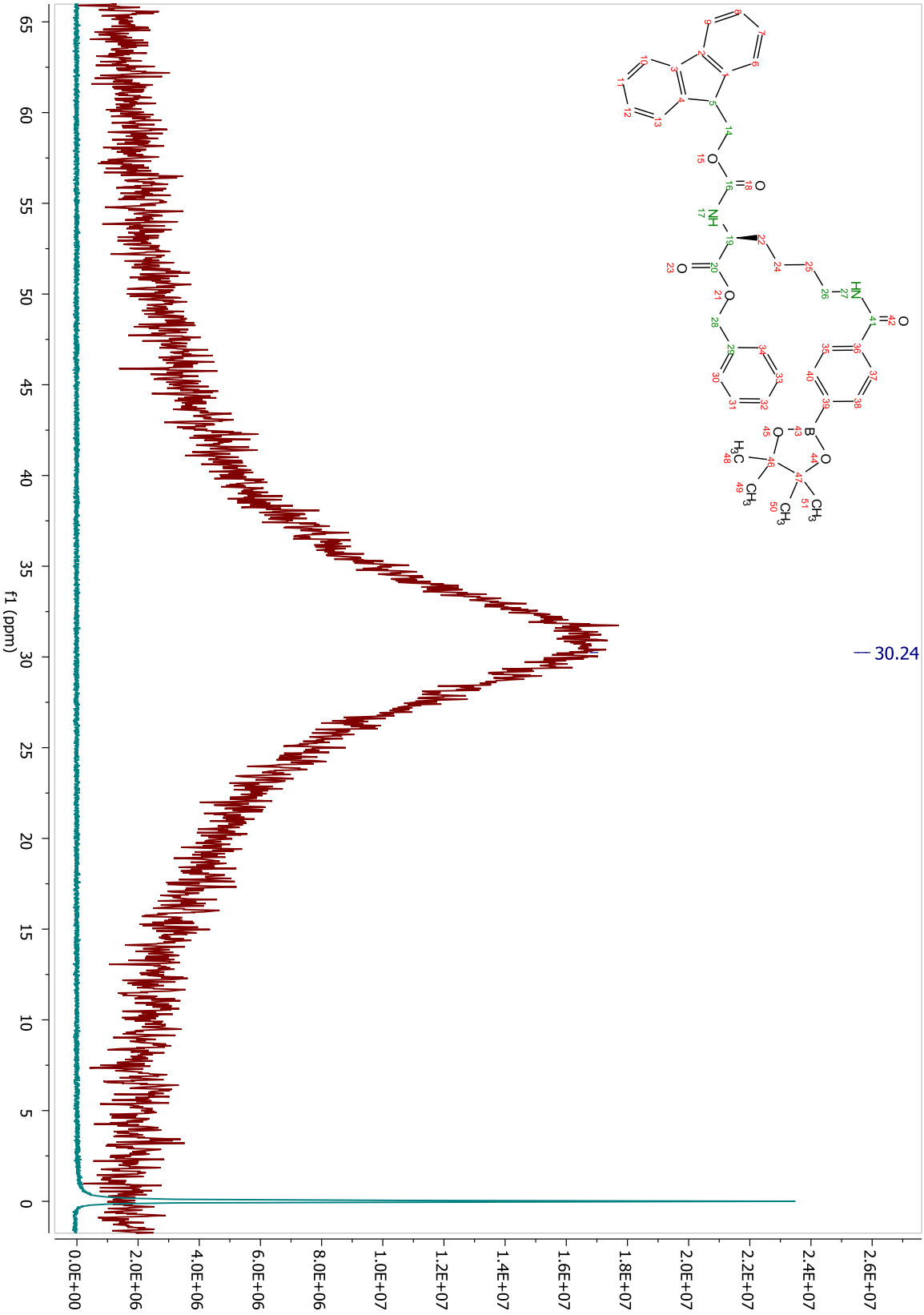
Compound 5.9


f1 (ppm)
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
2.08
4.09
1.90
2.15
7.01
0.62
0.60
0.81
0.92
2.53
0.79
1.70
0.56
0.71
1.80
14.24
7.84
7.82
7.75
7.73
7.57
7.56
7.40
7.38
7.33
7.29
7.28
6.42
5.59
5.57
5.21
5.18
5.15
5.13
4.42
4.40
4.39
4.38
4.37
4.35
4.34
4.20
4.18
4.17
3.41
3.40
3.39
3.38
1.90
1.88
1.88
1.87
1.78
1.76
1.75
1.74
1.73
1.72
1.70
1.61
1.60
1.59
1.57
1.56
1.34
-200000
-100000
0
100000
200000
300000
400000
500000
600000
700000
800000
900000
1000000
1100000
1200000
1300000
1400000
1500000
1600000
1700000
1800000
1900000
2000000
2100000
2200000

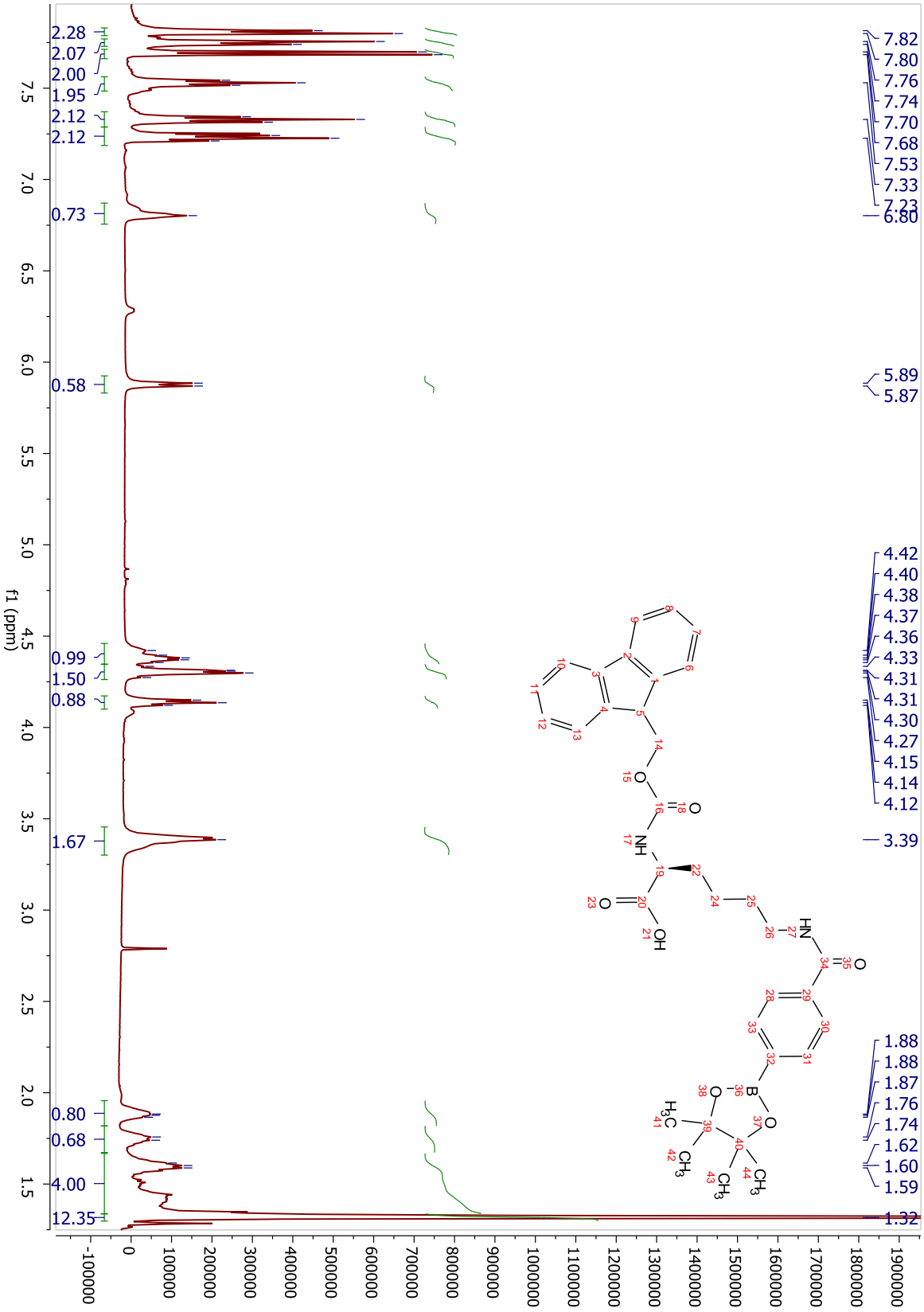
Compound 5.9


172.40
167.91
156.25
143.82
141.38
136.77
135.35
135.04
128.74
128.63
128.44
127.81
127.19
126.16
125.21
120.08
84.22
67.32
67.14
53.80
47.22
39.74
32.25
29.04
24.99
22.59
f1 (ppm)

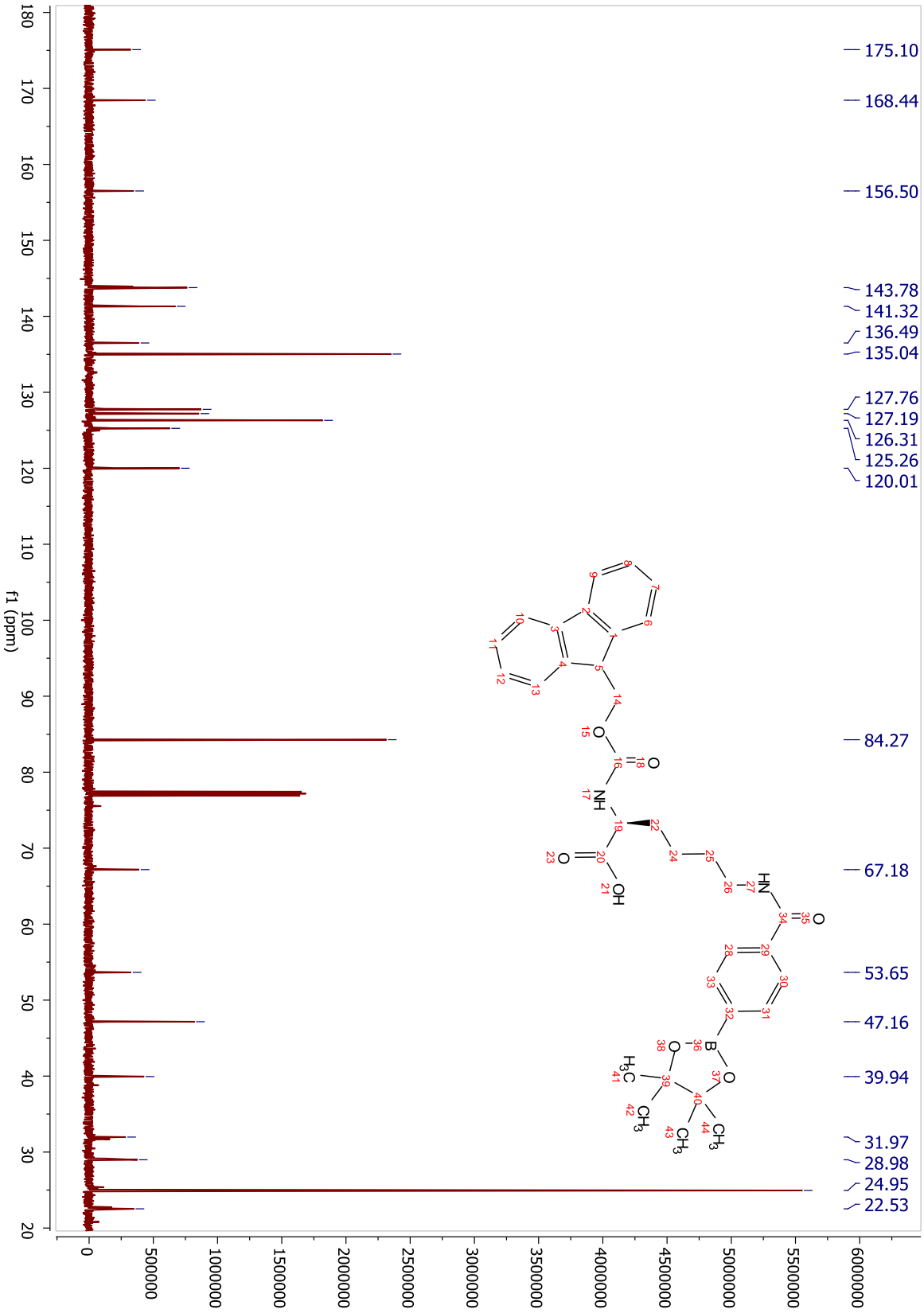
Compound 5.9



Compound 5.10



Compound 5.10



Compound 5.10