Design, Synthesis and Association Study of Universal Readers

for Recognition Tunneling

by

Sovan Biswas

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2016 by the
Graduate Supervisory Committee:

Stuart Lindsay, Co-Chair
Peiming Zhang, Co-Chair
Chad Borges
Ian Gould

ARIZONA STATE UNIVERSITY

May 2016

ABSTRACT

For reading DNA bases more accurately, a series of nitrogen-containing aromatic heterocycles have been designed and synthesized as candidates of universal reader to interact with all naturally occurring DNA nucleobases by hydrogen bonding interaction and eventually is used to read DNA by recognition tunneling. These recognition molecules include 6-mercapto-1H-benzo[d]imidazole-2-carboxamide, 5-(2-mercaptoethyl)-1H-imidazole-2-carboxamide, 5-(2-mercaptoethyl)-4H-1,2,4-traizole-3-carboxamide and 1-(2-mercaptoethyl)-1H-pyrrole-3-carboxamide. Their formation of hydrogen bonding complexes with nucleobases was studied and association constants were measured by proton NMR titration experiments in deuterated chloroform at room temperature. To do so, the mercaptoethyl chain or thiol group of these reading molecules was replaced or protected with the more lipophilic group to increase the solubility of these candidates in $CDCl_3$. The 3' and 5' hydroxyl groups of deoxyadenosine (dA), deoxyguanosine (dG), deoxycytidine (dC) and thymidine (dT) were protected with *tert*-butyldimethylsilyl (TBDMS) to eliminate hydrogen bonding competition from the hydroxyl protons with these candidates as well as to increase the solubility of the nucleosides in $CDCl_3$ for NMR titration experiment. Benzimidazole and imidazole containing readers exhibited the strongest H-bonding affinity towards DNA bases where pyrrole containing reader showed the weakest affinity. In all cases, dG revealed the strongest affinity towards the readers while dA showed the least.

The molecular complex formation in aqueous solution was studied by electrospray ionization mass spectrometry (ESI-MS) and tandem mass spectrometry. The formation of both 1:1 and 2:1 complexes between one or two reading molecules and a

DNA nucleotide were observed by ESI mass. A series of amino acids and carbohydrates were also examined by mass spectrometry to show the formation of non-covalent complexes with imidazole reader in aqueous solution. The experimental results were compared by calculating energies of ground state conformers of individual molecules and their complexes using computer modeling study by DFT calculations. These studies give insights into the molecular interactions that happen in a nanogap during recognition tunneling experiments.

DEDICATION

To My Family

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF ABBREVIATIONS

AcOH	acetic acid
Ac_2O	acetic anhydride
APCI	atmospheric pressure chemical ionization
anh	anhydrous
aq	aqueous
atm	atmosphere
Bn	benzyl
BnBr	benzyl bromide
br s	broad singlet
^{o}C	degrees Celsius
^{13}C NMR	carbon nuclear magnetic resonance spectroscopy
cat	catalytic
$CDCl_3$	deuterated chloroform
cm	centimeter
conc	concentrated
δ	chemical shift (ppm)
d	doublet
dd	doublet of doublet
DCM	dichloromethane
DMAP	dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide

DNA	deoxyribonucleic Acid
EtOAc	ethyl acetate
Et$_3$N	triethylamine
ESI	electrospray ionization
g	gram(s)
h	hour(s)
^{1}H NMR	proton nuclear magnetic resonance spectroscopy
HPLC	high-performance liquid chromatography
Hz	hertz
J	coupling constant
L	liter
m	multiplet
M	molar
M$^+$	molecular ion
MALDI-TOF	matrix assisted laser desorption ionization time of flight
mg	milligram(s)
μm	microgram(s)
MHz	mega Hertz
min	minutes
mL	milliliter
mM	millimolar
mmol	millimole(s)
μmol	micromole(s)

N	normal
nm	nanometer
NMR	nuclear magnetic resonance
ppm	parts per million
q	quartet
R_f	ratio of fronts
RNA	ribonucleic acid
s	singlet
t	triplet
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin Layer Chromatography
UV	ultraviolet
v	volume

CHAPTER 1

INTRODUCTION

1.1. DNA Sequencing

DNA sequencing is a tool to determine the exact order of nucleotides in a DNA strand. More precisely, it is a methodology to determine the order of four DNA bases namely adenine, thymine, guanine and cytosine in a DNA molecule. It is an indispensable tool in rapidly expanding biological and medicinal research fields such as forensic biology,[1,2] biotechnology,[3] virology[4,5] and medical diagnosis[6,7] to find diseases associated with genes and discovery of drugs. DNA sequencing is used to determine sequence of individual genes,[8,9] chromosomes[10] and entire genomes.[11] In 1953, James Watson and Francis Crick suggested the double helix structure of DNA.[12] Soon after the discovery, various types of DNA sequencing methods have been reported. Sanger sequencing is the standard method for DNA sequencing based on two dimensional chromatography and fluorescence-based detection.[13] The high demand for low cost sequencing has compelled the development of next-generation sequencing technologies which can parallelize the sequencing process producing thousands or millions of sequences simultaneously.[14,15] It is intended to lower the cost of DNA sequencing beyond what is possible with standard dye-terminator methods.[16]

1.2. Sanger Sequencing

In the 50's, Dr. Fred Sanger first determined the sequence of amino acids in a protein. He showed that the 51 amino acids of the insulin were arranged in a specific order.[17] Since the genetic code determines the order of amino acids, the DNA sequence is

collinear with the amino acid sequence. However, knowing the amino acid sequence of a protein does not tell us the exact nucleotide sequence of its gene. Because genetic code is redundant and more than one codon can code for an amino acid. In the early 70's, he developed the method to determine the exact sequence of nucleotides in a gene. It is known the famous Sanger method or the chain termination method for DNA sequencing or first generation sequencing.[18] He was awarded the noble prize in chemistry for this invention in 1980. The method is described in the following scheme (Figure 1.1).[19] It utilizes a high fidelity DNA-dependent polymerase to generate a complimentary copy to a single stranded DNA template.[20,21]

To sequence a DNA, a double stranded DNA is denatured to a single stranded DNA by heating. The DNA splits into a template strand (blue in step 1) and a complementary strand. A primer (yellow) is annealed to the template strand to add external nucleotides to the strand later. The template stranded DNA with attached primer was added into four reaction vials following the addition of DNA polymerase and dNTP's. One of these dNTP's is usually labelled with ^{31}P or ^{35}S atom to determine the DNA sequence later. Modified nucleotides (ddNTP's) are added to the reaction mixtures. Only one type of ddNTP is added to each reaction mixture (step 2). Chain termination occurs when a ddNTP is added to the growing strand because of the lack of 3'-OH group in ddNTP. The DNA strands are now separated based on their size by gel electrophoresis on acrylamide gel. A sample of each of the four reaction mixtures is added into separate lane on the gel. It can separate polynucleotide chains in size by 1 nucleotide. A current is set up across the gel and negative bias is applied to the added nucleotide end. As the DNA are negatively charged, it starts to move towards the positive electrode. Smaller fragments

travel faster than larger fragments. After it is developed, an autoradiograph is taken (step

3). Because one of the dNTPs is radiolabeled, the radiograph will show bands of DNA

strands. The bands are used to find out the DNA sequence. Each reaction mixture has the

same primer, therefore all strands begin with the same sequence. Each chain in a

particular flask ends with the corresponding ddNTP present in the flask. For example if a
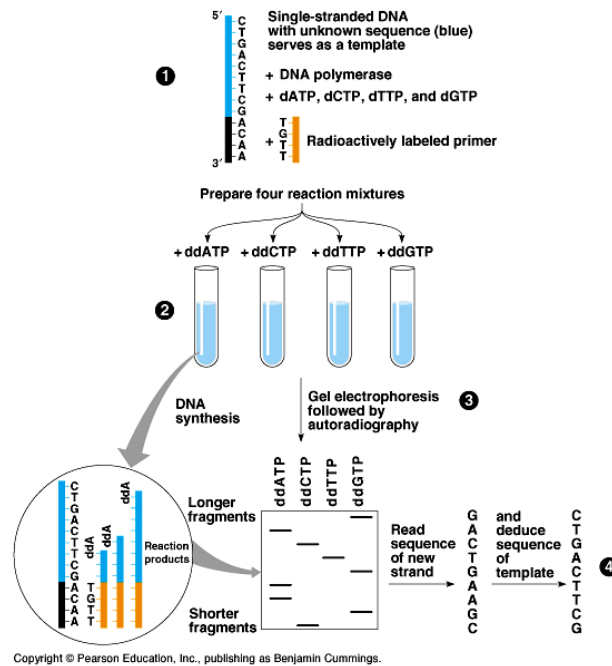
Figure 1.1. Schematic of DNA sequencing by Sanger method

DNA strand is taken from a ddATP flask, it will end with adenosine nucleotide. This

information is used to sequence the complete DNA by reading the various bands in the

radiograph (step 4). Sanger method is considered the "gold standard" for sequencing.

Modern day Sanger sequencing instrument is capillary based automated electrophoresis

which typically analyzes 8 to 96 sequencing reactions simultaneously.

1.3. Next Generation Sequencing

Over the past decade, DNA sequencing has rapidly transitioned from the Sanger method to next generation sequencing (NGS). NGS has become an indispensable tool for genomics, making great strides toward diagnosis of diseases in clinics.[22] NGS sequences in a massively parallel manner,[23] dominated by technologies that use polymerases to synthesize complementary strands along DNA targets so that DNA sequences can sequentially read out by optically imaging fluorescent signals resulting from incorporation of dye labeled nucleotides[24] or electronically measuring released protons during incorporation of natural triphosphates.[25] Since its advent, NGS has reduced the sequencing cost from about US $10 million to thousands of dollars.[26] With a state-of-the-art NGS machine, an individual human genome can be finished in a few days. Compared to Sanger sequencing (> 800 Q20 read length), [27] however, NGS has shorter read length (~150 bases for single end, www.illumina.com; or 200 bases, www.lifetechnologies.com) and lower raw sequencing accuracy.[23,28] These shortcomings present challenges for use of NGS. First, NGS requires much higher sequencing coverage than the Sanger method for de novo assembly of genomes with a comparable quality.[29] It generates sequencing data in a rate of 100 gigabases (Gb) per single genome for moderate coverage (~30-fold). The deluge of sequencing data requires a computing cluster or supercomputer for their analysis.[30] Secondly, short reads couldn't encapsulate long blocks of repetitive sequences, resulting in fragmented assemblies for repetitive sequences longer than the read length. Given the fact that nearly half of the human genome is filled with repeats (ranging in size from 1 - 2 bases to millions of bases),[31] a new platform is needed to address the repeat issue.

Current progress in nanopore sequencing has opened a new avenue to develop the sequencing technologies. A nanopore is an orifice with diameters comparable to that of a DNA double helix (ca. 2 nanometers) have been investigated as a physical means for the sequencing.[32,33] It should be able to read very long segments and simplyfy sample preparation elimination the use of costly biochemical reagents, such as polymerases or ligases. A nanopore can function as a fluidic channel to conduct ions under a volatage bias. When it is embedded in a thin membrane that separates two chambers filled with conductive electrolytes, DNA molecules can electrophoretically translocate through the nanopore, the ionic currents would transiently be reduced because the flow of ions is blocked by DNA.[34] This is a mechanism used by a commercial product MinION for sequencing DNA by protein nanopores (www.nanoporetech.com). Since there is no theoretic limit on length of the DNA translocation, the nanopore DNA sequencing will have potential to solve the assembly issues related to the short reads of NGS, providing a high speed and low cost process of sequenceing. However, the protein nanopore sequencing suffers from low accuracy (85%).[35] Gundlach and coworkers have demonstrated that the current blockade in a protein nanopore (Mycobacterium smegmatis porin A, referred to as MspA) is a collected event of four nucleotides (quadromer),[36] and the 256 possible quadromers produce a significant number of redundant current levels.[37] Despite many years of efforts, the nanopore has not achieved a single base resolution in DNA sequenicng. Branton *et al* pointed out that "even an infinitely short channel would not achieve the required resolution" and alternative readout methods are required for the nanopore DNA sequencing.[38]

1.4. Sequencing by Recognition Tunneling

An alternative method to the measurement of ion current blockades[39] is proposed by measuring the electron tunneling current across a translocating DNA molecule through a nanogap.[40] We have used a Scanning Tunneling Microscope (STM) where the two electrodes (gold or palladium) are functionalized by a reader molecule. The distance between these two electrodes are in nanometers (2-3 nm). When a DNA base passes through the electrodes, it is captured by the reader molecules due to the formation of hydrogen bonding complexes (Figure 1.2) and generates electronic signals as spikes.

When benzamide reader is used for recognition of DNA bases, it is found that it can read all DNA bases, A, C, G except T.[41] In addition, there is significant overlaps of the electric signal peaks. In order to be able to read all DNA bases accurately we designed universal readers following a set of guidelines. The molecule should contain multiple hydrogen bonding sites with multiple donors and acceptors sites. It should have
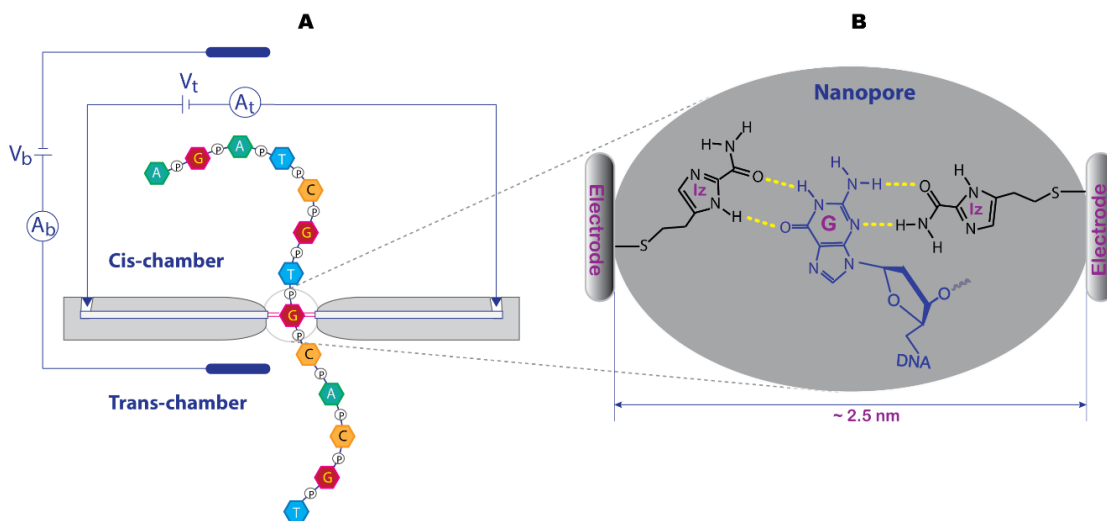
Figure 1.2. Cartoon illustrating **(A)** A tunneling device embedded in a nanopore to read DNA bases when they sequentially translocate through a nanopore; **(B)** Recognition interactions in the nano-gap where read molecules (universal reader) attached to the electrodes catch a DNA base by forming a hydrogen bonding complex to cause electronic spikes.

high conductance and small size to fit into a nanopore without impeding the DNA translocation. It should have variable conformations in order to catch all DNA bases. It has to be electrochemically stable under physiological conditions. We selected five membered aromatic nitrogen heterocycles because of their π-excessive nature (6 π electron shared by 5 atoms).[42] Based on these principles we have designed four universal readers ICA, BCA, TCA and PCA which can read all four naturally occurring nucleobases by recognition tunneling.

1.5. Determining Association Constants from Titration Experiment

A common approach to study supramolecular interaction is a titration of the guest to a solution of the host, and measuring changes of some physical property through NMR,[43,44] UV-Vis,[45,46] fluorescence[47,48] and other techniques such as isothermal calorimetry[49,50] etc. Despite the simplicity of the techniques, there are important concerns that need to be taken care of to obtain reliable results. Choice of method (e.g. NMR or UV-Vis) and choice of stoichiometric binding model are very important depending on the type of systems and environments are used. Another aspect is to select a suitable concentration range of host and guest before experiment. Finally a non-linear regression method[51,52] is preferred over a linear regression method for better result. In a NMR titration method, a solution of guest molecule is gradually added to a solution of host molecule while the change in chemical shift of certain nucleus is measured. The resulting information is fitted to a suitable binding model and association constant K_a is obtained from the equation and fitting curve. It is then compared against literature results available from a different measurement.

1.6. HypNMR Program for Titration Curve Fitting

In our present study we have used HypNMR 2008 program from hyperquad protonic software[53] to measure association constants from average chemical shifts for 1:1 and 1:2 complex formation between one host and one or two guest molecules respectively. It is assumed that the equilibrium is attained very rapidly (fast-exchange) on the NMR time scale so that the chemical shift for a given nucleus is a mole fraction-weighted average over all of the chemical species in which the nucleus is present. Data input is comprised of the chemical shifts of the NMR peaks in relation to the analytical concentrations of reagents in the solution and sometimes it pH. The program can handle data in which the chemical shift(s) of one or more nuclei may be assigned in some spectra and not in others. It is used to determine the stability constant of host-guest reaction, dimerization reaction,[54] natural polyprotic bases in solution,[55] fluorinated polyamines from ^{13}C NMR data[56] etc. The program was originally written for the DOS operating system as a set of stand-alone FORTRAN programs and enhanced later to run on Windows operating systems. The stability constant refinement part is written using FORTRAN program and compiled with 32-bit compiler for additional precision. It can be used on multi reagent system with many stability constant involved. There is no restriction on the number of resonant nuclei being accounted for and the number of data points being used. Finally it generates cumulative stability constants in terms of common logarithm and from which stepwise stability constants are derived.

1.7. Cumulative and Stepwise Stability Constants

The stability constant (K_a) for an association reaction $H + G = HG$ is defined by the

equation [HG] = K_a[H][G] where H is a host molecule and G is a guest molecule and [H], [G], [HG] are the concentrations of host, guest and host-guest complex respectively in equilibrium. Cumulative stability constants also known as overall stability constants refer to the formation of a species from the reagents. For example, the cumulative formation constant for HG_2, β_2, is defined by $[HG_2] = \beta_2[H][G]^2$ whereas the stepwise constant, K_2, for the formation of HG_2 from HG and G is defined by $[HG_2] = K_2[HG][G]$.

1.8. NMR Titration Equation

In an NMR titration experiment the host concentration is usually kept constant while the guest concentration is varied from zero to fairly high. Host is usually more expensive (or more synthetically challenging) compared to guest. For a simple 1:1 equilibrium system between H and G (Figure 1.3, a), the equilibrium constant for the formation of HG complex can be written as $K_a = [HG]/[H][G]$.
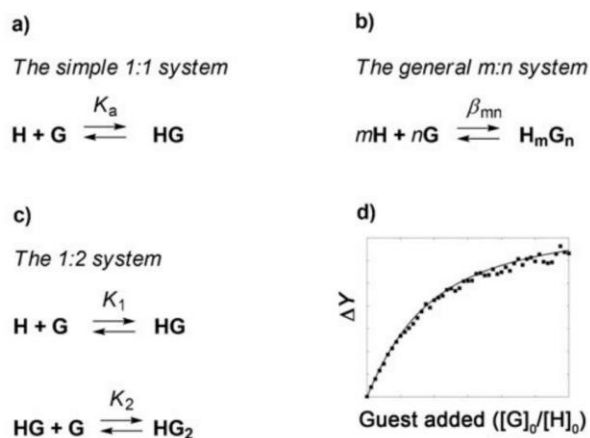
Figure 1.3. (a)–(c) Three typical equilibrium for 1:1, 1:2 and m:n systems: **H** = host, **G** = Guest and $\mathbf{H_xG_x}$ = host–guest complex of interest, K_x = the thermodynamic association constant for a particular interest and β_{mn} = overall association constant for an m : n host–guest complex formation. **(d)** A typical binding isotherm.

9

Similarly, for a 1:2 equilibrium system between one H and two G molecules (Figure 1.3, c), the stepwise equilibrium constants for the formation of [HG] and [HG$_2$] complexes can be written in terms of K_1 and K_2.

$$K_1 = \frac{[HG]}{[H][G]} \qquad \beta_{mn} = \frac{[H_m G_n]}{[H]^m [G]^n}$$

$$K_2 = \frac{[HG_2]}{[H][HG]}$$

In general in case of the formation of a supramolecular complex involving m Host and n Guest molecules, the overall stability constant can be written as β_{mn} above.

Now the equation for the 1:1 equilibrium can be rewritten in terms of concentration of

$$K_a = \frac{[HG]}{([H]_0 - [HG])([G]_0 - [HG])}$$

$$K_a = \frac{[HG]}{[H]_0[G]_0 - [HG][G]_0 - [HG][H]_0 + [HG]^2}$$

$$K_a([H]_0[G]_0 - [HG][G]_0 - [HG][H]_0 + [HG]^2) = [HG]$$

$$K_a[HG]^2 - (K_a[G]_0 - K_a[H]_0 + 1)[HG] + K_a[H]_0[G]_0 = 0$$

thus

$$[HG] = \frac{1}{2}\left(G_0 + H_0 + \frac{1}{K_a}\right) - \sqrt{\left(G_0 + H_0 + \frac{1}{K_a}\right)^2 + 4[H_0][G_0]}$$

Figure 1.4. Expression of [HG] as a function of [H]$_0$, [G]$_0$ and Ka

[HG] where [G]$_0$ and [H]$_0$ are the total concentrations of guest and host respectively according to the equations, [H]$_0$ = [H] + [HG] and [G]$_0$ = [G] + [HG] following the mass balance equation (Figure 1.4)

Now, if n_c is mole fraction of complex [HG], then n_c = [HG]/[H]$_0$

If, δ_h and δc are the chemical shifts of H and HG, then at fast exchange condition,

$\delta = n_h\delta_h + n_c\delta_c = n_c\delta_c + (1-n_c)\delta_h = (\delta_c- \delta_h)n_c + \delta_h$; where δ is the population average

chemical shift or observed chemical shift (δ_{obs})

Now rewriting the above equation we get, $n_c = (\delta - \delta_h)/(\delta_c - \delta_h)$ and replacing n_c by [HG]

we get, [HG]/[H]$_0$ = $(\delta - \delta_h)/(\delta_c - \delta_h)$

Now if $\delta - \delta_h$ = y and $\delta_c - \delta_h$ = n; where "n" is a number, not a mole fraction; [H]$_0$ = A,

[G]$_0$ = x and K_a = K;

Then, [HG]/[H]$_0$ = y/n, or [HG] = y[H]$_0$/n = yA/n,

Now putting this in the above final equation (Figure 1.4) we get

$$y = [0.5*(A+x+1/K)-(sqrt(0.25*(A+x+1/K)^2-A*x))]*n/A$$

This equation can be solved using origin curve fitting. Using this method the binding

constant of a 1:1 equilibrium can be solved. But solving a 1:2 equilibrium requires further

analysis of a very complex equation (not shown here). We need a program like HypNMR

which can handle multi equilibrium equation and can calculate association constant.

1.9. Electrospray Ionization Mass Spectrometry (ESI-MS)

Electrospray ionization (ESI) is an analytical technique used in mass spectrometry

that can provide information about molecular mass and structure of analyte molecules

after their conversion to ions. The analyte molecules are introduced into ionization source

by electrospray in which high voltage is applied to liquid solution to create positive and

negative charged ions. The ions travel through the different parts of the mass

spectrometer and being analyzed by quadrupole analyzer according to their mass/charge

(m/z) ratio. A typical mass spectrum show the relative abundance of the signals according to their m/z ratio.[57] The basic components of the ESI-mass spectrometer is shown (Figure 1.5).[58] It has been extensively used in clinical biochemistry,[59] qualitative and quantitative analysis of antibiotics in pharmaceutical formulation,[60] and drug discovery.[61] It has emerged as a powerful tool in the life science to determine the identity,[62,63] quantity,[64,65] and structural properties[66,67] of the protein molecules.
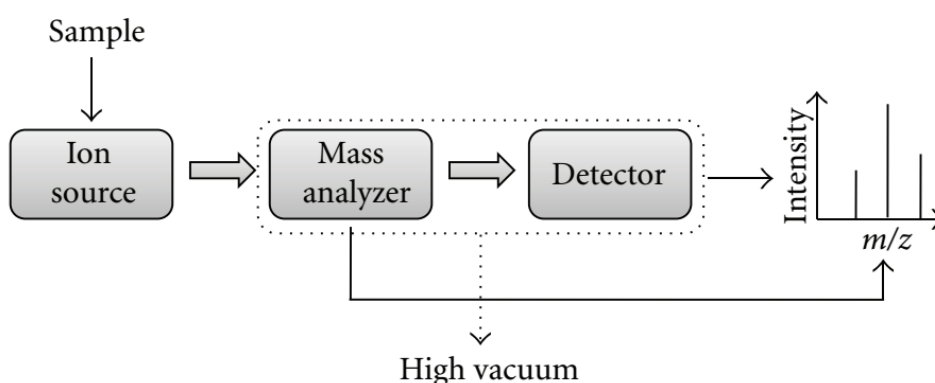
Figure 1.5. The basic components of the ESI-mass spectrometer

The electrospray ionization technique was first reported by Masamichi Yamashita and John Bennett Fenn in 1984.[68] John Fenn was awarded the Noble Prize in Chemistry in 2002 for the development of electrospray ionization for the analysis of biological macromolecules.[69] The following is a diagram of a second-generation ESI-MS apparatus from Dr. Fenn's lab.[70] It shows a liquid sample is injected through a needle spray syringe into the instrument, desolvation process occurs, charged droplets evaporates, then goes to the gas phase and being analyzed by quadrupole analyzer (Figure 1.6). ESI is a soft ionization technique and proved to retain the identity of a non-covalent complex in the gas phase after ionization (without fragmentation). This technique has been extremely

12

helpful and universally applied to detect and identify molecular complexes. However a single mass spectrum does not provide any structural information about the molecular complex being analyzed. To overcome this problem, ESI is coupled with a tandem mass spectrometry (ESIMS/MS).
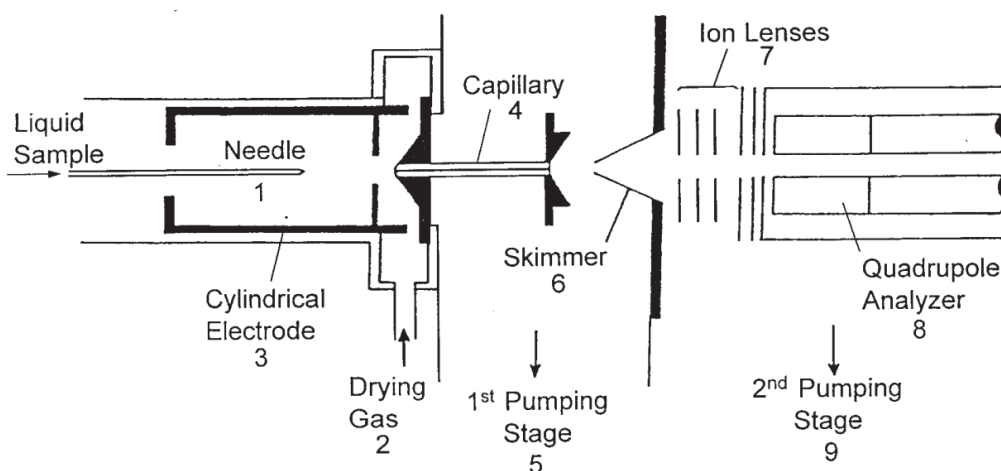
Figure 1.6. Diagram of a Second-generation ESI-MS Apparatus from Fenn's Lab (Late 1980's) showing ion desolvation process. Small, charged droplets produced by the electrospray evaporate, generating a high electric field at the droplet surface. Analyte molecules that were dissolved in the droplet can attach to charges and be lifted into the gas phase by this field

1.10. Ionization Mechanism in Electrospray Mass Spectrometry

Sample solution is prepared by mixing the analyte(s) of interest with a solvent. The choice of solvent is generally a mixture of water and methanol[71] or acetonitrile. The sample solution is infused into the ES chamber through a stainless steel hypodermic needle spray syringe with a flow rate between 1 and 20 µL/min. In some cases, small amount of acetic acid is added into the analyte solution to increase the concentration of protons that facilitate the ionization process. A voltage is applied to the needle tip. This potential at the needle tip charges the surface of the emerging liquid, dispersing it by Coulomb forces into a fine spray of charged droplets. Due to the presence of the electric

13

field, the charged droplets migrate toward the inlet end of the glass capillary at the end

wall of the chamber (see label 4 in figure 1.6). There is a continuous flow of heated inert

gas or bath gas of nitrogen or carbon dioxide in addition to the high temperature of the

ESI source.[72] This drying gas evaporates solvent from each droplet and decreases its size.

During this process, the charge density on a droplet surface increases until the Rayleigh

limit is reached at which the Coulomb repulsion becomes of the same order as the surface

tension. The resulting instability called "Coulomb explosion" causes to tear apart the

original droplet producing many smaller, more stable charged droplets known as daughter

droplets or progeny droplet (Figure 1.7).[58] The new droplets undergo desolvation and

Coulomb fission again. This sequence repeats until the radius of curvature of a daughter

droplet becomes small enough that the field due to the surface charge density is strong

enough to desorb ions from the droplet into the "quasi-molecular" ions which are

required for mass analysis. Final production of gas-phase ions can be explained by two

major theories: the ion evaporation model[73,74] and the charge residue model.[75] Some of

these ions that enters the glass capillary, emerges as a supersonic free jet in the first two

vacuum chambers. A major portion of this free jet passes through a skimmer into a

second vacuum chamber and delivers ions to the analyzer with a quadrupole mass filter.

Any solvent molecule or uncharged molecule are removed from the capillary inlet by the

continuous flow of bath gas. Depending on the objective of the experiment and species

being analyzed, suitable solvent, dry gas temperature and flow rate of bath gas is

designed. The ions observed by mass spectrometry are protonated molecular form

$[M+H]^+$ and/or single or multiply sodium ions $[M+nNa-(n-1)H]^+$ form. Multiply charged

ions such as $[M+nH]^{n+}$ are often observed for large macromolecules.
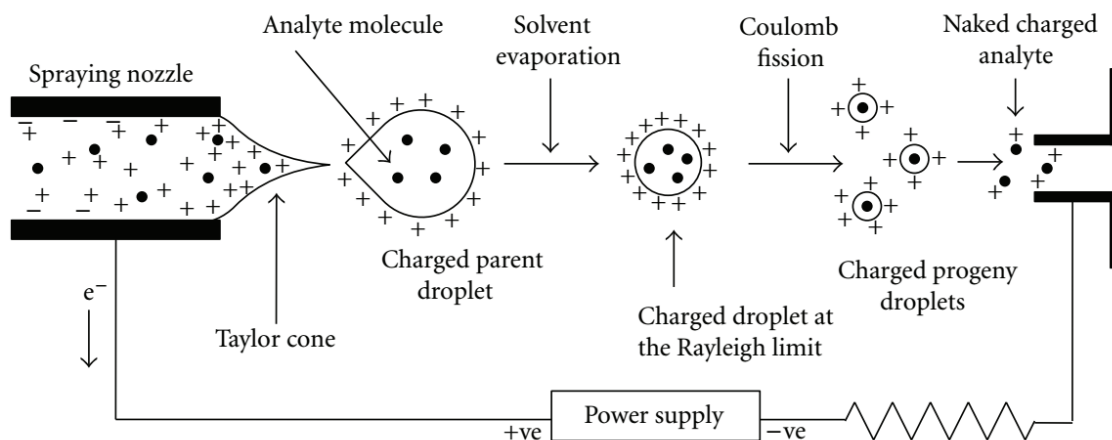
Figure 1.7. Schematic representation of the electrospray ionization

1.11. Tandem Mass Spectrometry (MS2): Collision-Induced Dissociation

Tandem means arrangement of two or more objects one behind another.[58] Tandem

mass spectrometry also known as MS/MS or MS2 is a method where the parent ions are
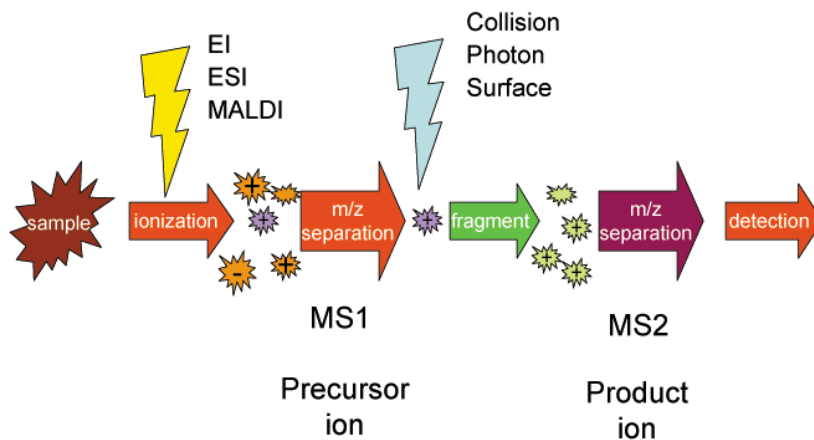
Figure 1.8. Schematic of a tandem mass spectrometry

subjected to two or more sequential stages of mass analysis according to their m/z ratio.

In the first stage of mass spectrometry (MS1) ions are formed in the ion source and

separated according to m/z ratio. In the second stage (MS2), a parent ion of interest is

15

selected and fragmented by collision-induced dissociation to generate product ions which

are detected by a second detector in the mass spectrometry (Figure 1.8).[76]

CHAPTER 2

DESIGN AND SYNTHESIS OF UNIVERSAL READERS FOR RECOGNITION OF

DNA BASES TRHOUGH ELECTRON TUNNELING

2.1. Design of Universal Reader Candidates for Hydrogen Bonding interactions with DNA Nucleobases

We intially used benzamide (**Ba,** Figure 2.1) for recognition tunneling (RT) because a majority of hydrogen bonding motifs existing in DNA bases are a form of donor and acceptor alternation so that the amide group is a good moiety for the recogntion interactions. Our data show that the benzamide moiety reads DNA base A, C, G, and methylated C but T.[77] This could be because the molecule lacks flexibility and suffcient hydrogen bonding sites to interact with the DNA base. Therefore, we designed and synthesized 4(5)-2-mercaptoethyl-1*H*-imidazole-2-carboxamide (**ICA,** Figure 2.1) for RT, which bears mutiple hydrogen bonding sites and a flexible linker. As metioned above, the imidazole-2-carboxamide functions as a universal reader to interact with DNA bases. It provides us a framework to explore new structures for RT. As illustrated in Figure 2.1. Three candidates were derived from tuning the imidazole ring. First one is 1-(2-Mercaptoethyl)-1*H*-pyrrole-3-carboxamide (**PCA**), the pyrrole ring of which has higher π electron densities on the aromatic carbons than the imidazole ring, second one is 5-mercapto-1*H*-benzo[*d*]imidazole-2-carboxamide (**BCA**) that extends the π system of the imidazole ring and is more rigid, and third one is 3-(2-mercaptoethyl)-1*H*-1,2,4-triazole-5-carboxamide (**TCA**) that has one more hydrogen bonding sites than imidazole. The studies on these molecules should give us more insights into effects of chemical structures on recognition of DNA bases. It should be noted that each of these molecules is

connected with a thiol function either through a two-carbon chain or its equivalent in length (see **BCA** in Figure 2.1 where the thiol is placed at a position of two carbon-carbon bonds away from the imidazole ring as drawn in red) for their attachment to electrodes.
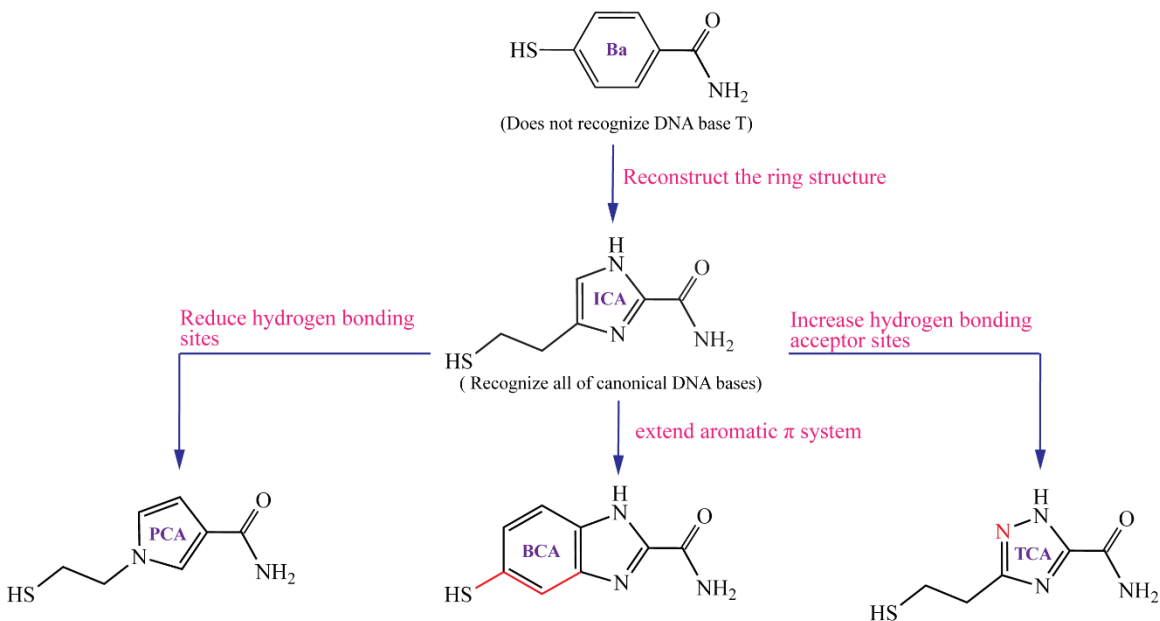
Figure 2.1. Universal reader candidates derived from tuning the imidazole-2-carboxamide molecule

2.2. Hydrogen Bonding Properties of Universal Reader Candidates

In our design, we render these molecules flexible in their conformations to facilitate their interactions with DNA bases. For example, the amide group is connected to the heterocyclic ring via a σ bond so it can free rotate. As a result, the pyrrole carboxamide (**PCA**) exists in two different conformations, designated as *Syn* and *Anti,* (Figure 2.2). Furthermore, the azole ring exists in tautomeric forms due to interconversion of the N-H proton between the ring nitrogen atoms. Our previous report indicated that the tautomeric proton of **ICA** would preferably take a configuration with

18

the NH_2 of the amide at its trans position,[40] which is stabilized by the intramolecular

hydrogen bond. Our ^{1}H NMR data show that the **ICA** has two tautomer more uniformly

distributed with a ratio of 1 : 0.95, compared to **BCA** (1 : 0.72) and **TCA** (1 : 0.60 : 0.24)

(Figure 2.2). It should also be noted from the chemical shifts that the tautomeric protons

in the molecules have different acidities with an order of triazole's (**TCA**) >

benzoimidzole's (**BCA**) > imidazole's (**ICA**).
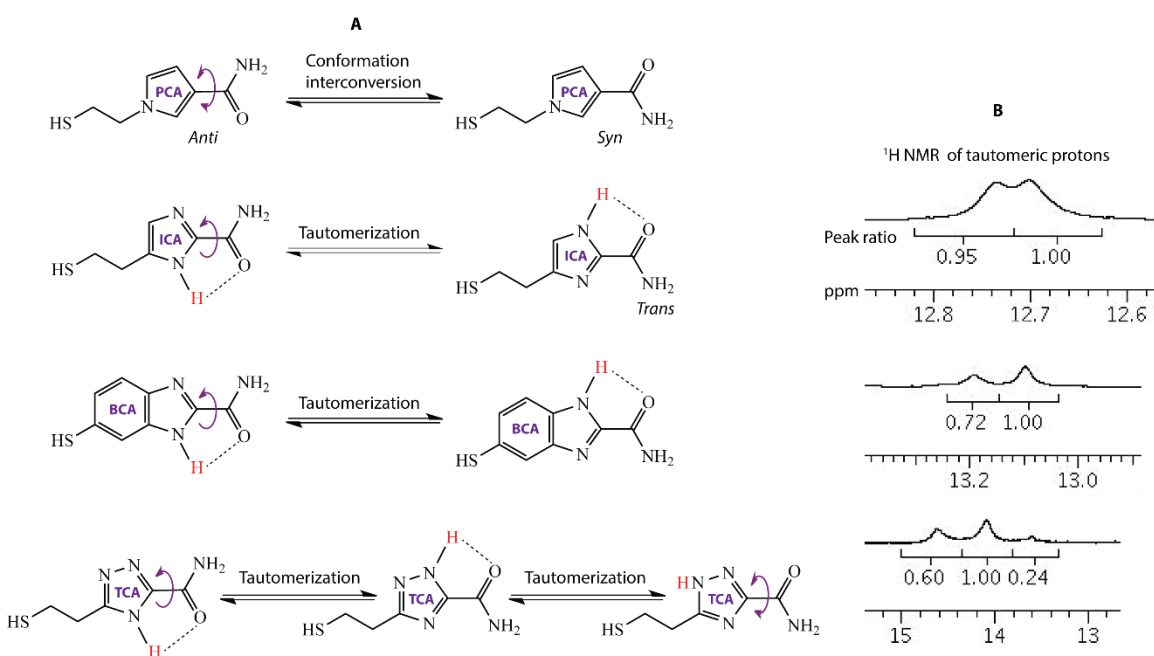
Figure 2.2. (**A**) Interconvertible conformation of the universal reader **PCA**, and proton tautomer of **ICA**, **BCA**, and **TCA**; (**B**) ^{1}HNMR spectra of tautomer (10 mM DMSO-d_6).

2.3. Recognition of DNA Bases by Electron Tunneling through Hydrogen Bonding Complexes in Nanogaps

We have used Scanning Tunneling Microscope (STM) to quickly create a nanogap

for studies of recognition tunneling (RT). In a typical RT experiment, a tunneling current

was set at 4 pA with a voltage bias of 0.5 V, which corresponded to a nanogap of ~ 2.4

19

nm distance.[78] The measurement followed a process of mounting a Pd probe and a Pd substrate both functionalized with a molecular reader to a PicoSPM scanning tunneling microscope, stabilizing the tunnel junction in a phosphate buffer (1.0 mM, 7.4 pH) until a clean baseline was generated (~ 2 h), introducing an analyte solution (typically 100 μM in 1.0 mM phosphate buffer, pH 7.4) to the liquid cell, and collecting current recordings for ~ 20 min under a predefined tip-substrate bias (Figure 2.3). In a typical RT spectra, peak amplitude is defined as the height of the current spike from the tunnel current baseline and peak width is define as the width of the current spike at half of its maximum height (see Figure 2.3).
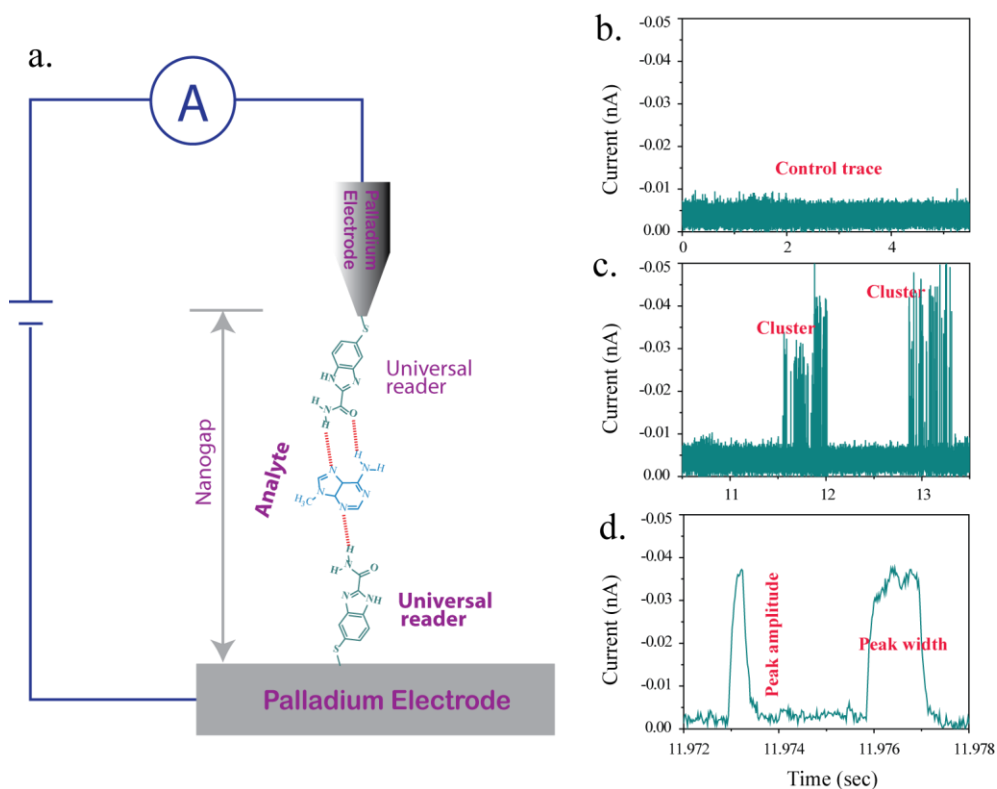
Figure 2.3. (a) A schematic of scanning tunneling microscope (STM) for recognition tunneling experiment; **(b)** Control trace in absence of any analyte molecule; **(c)** Typical RT spectra showing clusters; **(c)** Parameters of peaks obtained during a RT experiment.

In our present study, four naturally occurring DNA nucleoside monophosphates (referred to as **dAMP**, **dCMP**, **dGMP** and **dTMP**) were used as analytes (See Figure 2.4). For each analyte, four separate measurements were carried out separately with freshly made probes, substrates, and samples. Figure 2.4 presents typical recognition spectra of DNA nucleotides recorded in STM tunneling junctions functionalized with different molecular readers.
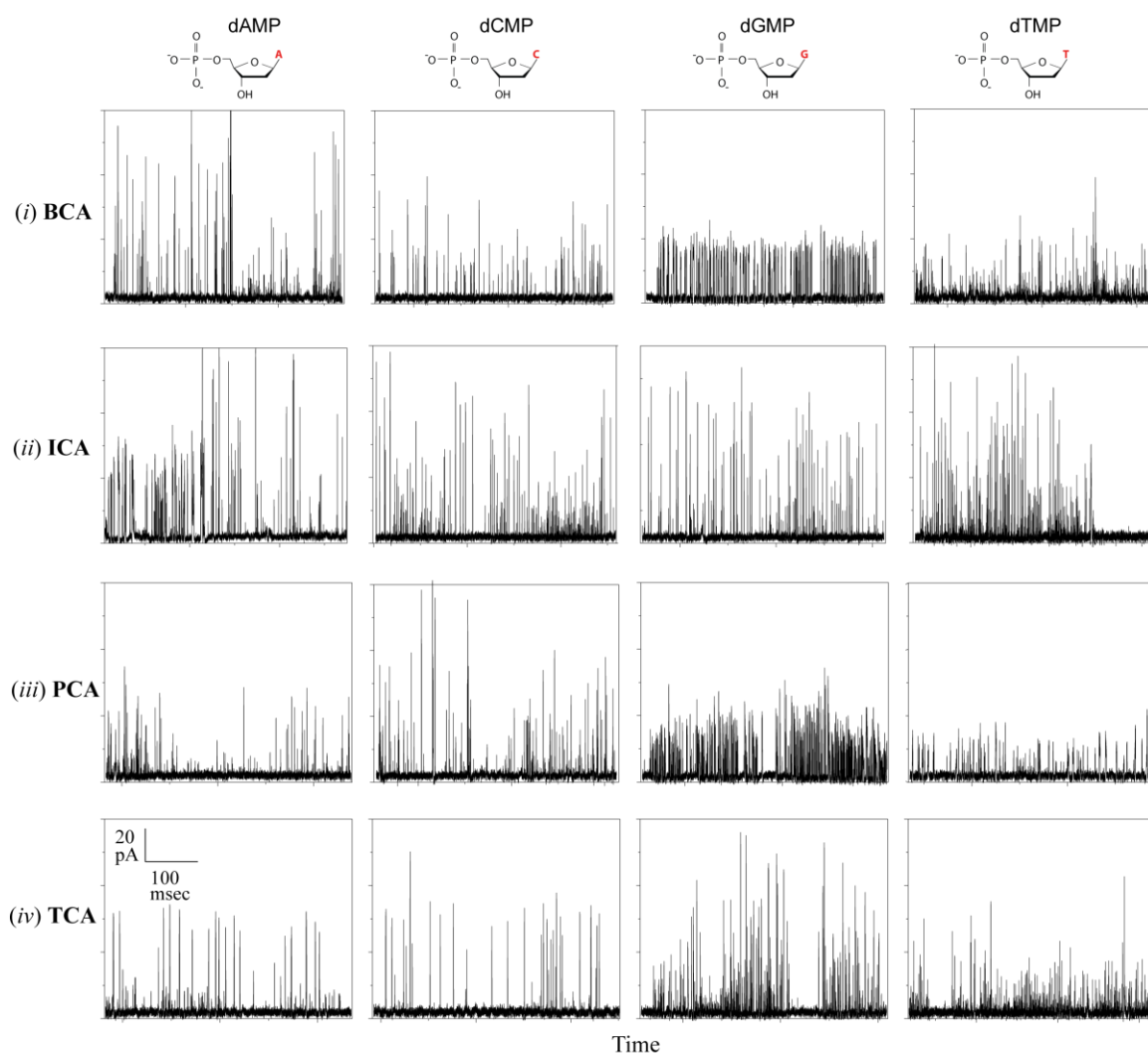
Figure 2.4. RT spectra generated with: (*i*) **BCA**, (*ii*) **ICA**, (*iii*) **PCA**, (*iv*) **TCA** functionalized tip and substrate at a setting point of 4 pA and 0.5 V.

Initially the spikes were analyzed by their averaged peak amplitude and peak width (date is not shown). The distributions of amplitude and peak width were very similar among different readers as well as individual DNA nucleotides. This may be explained by that these nucleotides form similar structures with different reader molecules in the nanogap, resulting in similar tunneling pathway. Nonetheless, a RT spectrum bears rich information on the trapped molecules beyond the above-mentioned parameters.

In order to call DNA bases, the tunneling current data were sequentially subjected to Fourier transform and cepstrum conversion, which produced a plethora of features for each of spikes and clusters. When those features were utilized to identify individual DNA bases, we found that there was no single feature that can be employed alone to effectively distinguish any two of DNA nucleotides from one another. However, a combination of two individual features has given a leap in the calling rate. As shown in Figure 2.5, a two-D plot can separated two DNA nucleotides **dAMP** (red dots) and **dGMP** (green dots) from each other with an efficiency (P, defined as accuracy) of 0.68 for the reader molecule **BCA**, 0.7 for **ICA**, 0.62 for **PCA**, and 0.79 for **TCA**.[79]
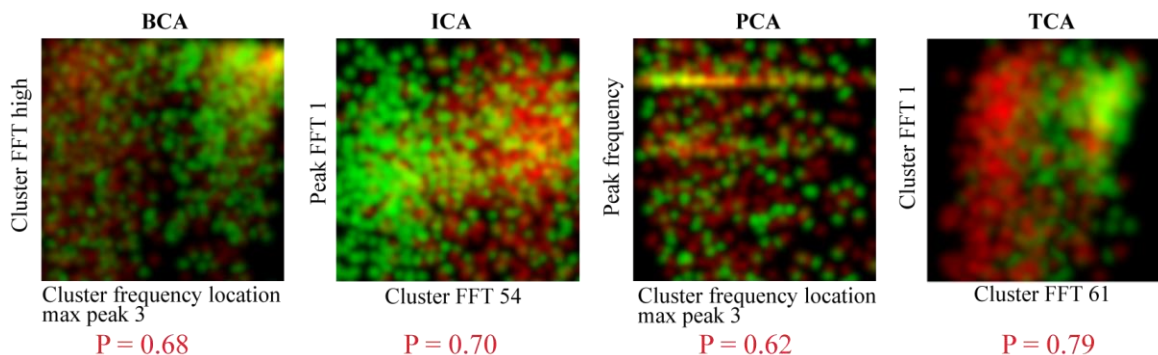
Figure 2.5. Two-D histograms of different readers' features where the brightness of each point represents the frequency value of the pair of features for **dAMP** (red) and **dGMP** (green), the accuracy (P) with which data can be assigned increases compared to one-D plot. Colors are yellowed with overlapped points.

The two-dimensional plot demonstrates an effective approach to calling DNA bases from the tunneling data. We have adapted a Support Vector Machine, SVM, a machine-learning algorithm to carry out the multidimensional analysis, which was previously applied to analyze the tunneling data generated with **ICA**.[80] Using the SVM method, the highest accuracy for each individual DNA nucleotide a reader molecule can achieve is listed in Table 2.1. These reader molecules read DNA bases differently with an accuracy order of **BCA** > **ICA** > **TCA** > **PCA** on average. Thus, **PCA** is removed from as a universal reader candidate due to its low accuracy. In contrast, **BCA** does not only have higher accuracy but also read DNA bases much less discriminately than **ICA** and **TCA**, based on their mean values and standard deviations of accuracy for four nucleotides. As a result, **BCA** is an ideal candidate of the universal reader for identification of DNA bases by RT.

Table 2.1. Highest accuracy (%) that can be achieved with different readers for determining individual DNA nucleotides by RT

	dAMP	dCMP	dGMP	dTMP	Mean ± σ
BCA	98.5	98.8	98.7	98.9	98.7 ± 0.1
ICA	96.5	97.4	96.4	98.1	97.1 ± 0.8
PCA	90.1	89.8	89.2	88.2	89.3 ± 0.8
TCA	94.3	95.5	96.5	99.0	96.3 ± 2.0

2.4. Recognition of DNA Bases by π-π Stacking as an Alternative to Hydrogen Bonding in Nanogaps by Electron Tunneling

We have studied hydrogen bonding interactions to read DNA bases and other molecules such as amino acids[81] and sugars by recognition tunneling. We also designed a new universal reader 1-(2-mercaptoethyl)pyrene (**Py**) based on π-π interactions, which should be more specific to the canonical DNA bases (see Figure 2.6). We found that the pyrene reader (**Py**) identified DNA bases with accuracy of 96.7% for **dTMP**, 97.1% for **dGMP**, 98.8% for **dAMP**, and 99.4% for **dCMP**, significantly higher than the imidazole-2-carboxamide reader **ICA** that had the accuracy of 96.4% for **dGMP**, 96.5% for **dAMP**, 97.4% for **dCMP**, and 98.1% for **dTMP**. However, **ICA** can read an abasic (AP) monophosphate, a product from spontaneous base hydrolysis or an intermediate of base excision repair. Our data analysis indicates that the signals of AP cannot be distinguished from those of DNA bases generated by **ICA**, but those by **Py** (see Figure 2.7).[82] Thus, sequencing DNA using both π-π stacking and hydrogen bonding based universal readers in parallel should generate more comprehensive genome sequences than current technologies.
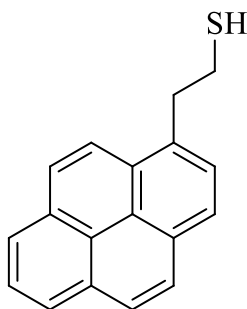
Figure 2.6. 1-(2-mercaptoethyl)pyrene (**Py**)
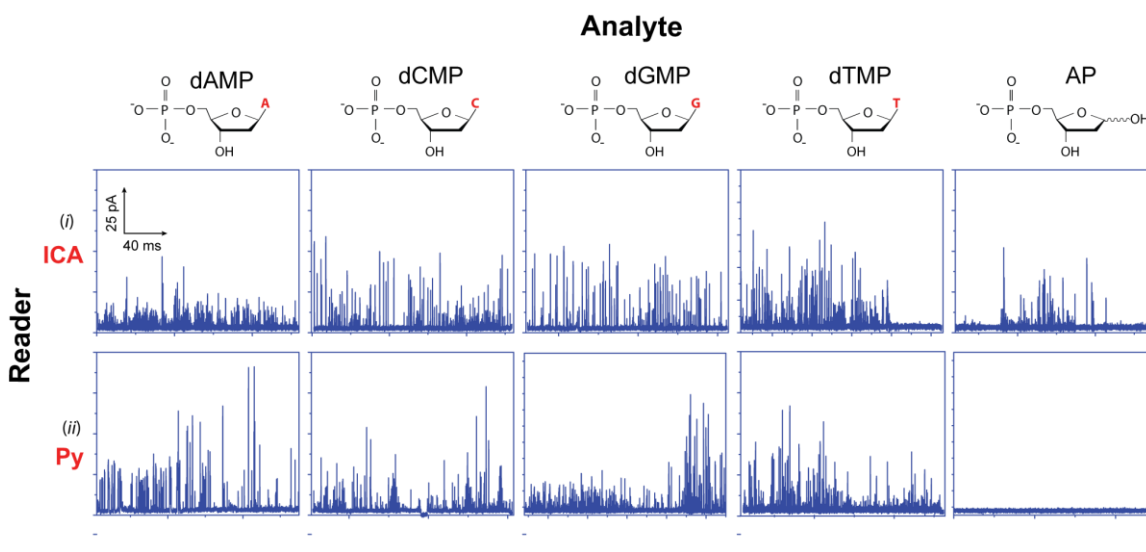
Figure 2.7. Examples of RT spectra generated with: (i) **ICA** functionalized tip and substrate at set point 4 pA and 0.5 V; (ii) **Py** functionalized tip and substrate at a set point of 2 pA and 0.5 V.

2.5. Synthesis of Universal Reader Candidates

We have developed new methods to synthesize these universal reader candidates. **PCA** was synthesized by a six-step process starting from commercially available 2-trichloroacetyl pyrrole (Scheme 2.1). This compound was reacted with chlorosulfonyl isocyanate (CSI) to produce corresponding cyano compound **1.1** in 80% yield. Trichloroacetyl group was converted to carboxylic acid by hydrolysis using sodium hydroxide to produce **1.2** in 90% yield. Decarboxylation under high temperature and using copper chromite catalyst produced **1.3** in 65% yield.[83] It was converted to its sodium form by reacting with sodium hydride, followed by reacting with benzyl 2-bromoethyl sulfide, which produced 1-(2-(benzylthio)ethyl)-1*H*-pyrrole-3-carbonitrile (**1.4**) in a 45% yield. The compound **1.4** was converted to 1-(2-(benzylthio)ethyl)-1*H*-pyrrole-3-carboxamide (**1.5**) by hydrolysis of the cyano group in a mixed solution of

25

sulfuric acid and trifluoroacetic acid with a 80% yield, based on a method reported in literature.[84] Removing the benzyl group of compound **1.5** with sodium in liquid ammonia at -78°C produced the desired compound **PCA** in a 50% yield.
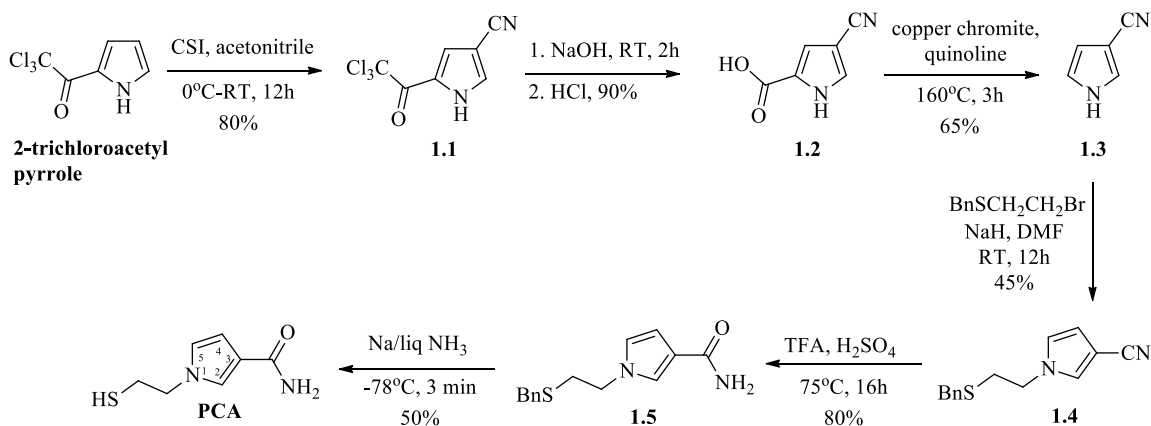
Scheme 2.1. Synthesis of 1-(2-mercaptoethyl)-1*H*-pyrrole-3-carboxamide (**PCA**)

BCA was synthesized by a five step process starting from commercially available 2-nitro-4-thiocyanatoaniline (Scheme 2.2). It was converted to free thiol according to literature reported procedure[85] to produce **2.1** in 96% yield. Thiol group was protected by benzyl group by reacting with benzyl bromide to provide compound **2.2** in 70% yield. The nitro group was reduced to amino group by reacting with sodium dithionite to provide compound **2.3** in 83% yield. The next step cyclization reaction was performed following the literature reported by Zavarzin and coworkers using oxamic acid monothiooxamide as an intermediate.[86] Compound **2.4** was synthesized in a one-pot reaction by heating a mixture of 4-(benzylthio)benzene-1,2-diamine (**2.3**)[87] with chloroacetamide and sulfur in the presence of triethyl amine in a fair yield (53%). We

26

fully characterized the compound **2.4** with ^{1}H and ^{13}C NMR, and high-resolution mass

spectrometry. Debenzylation of **2.4** with sodium readily furnished the universal reader

BCA with a 68% yield.
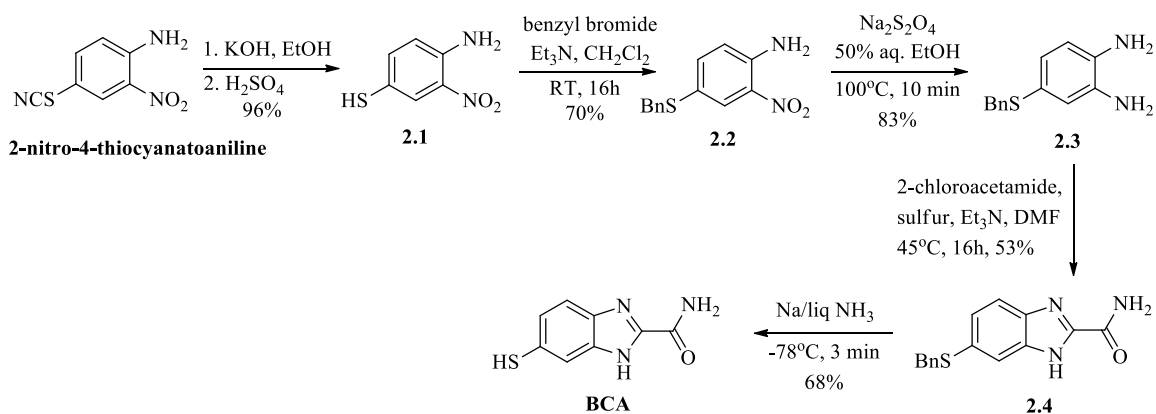
Scheme 2.2. Synthesis of 6-mercapto-1*H*-benzo[*d*]imidazole-2-carboxamide (**BCA**)

In a similar way, **TCA** was synthesized by four steps starting from 3-bromopropane

nitrile as the commercial available starting material (Scheme 2.3). It was converted to

compound **3.1** where bromine was replaced by benzylthio group in 65% yield. It was

converted to 3-(benzylthio)propanenitrile salt (**3.2**)[88] by reacting with anhydrous HCl in

ether and benzyl mercaptan in 97% yield. Then a key reaction of imido thioester with

hydrazide was followed.[89] In this step compound **3.2** was reacted with oxamic hydrazide

to form 1,2,4-triazole-5-carboxamide (**3.3**) in 40% yield. Finally, debenzylation of **3.3**

with sodium furnished the compound **TCA in** 41% yield.

Synthesis of 4(5)-(2-(mercaptoethyl)-1-*H*-imidazole-2-carboxamide (**ICA**) was

carried out following the literature.[40]

27
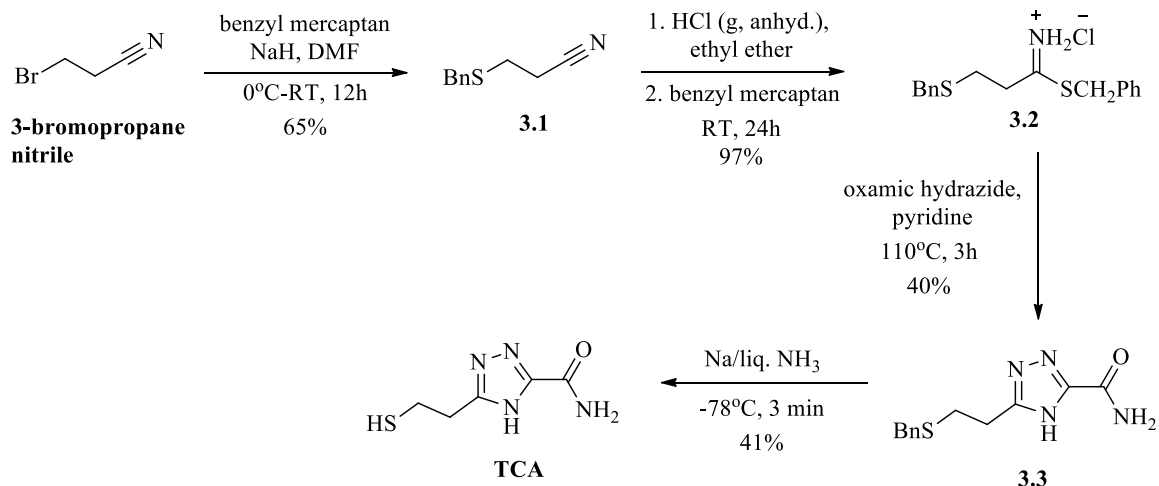
Scheme 2.3. Synthesis of 5-(2-mercaptoethyl)-4*H*-1,2,3-triazole-3-carboxamide (**TCA**)

2.6. Synthesis of Pyrene Reader Candidate

Synthesis of pyrene reader was reported in the following scheme (Scheme 2.4). It was synthesized in three steps from commercially available 1-bromopyrene. Purchased 1-bromopyrene (95% purity) was first purified by silica gel flash chromatography eluting with hexane, dried at 40°C overnight, and stored over drierite under vacuum. THF was freshly distilled over sodium prior to use. Nitrogen was flowed through drierite before it went into the reaction vessel. Ethylene oxide (1.2 M solution in dichloromethane) was stored over molecular sieves for two days before use. In the first step a Grignard reaction was performed between 1-bromopyrene and ethylene oxide to provide compound **4.1** in 40% yield. The hydroxyl group was protected by tosyl group by reacting with tosyl chloride to provide **4.2** in 82% yield. The tosyl was replaced by thioacetate to produce **4.3** in 78% yield. It was converted to free thiol by in situ addition of pyrrolidine to produce **Py** in quantitative yield (not shown in the scheme but synthesis is reported).
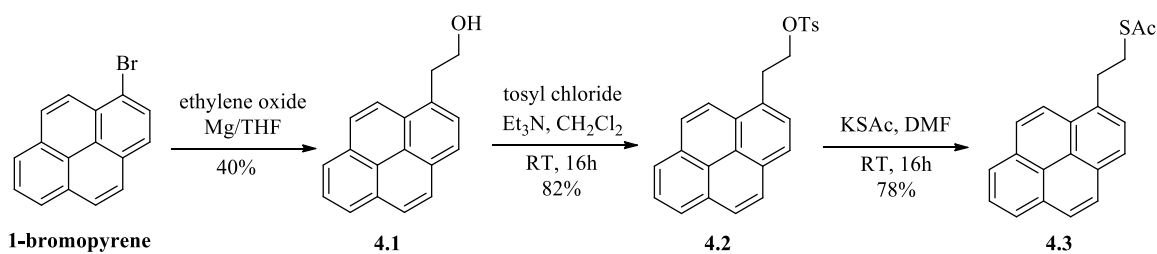
Scheme 2.4. Synthesis of S-(2-(pyren-1-yl)ethyl) ethanethioate

2.7. Experimental Procedures

Reagents and solvents were purchased from commercial suppliers (Sigma-Aldrich, Alfa Aesar, Fluka, TCI America) and used as received unless otherwise noted. All experiments requiring anhydrous conditions were performed in flame-dried glassware under nitrogen atmosphere. Reactions were monitored by thin layer chromatography (TLC) using glass plates precoated with silica gel (EMD Chemicals Inc.). Flash chromatography was performed in an automated flash chromatography system (CombiFlash R_f, Teledyne Isco, Inc.) with silica gel columns (60-120 mesh). ^{1}H NMR and ^{13}C NMR spectra were recorded on Varian INOVA 400 (400 MHz) and Varian INOVA 500 (500 MHz) spectrometers at 25°C at the Magnetic Resonance Research Center at Arizona State University. Chemical shifts (δ) are given in parts per million (ppm) and are referenced to the residual solvent peak (CDCl$_3$: $\delta_H = 7.26$ ppm, CD$_3$OD: $\delta_H = 3.31$ ppm, DMSO-d$_6$: $\delta_H = 2.50$ ppm). Coupling constants (J) are expressed in hertz (Hz) and the values are rounded to the nearest 0.1 Hz. Splitting patterns are reported as follows: br, broad; s, singlet; d, doublet; dd, doublet of doublets; t, triplet; dt, doublet of

triplets; q, quartet and m, multiplet. High resolution mass spectra (HRMS) are acquired at the Arizona State University CLAS High Resolution Mass Spectrometry Facility.
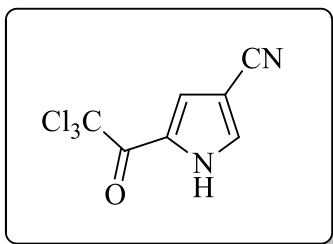
2-trichloroacetylpyrrole-4-carbonitrile (1.1). 2-Trichloroacetylpyrrole (6.0 g, 28.23 mmol) was dissolved in acetonitrile (30 mL) and cooled in ice bath. A solution of chlorosulfonyl isocyanate or CSI (9.6 g, 67.84 mmol) in acetonitrile (9 mL) was added dropwise into the stirred solution. The reaction mixture was warmed to room temperature and stirred for 12h. It was then cooled in ice bath, DMF (24 mL) was added, heated first to 50°C for 1h and then stirred at room temperature for 3h. The mixture was poured onto crushed ice (~100 g) and extracted with CH_2Cl_2 (3×40 mL). The combined organic layers were washed with aqueous sodium hydrogen carbonate (1% 30 mL), brine (30mL), dried over MgSO4, filtered and concentrated by rotary evaporator. The residue was separated in a silica gel column by flash chromatography using a gradient of ethyl acetate (0-30% for 3h) in hexane. Compound **1.1** was obtained as white solid (5.35 g, 80%). ^{1}H NMR (400 MHz, CDCl$_3$): δ 9.80 (s, br, 1H, pyrrole-N<u>H</u>), 7.59 (s, 1H, Ar-<u>H</u>), 7.58 ppm (s, 1H, Ar-<u>H</u>); HRMS (EI+) *m/z*: calculated for C$_7$H$_3$N$_2$OCl$_3$: 235.9311; measured: 235.9318.
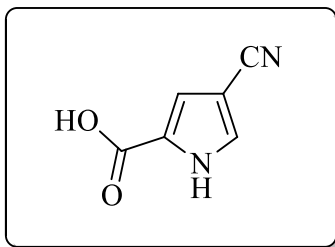
4-cyano-1*H*-pyrrole-2-carboxylic acid (1.2). Compound **1.1** (4.0 g, 16.84 mmol) was

added in an aqueous sodium hydroxide solution (2M, 40 mL) cooled in ice bath. It was

warmed to room temperature and stirred for 2h during which solution became cloudy.

The solution was acidified with concentrated HCl and the precipitate of product was

collected by vacuum filtration, washed with cold water and dried at 110°C overnight. The

product **1.2** (2.07 g, 90%) was obtained as a pale brown solid and used without further

purification. ^{1}H NMR (400 MHz, DMSO-d_6): δ 12.98 (s, br, 1H, COO<u>H</u>), 12.70 (s, br,

1H, pyrrole-N<u>H</u>), 7.76 (s, 1H, Ar-<u>H</u>), 7.12 ppm (s, 1H, Ar-<u>H</u>); ^{13}C NMR (100 MHz,

DMSO-d_6): δ 160.9, 130.6, 125.0, 117.1, 116.2, 92.9 ppm; HRMS (EI+) *m/z*: calculated

for $C_6H_4N_2O_2$: 136.0273; measured: 136.0269.
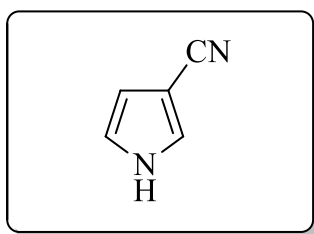
1*H*-pyrrole-3-carbonitrile (1.3). Compound **1.2** (1.8 g, 13.24 mmol) was mixed with

quinoline (12 mL) followed by the addition of copper chromite catalyst (1.2 g). The

stirred mixture was heated in an oil bath at 160°C for 3h. The resulting brown mixture

was cooled, diluted with ether (200 mL) and filtered through celite bed to remove

insoluble materials. The ether solution was washed with four portions (30 mL each) of

1M HCl and then with aqueous sodium bicarbonate solution (5%). The pale yellow ether

layer was collected, dried over MgSO$_4$, filtered and concentrated by rotary evaporator.

The residue was separated in a silica gel column by flash chromatography using a

gradient of ethyl acetate (0-30% for 3.5h) in hexane. Compound **1.3** was obtained as pale

yellow solid (0.8 g, 65%). ^{1}H NMR (400 MHz, CDCl$_3$): δ 9.29 (s, br, 1H, pyrrole-N$\underline{H}$),

7.31 (m, 1H, pyrrole-$\underline{H}$), 6.80 (m, 1H, pyrrole-$\underline{H}$), 6.47 ppm (m, 1H, pyrrole-$\underline{H}$); ^{13}C

NMR (100 MHz, DMSO-d_6): δ 126.2, 119.7, 117.6, 111.8, 92.7 ppm; HRMS (APCI+)

m/z: calculated for C$_5$H$_4$N$_2$+H: 93.0453; measured: 93.0455.
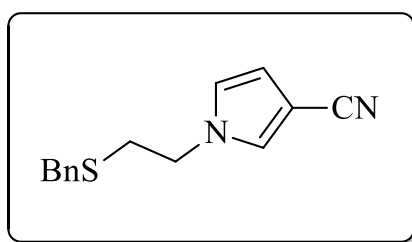
1-(2-(benzylthio)ethyl)-1*H*-pyrrole-3-carbonitrile (1.4). A solution of 1*H*-pyrrole-3-

carbonitrile (100 mg, 1.08 mmol) in anhydrous DMF (0.5 mL) was added dropwise into a

suspension of NaH (39 mg, 1.62 mmol) in anhydrous DMF (1.0 mL) at 0 °C under

nitrogen. The mixture was stirred for 30 min, to which a solution of 2-bromoethyl benzyl

sulfide (301 mg, 1.30 mmol) in anhydrous DMF (0.5 mL) was added dropwise. The

reaction mixture was stirred at room temperature for 12 h, followed by the addition of a

saturated NH$_4$Cl solution of (10 mL) and extracting the mixture with ethyl acetate (3 × 10

mL). The combined organic phase was washed with brine three times (each 10 mL), dried

over MgSO$_4$, filtered, and concentrated by rotary evaporator. The residue was separated

in a silica gel column by flash chromatography using a gradient of ethyl acetate (0-30%

in 3 h) in hexane. Compound **1.4** was obtained as a colorless liquid (117 mg, 45%). ^{1}H

NMR (500 MHz, CDCl$_3$): δ 7.27-7.36 (m, 5H, Ar$\underline{H}$), 7.07 (t, J = 1.8 Hz, 1H, pyrrole-H2), 6.58 (dd, J = 2.7 Hz, J = 1.8 Hz, pyrrole-H5), 6.40 (dd, J = 2.7 Hz, J = 1.7 Hz, pyrrole-H4), 3.90 (t, J = 7.0 Hz, 2H, SCH$_2$C$\underline{H}_2$), 3.58 (s, 2H, PhC$\underline{H}_2$), 2.71 ppm (t, J = 7.0 Hz, 2H, SC$\underline{H}_2$CH$_2$); ^{13}C NMR (125 MHz, CDCl$_3$): δ 137.8, 129.1, 128.9, 128.1, 127.6, 122.2, 116.9, 112.5, 93.1, 50.3, 36.7, 32.4 ppm; HRMS (APCI+): found m/z 243.0962; calculated for C$_{14}$H$_{14}$N$_2$S+H: 243.0956.
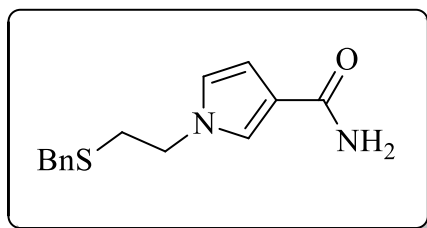
1-(2-(benzylthio)ethyl)-1H-pyrrole-3-carboxamide (1.5). Compound **1.4** (100 mg, 0.413 mmol) was added into a mixture of trifluoroacetic acid (0.8 mL) and aqueous solution of sulfuric acid (1.0 mL, 20% v/v) at room temperature. The resulting mixture was heated at 75°C with stirring for 16 h. TLC analysis indicated the reaction was completed (eluent: 5% methanol in dichloromethane; R$_f$ = 0.37). The solution was neutralized by the addition of a saturated NaHCO$_3$ solution (10 mL), extracted with ethyl acetate (3 × 10 mL). The combined organic phases were washed with brine (10 mL), dried over MgSO$_4$, filtered and concentrated by rotary evaporator. The residue was separated in a silica gel column by flash chromatography with a gradient of methanol (0-5% in 3 h) in dichloromethane to obtain compound **1.5** (86 mg, 80%). ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.22-7.30 (m, 5H, Ar$\underline{H}$), 7.19 (t, J = 1.6 Hz, 1H, pyrrole-H2), 6.52 (dd, J = 2.8 Hz, J = 2.0 Hz, pyrrole-H5), 6.38 (dd, J = 2.8 Hz, J = 1.6 Hz, pyrrole-H4), 5.98 (s, br, 2H, amide N$\underline{H}_2$), 3.85 (t, J = 7.0 Hz, 2H, SCH$_2$C$\underline{H}_2$), 3.52 (s, 2H, PhC$\underline{H}_2$), 2.66 (t, J =

7.0 Hz, 2H, SC$\underline{H}_2$CH$_2$) ppm; ^{13}C NMR (100 MHz, CDCl$_3$): δ 167.2, 138.0, 129.2, 128.9, 127.6, 124.4, 122.0, 119.4, 108.4, 50.2, 36.7, 32.5 ppm; HRMS (APCI+): found m/z 261.1064; calculated for C$_{14}$H$_{16}$N$_2$OS+H: 261.1062.
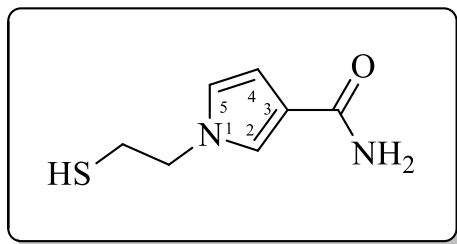
1-(2-mercaptoethyl)-1*H*-pyrrole-3-carboxamide (PCA). A solution of compound **1.5** (50 mg, 0.19 mmol) in anhydrous THF (1.0 mL) was added to liquid ammonia (~2 mL) at -78°C under nitrogen and the resulting solution was stirred for 10 min. Small pieces of freshly cut sodium were added one after the last one disappeared until a blue color was able to remain unfading for 3 min, and then NH$_4$Cl was added until the blue color disappeared. The ammonia was allowed to evaporate at room temperature under a flow of nitrogen. The organic solvent was removed by rotary evaporation. The residue was separated in a silica gel column by flash chromatography with a gradient of methanol (0-10% in 2.5 h) in dichloromethane. Product **PCA** was obtained as a white solid (16 mg, 50%). ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.28 (t, J = 2.0 Hz, 1H, pyrrole-H2), 6.63 (dd, J = 2.9 Hz, 2.3 Hz, pyrrole-H5), 6.40 (dd, J = 2.9, 1.9 Hz, pyrrole-H4), 5.81 (s, br, 2H, amide N$\underline{H}_2$), 4.06 (t, J = 6.6 Hz, 2H, SCH$_2$C$\underline{H}_2$), 2.84 (dt, J = 8.4, 6.6 Hz, 2H, SC$\underline{H}_2$CH$_2$), 1.33 (t, J = 8.6, 1H, S$\underline{H}$) ppm; ^{13}C NMR (100 MHz, CDCl$_3$): δ 167.0, 124.5, 122.1, 119.5, 108.4, 53.3, 26.3 ppm; HRMS (APCI+): found m/z 171.0595; calculated for C$_7$H$_{10}$N$_2$OS+H: 171.0592.
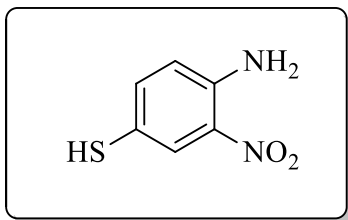
4-amino-3-nitrobenzenethiol (2.1). 2-nitro-4-thiocyanatoaniline (3.51 g, 18 mmol) was

used as the commercially available starting material to synthesize compound **2.1** (2.94 g,

96%) following the procedure reported in literature.[85]
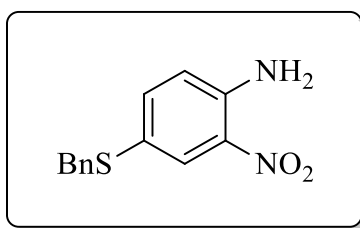
4-(benzylthio)-2-nitroaniline (2.2). Triethylamine (2.44 mL, 17.65 mmol) was added

into a solution of **2.1** (2.0 g, 11.76 mmol) in dichloromethane (20 mL) and stirred for 30

min. Benzyl bromide (1.68 mL, 14.12 mmol) was added into the solution and it was

stirred for 16 h at room temperature. The reaction mixture was diluted by adding

dichloromethane (80 mL), washed with saturated sodium bicarbonate solution (50 mL)

and brine (50 mL), dried over $MgSO_4$, filtered and concentrated by rotary evaporator.

The residue was separated in a silica gel column by flash chromatography using a

gradient of ethyl acetate (0-10% in 3 h) in hexane. Compound **2.2** was obtained as red

solid (2.1 g, 70%). ^{1}H NMR (400 MHz, CDCl$_3$): δ = 8.10 (d, J = 2 Hz, 1H, Ar$\underline{H}$), 7.19-

7.29 (m, 6H, Ar$\underline{H}$), 6.67 (d, J = 8.8 Hz, 1H, Ar$\underline{H}$), 6.08 (s, br, 2H, N$\underline{H}_2$), 3.99 (s, 2H,

SC$\underline{H}_2$Ph); ^{13}C NMR (100 MHz, CDCl$_3$): δ = 144.3, 140.5, 137.7, 132.4, 130.5, 129.3,

128.8, 127.6, 122.8, 119.5, 41.3 ppm; HRMS (APCI+): m/z calculated for

$C_{13}H_{12}N_2O_2S$+H: 261.0698; measured: 261.0694.
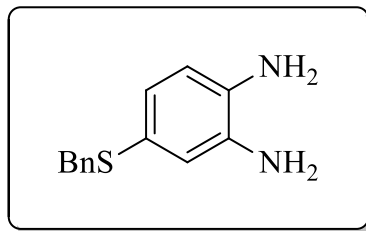
4-(benzylthio)benzene-1,2-diamine (2.3). Compound **2.2** (1.0 g, 3.85 mmol) was dissolved in 50% aqueous ethanol (total volume 40 ml). Sodium dithionite (4.02 g, 23.10 mmol) was added portion wise into the solution over a period of 20 min. The stirred solution was gradually heated to 100°C and refluxed for about 10 min while the red solution became colorless. It was cooled at room temperature and the solvents were evaporated in rotary evaporator. The crude was extracted with boiling methanol (3×50 ml) and filtered through celite bed under vacuum suction. Silica gel was added to the solution and concentrated by rotary evaporator. The residue was separated in a silica gel column by flash chromatography using a gradient of methanol (0-2% in 2 h) in dichloromethane to obtain **2.3**. Yield: 0.73 g (83%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta =$ 7.18-7.24 (5H, m, ArH), 6.67-6.72 (2H, m, ArH), 6.55 (1H, d, $J = 8.0$ Hz, ArH), 3.95 ppm (2H, s, CH$_2$), 3.33 (4H, broad, NH$_2$),; ^{13}C NMR (100 MHz, CDCl$_3$): $\delta = 138.7$, 135.2, 134.9, 129.2, 128.6, 127.2, 125.4, 125.2, 121.3, 117.1, 41.7 ppm; HRMS (APCI+): m/z calculated for C$_{13}$H$_{14}$N$_2$S+H: 231.0956; measured: 231.0957.
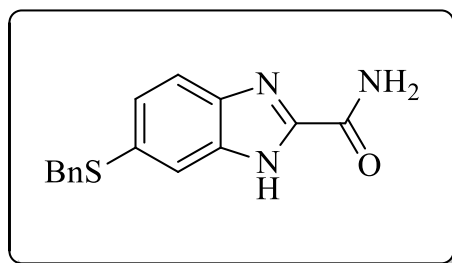
6-(benzylthio)-1*H*-benzo[*d*]imidazole-2-carboxamide (2.4). 2-Chloroacetamide (0.19

g, 2.0 mmol) was added to a mixture of **2.3** (0.46 g, 2.0 mmol), sulfur (0.26 g, 8.1 mmol)

and triethyl amine (0.5 mL) in DMF (5 mL). The reaction mixture was stirred at 45°C for

16 h, cooled to room temperature, diluted with water (20 mL), and extracted with ethyl

acetate (3 × 20 mL). The combined organic phases were washed with brine three times

(each 10 mL), dried over $MgSO_4$, filtered and concentrated by rotary evaporator. The

residue was separated in a silica gel column by flash chromatography with a gradient of

methanol (0-5% in 4 h) in dichloromethane to obtain compound **2.4** (0.3 g, 53%). [1]H

NMR (400 MHz, DMSO-d$_6$): δ 13.15 (s, 1H, imidazole N<u>H</u>), 8.24 (s, broad, 1H, amide

N<u>H</u>$_2$), 7.82 (s, broad, 1H, amide N<u>H</u>$_2$), 7.63 (d, *J* = 8.0 Hz, 1H, Ar*H*), 7.43 (d, *J* = 8.8 Hz,

1H, Ar<u>H</u>), 7.19-7.32 (m, 6H, ArH), 4.21 (2H, d, *J* = 8.8 Hz, CH$_2$); HRMS (APCI+) *m/z*:

calculated for $C_{15}H_{13}N_3SO+H$: 284.0858, measured: 284.0853.
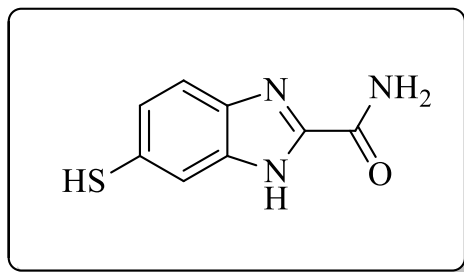
6-mercapto-1*H*-benzo[*d*]imidazole-2-carboxamide (BCA). A solution of compound

2.**4** (0.1 g, 0.35 mmol) in anhydrous THF (1.0 mL) was added to liquid ammonia (~3

mL) at -78°C under nitrogen and the resulting solution was stirred for 10 min. Small

pieces of freshly cut sodium were added one after the last one disappeared until a blue

color be able to remain unfading for 3 min, and then NH$_4$Cl was added until the blue

color disappeared. The ammonia was allowed to evaporate at room temperature under a

flow of nitrogen. The organic solvent was removed by rotary evaporation. The residue

was separated in a silica gel column by flash chromatography with a gradient of methanol (0-10% in 3 h) in dichloromethane to obtain compound **BCA** (47 mg, 68%). ^{1}H NMR (400 MHz, DMSO-d_6): δ 13.10 (m, 1H, imidazole N$\underline{H}$), 8.20 (s, broad, 1H, amide N$\underline{H}_2$), 7.79 (s, broad, 1$\underline{H}$, amide NH$_2$), 7.58-7.65 (m, 1H, ArH), 7.40-7.45 (m, 1H, ArH), 7.15-7.22 (m, 1H, ArH), 5.59 (m, 1H, SH) ppm; HRMS (APCI+) m/z: calculated for $C_8H_7N_3OS+H$: 194.0388; measured: 194.0383.
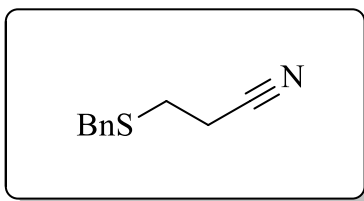
3-(benzylthio)propanenitrile (3.1). Benzyl mercaptan (1.05 g, 19.0 mmol) was added into a stirred solution of sodium hydride (60% in mineral oil, 1.16 g, 24.0 mmol) in anhydrous DMF (50 mL) at 0 ^{0}C under inert atmosphere. The reaction mixture was stirred for another 30 min followed by the slow addition of 3-bromopropanenitrile (2.68 g, 20.0 mmol). The resulting mixture was allowed to warm at room temperature and stirred for 12 h while one of the starting materials was consumed completely. The solvent was removed by rotary evaporation. A saturated solution of aqueous NH$_4$Cl was added into the mixture and extracted with chloroform (3×20 mL). The combined organic extracts were washed with brine (30 mL) and dried over magnesium sulfate. The solution was then filtered and concentrated by rotary evaporator. The crude product was purified by silica gel flash column chromatography. Pure product **3.1** (2.25 g, 65%) was obtained as pale yellow liquid. ^{1}H NMR (400 MHz, CDCl$_3$): δ = 7.24-7.33 (5H, m, ArH), 3.78 (2H, s, PhCH$_2$), 2.64 (2H, t, J = 8.0 Hz, CH$_2$), 2.47 ppm (2H, t, J = 8.0 Hz, CH$_2$); [13]C

NMR (100 MHz, CDCl₃): δ = 137.2, 128.9, 127.2, 118.3, 36.0, 26.2, 18.3 ppm; HRMS (APCI+): m/z calculated for $C_{10}H_{11}NS+H$: 178.0690; found: 178.0688.
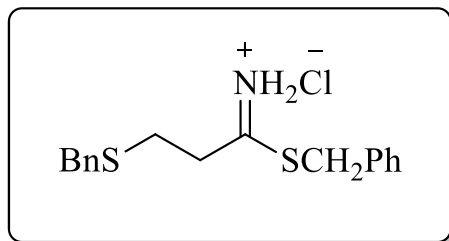
benzyl 3-(benzylthio)propanimidothioate hydrochloride (3.2). Hydrogen chloride was bubbled into a solution of **3.1** (2.0 g, 11.3 mmol) and benzyl mercaptan (2.0 mL, 17.0 mmol) in anhydrous ethyl ether (120 mL) in an ice bath under nitrogen for 2 h. The solution was capped and allowed to warm to room temperature, stirred for another 24 h. The reaction solution then stood still for 2 h, from which the compound **3.2** crystallized out. The product was filtered through a Buchner funnel, washed with cold ethyl ether three times (each 20 mL), and dried in vacuum at room temperature overnight. It weighed 3.7 g (yield: 97%). ¹H NMR (400 MHz, CDCl₃): δ 12.59 (s, br, 1H, NH₂), 11.75 (s, br, 1H, NH₂), 7.21-7.39 (m, 10H, ArH), 4.78 (s, 2H, CSCH₂Ph), 3.86 (s, 2H, PhCH₂SCH₂), 3.20 (t, J = 7.2 Hz, 2H, SCH₂CH₂), 2.88 (t, J = 7.0 Hz, 2H, SCH₂CH₂) ppm; ¹³C NMR (100 MHz, CDCl₃): δ 193.3, 137.8, 131.3, 129.8, 129.6, 129.5, 129.3, 129.0, 127.6, 39.3, 37.5, 36.4, 29.8 ppm; HRMS (APCI+) m/z: calculated for $C_{17}H_{19}NS_2+H$: 302.1037; measured: 302.1036.
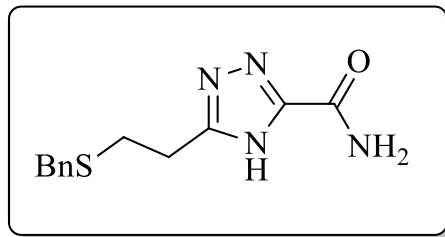
5-(2-(benzylthio)ethyl)-*4H*-1,2,4-triazole-3-carboxamide (3.3). A solution of **3.2** (1.01 g, 3.0 mmol) and oxamic hydrazide (0.31 g, 3.0 mmol) in anhydrous pyridine (10 mL) was refluxed at 110°C for 3 h. The solvent was removed by co-evaporating with toluene (5 mL × 2) by rotary evaporation. The yellow oily residue was dissolved in DMSO (15 mL), to which water (50 mL) was added, resulting in a white precipitate. The product was filtered through a Buchner funnel, washed thoroughly with cold water (40 mL), cold ethyl ether (40 mL), and dried in vacuum at room temperature. It was then dissolved in boiling ethanol (~25 mL) and allowed to recrystallize by slowly cooling the solution down to room temperature, filtered, and dried in vacuum at 40 $^\circ$C for overnight. Compound **3.3** was obtained as white crystals (0.31 g, 40%). ^{1}H NMR (400 MHz, DMSO-d_6): δ 14.26 (s, broad, 1H, N<u>H</u>), 7.83 (s, broad, 1H, N<u>H</u>$_2$), 7.61 (s, broad, 1H, N<u>H</u>$_2$), 7.21-7.32 (m, 5H, Ar<u>H</u>), 3.74 (s, 2H, PhC<u>H</u>$_2$), 2.97 (t, J = 7.2 Hz, 2H, SCH$_2$C<u>H</u>$_2$), 2.77 ppm (t, J = 7.2 Hz, 2H, SC<u>H</u>$_2$CH$_2$); ^{13}C NMR (100 MHz, DMSO-d_6): δ = 160.5, 160.1, 156.1, 138.4, 128.8, 128.4, 126.8, 34.9, 28.7, 26.7 ppm; HRMS (APCI+) *m/z*: calculated for C$_{12}$H$_{14}$N$_4$OS+H: 263.0967; measured: 263.0972.
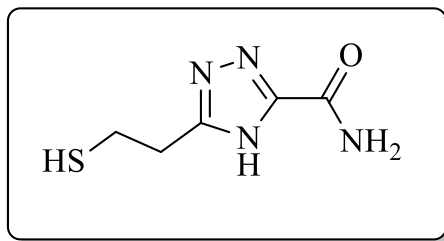
5-(2-mercaptoethyl)-4_H_-1,2,4-triazole-3-carboxamide (TCA). A solution of compound

3.3 (150 mg, 0.57 mmol) in anhydrous THF (1.0 mL) was added to liquid ammonia (~2

mL) at -78°C under nitrogen and the resulting solution was stirred for 10 min. Small

pieces of freshly cut sodium were added one after the last one disappeared until a blue

color was able to remain unfading for 3 min, and then NH_4Cl was added until the blue

color disappeared. The ammonia was allowed to evaporate at room temperature under a

flow of nitrogen. The organic solvent was removed by rotary evaporation. The residue

was separated in a silica gel column by flash chromatography with a gradient of methanol

(0-10% in 2 h) in dichloromethane. Product **TCA** was obtained as a white solid (40 mg,

41%). ^{1}H NMR (400 MHz, DMSO-d_6): δ = 6.79 (s, broad, 1H, N$\underline{H}_2$), 6.71 (s, broad, 1H,

N$\underline{H}_2$), 2.57 (t, J = 6.8 Hz, 2H, SC$\underline{H}_2$CH$_2$); 2.43 (t, J = 6.8 Hz, 2H, SCH$_2$C$\underline{H}_2$), 2.08 (t, J =

2.0 Hz, 1H, S$\underline{H}$); HRMS (APCI+) m/z: calculated for $C_5H_8N_4OS$+H: 173.0497;

measured: 173.0493.
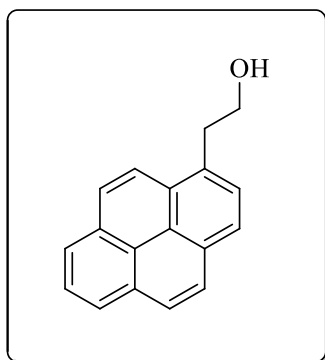
2-(pyren-1-yl)ethanol (4.1). A solution of 1-bromopyrene (0.4 g, 1.42 mmol in 12 mL THF) was added onto magnesium turnings (0.1 g, 4.27 mmol) in a flame-dried Schlenk flask.[90,91] It was refluxed at 70°C while the solution turned into dark brown color and continued to reflux for another 2 h. The resulting solution was cooled in ice bath followed by addition of ethylene oxide solution (3.6 mL, 4.27 mmol in 6mL THF). The mixture was allowed to warm to room temperature and stirred 12 h. It was cooled in ice bath then hydrolyzed by careful addition of HCl (5 mL 10%). Organic compounds were extracted with ethyl acetate (20 mL × 2). The combined organic layers were washed with brine (40 mL), dried over $MgSO_4$, filtered and dried in rotary evaporator. Product was purified through silica gel column by flash chromatography using a gradient of ethyl acetate (0 - 20% for 3 h) in hexane. Compound **4.1** was obtained as yellow solid (0.14 g, 40%). ^{1}H NMR (500 MHz, $CDCl_3$): δ 8.29 (d, J = 9.0 Hz, 1H, Ar$\underline{H}$), 8.18 (d, J = 8.0 Hz, 2H, Ar$\underline{H}$), 8.10-8.13 (m, 2H, Ar$\underline{H}$), 7.99-8.04 (m, 3H, Ar$\underline{H}$), 7.90 (d, J = 8.0 Hz, 1H, Ar$\underline{H}$), 5.29 (s, br, 1H, O$\underline{H}$), 4.09 (t, J = 6.5 Hz, 2H, C$\underline{H}_2$CH$_2$OH), 3.61 ppm (t, J = 6.5 Hz, 2H, CH$_2$C$\underline{H}_2$OH); ^{13}C NMR (125 MHz, $CDCl_3$): δ 132.53, 131.52, 130.98, 130.42, 129.36, 128.07, 127.68, 127.57, 127.09, 126.06, 125.24, 125.19, 125.00, 124.97, 123.30, 63.95, 36.78, 29.85 ppm; HRMS (FAB+) m/z: calculated for $C_{18}H_{14}O$+H: 247.1123; measured: 247.1129.

42
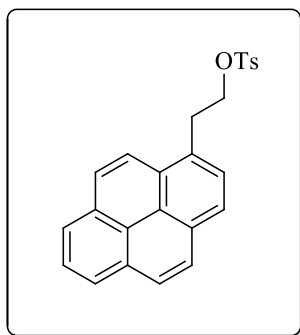
2-(pyren-1-yl)ethyl 4-methylbenzenesulfonate (4.2). Triethyl amine (0.08 mL, 0.55

mmol) was added into a solution of compound **4.1** (45 mg, 0.18 mmol) and tosyl chloride

(52 mg, 0.28 mmol) in 1.5 mL dichloromethane at room temperature. The resulting

solution was stirred for 16 h followed by addition of saturated sodium bicarbonate

solution (5 mL). Organic compounds were extracted by dichloromethane (3×10 mL).

Combined organic extract was washed with brine (30 mL), dried over $MgSO_4$, filtered

and evaporated to dryness in rotary evaporator. Product was purified through silica gel

column by flash chromatography using a gradient of ethyl acetate (0 - 20% for 3 h) in

hexane. Compound **4.2** was obtained as white solid (60 mg, 82%). ^{1}H NMR (500 MHz,

CDCl$_3$): δ 8.17 (q, $J = 7.5$ Hz, 2H, Ar<u>H</u>), 7.95-8.05 (m, 6H, Ar<u>H</u>), 7.76 (d, $J = 7.5$ Hz,

1H, Ar<u>H</u>), 7.33 (d, $J = 8.5$ Hz, 2H, Ar<u>H</u>), 6.68 (d, $J = 8.5$ Hz, 2H, Ar<u>H</u>), 4.44 (t, $J = 7.0$

Hz, 2H, C<u>H</u>$_2$CH$_2$OTs), 3.64 (t, $J = 7.0$ Hz, 2H, CH$_2$C<u>H</u>$_2$OTs), 1.85 ppm (s, 3H, C<u>H</u>$_3$); [13]C

NMR (125 MHz, CDCl$_3$): δ 144.28, 132.23, 131.37, 130.73, 130.65, 129.96, 129.17,

128.87, 128.28, 127.80, 127.47, 127.40, 127.23, 126.08, 125.32, 125.13, 125.06, 124.80,

124.79, 122.52, 70.41, 33.06, 21.05 ppm; HRMS (FAB+) m/z: calculated for

$C_{25}H_{20}O_3S$+H: 401.1211; measured: 401.1213.
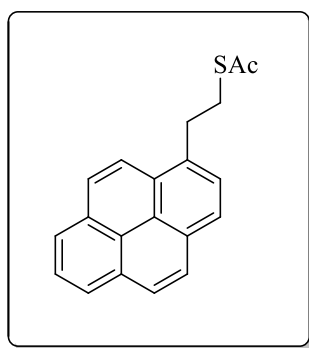
S-(2-(pyren-1-yl)ethyl) ethanethioate (4.3). Compound **4.2** (55 mg, 0.138 mmol) was

dissolved in 1.5 mL DMF followed by the addition of potassium thioacetate (24 mg,

0.206 mmol). The resulting mixture was stirred for 16 h at room temperature. Brine (10 mL) was added into the reaction mixture and organic compounds were extracted with dichloromethane (2×10 mL). Combined organic extract was dried over $MgSO_4$, filtered and evaporated to dryness in rotary evaporator. Product was purified through silica gel column by flash chromatography using a gradient of ethyl acetate (0 - 5% for 3 h) in hexane. Final product **4.3** was obtained as white solid (32 mg, 78%). ^{1}H NMR (500 MHz, CDCl$_3$): δ 8.42 (d, J = 9.5 Hz, 1H, Ar$\underline{H}$), 8.15-8.20 (m, 3H, Ar$\underline{H}$), 8.12 (d, J = 8.0 Hz, 1H, Ar$\underline{H}$), 7.99-8.04 (m, 3H, Ar$\underline{H}$), 7.89 (d, J = 8 Hz, 1H, Ar$\underline{H}$), 3.60 (t, J = 8.0 Hz, 2H, C$\underline{H}_2$CH$_2$S), 3.32 (t, J = 8.0 Hz, 2H, CH$_2$C$\underline{H}_2$S), 2.41 ppm (s, 3H, C$\underline{H}_3$); ^{13}C NMR (125 MHz, CDCl$_3$): δ 196.17, 134.14, 131.49, 131.00, 130.47, 129.01, 127.86, 127.56, 127.10, 126.02, 125.18, 125.17, 125.06, 125.03, 124.96, 123.30, 33.95, 30.87 ppm (two carbons were not identified); HRMS (FAB+) m/z: calculated for $C_{20}H_{16}OS+H$: 305.1000; measured: 305.1001.
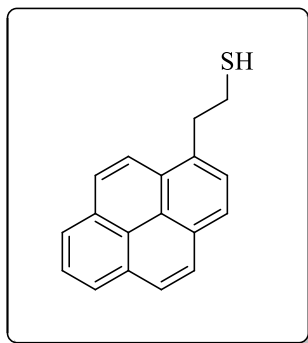
1-(2-Mercaptoethyl)pyrene (Py). Pyrrolidine (2 μL, 24.6 μmol) was added into a solution of **4.3** (5 mg, 16.4 μmol) in ethanol (1 mL) and stirred for 30 min at room temperature. Solvent was evaporated to dryness by rotary evaporator to obtain **Py** (4.3 mg, 100%). R_f on TLC: 0.18 (9:1 hexane/ethyl acetate). HRMS (APCI+): found m/z 263.0886; calculated for $C_{18}H_{14}S+H$: 263.0894.

CHAPTER 3

ASSOCIATION STUDY BY NMR TITRATION

3.1. Hydrogen Bonding Association Study

We have determined associations of the hydrogen-bonding moieties of the reading molecules with DNA bases in an aprotic solvent (deuterated chloroform) by NMR titration. For doing so, the mercaptoethyl chains or thiol of these reading molecules were replaced with more lipophilic groups and designated as **dICA**, **dBCA**, **dTCA**, and **dPCA** (Figure 3.1). The synthesis of these modified readers were discussed later in this chapter. Four naturally occurring DNA nucleosides (designated as dA, dC, dG and dT) were protected on their hydroxyls with *tert*-butyldimethylsilyl (TBDMS) group to render them soluble in chloroform (Figure 3.2). Synthesis of these modified nucleosides discussed in literature.[92] Because of their limited solubility in chloroform, the modified reader molecules were used as substrates for the NMR titration, where their concentrations held constant with minimum self-association.
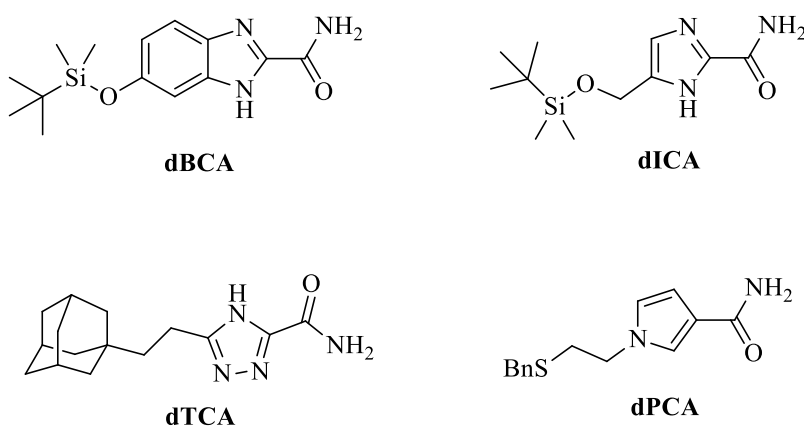
Figure 3.1. List of modified universal readers for hydrogen bonding association study by NMR titration
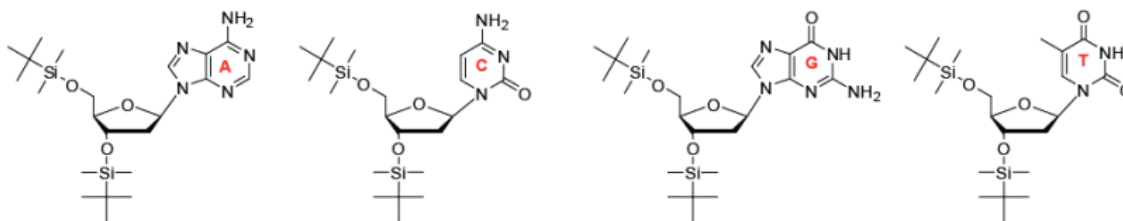
Figure 3.2. List of modified nucleosides where 3' and 5' hydroxyl groups are protected by TBDMS

3.2. Determination of Association Constants by Studying Chemical Shift Change

Previously, we determined the association constants of **dICA** with DNA bases by monitoring changes in chemical shifts of the amide protons (see Table 3.1).[40] In the same manner, we determined association constants of **dTCA** and **dPCA** interacting with DNA bases by monitoring changes in chemical shifts of the amide protons. Figure 3.3 shows

dTCA = 1.0 mM Amide-H	Concentration of dT (mM)	Chemical Shift (ppm)
	62.5	5.924
	51.3	5.891
	39.5	5.845
	27.0	5.783
	20.6	5.747
	13.9	5.706
	7.0	5.647
	0.0	5.579

6.0   5.9   5.8   5.7   5.6   5.5 ppm
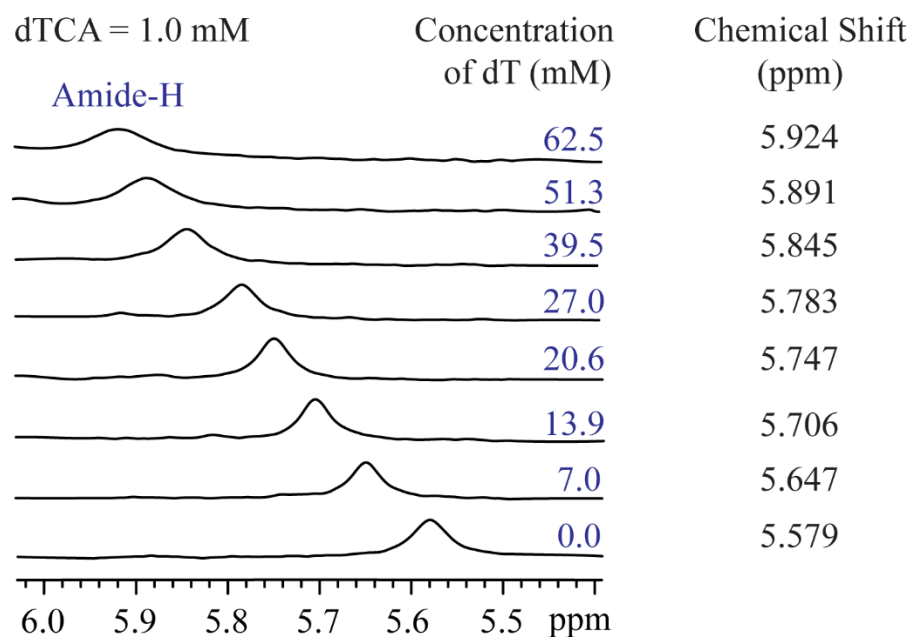
Figure 3.3. NMR titration spectra of **dTCA** amide proton in different concentrations of dT (0.0 to 62.5 mM) and corresponding chemical shift values at 297 K

changes in amide chemical shift of **dTCA** when titrated with dT. The amide proton of a 1.0 mM **dTCA** solution was observed at 5.579 ppm in the absence of any dT titrator. The amide proton peak was detected very clearly throughout the titration range. A known volume of dT solution was added into the NMR tube containing **dTCA** and the chemical shift was recorded after each addition. The peaks shifted to the left with the increase of dT concentration. The observed chemical shift was 5.647 ppm when 7.0 mM dT was added with 0.068 ppm chemical shift change. When dT concentration was 50 times higher, the chemical shift change was 0.3 ppm. Similarly all the chemical shifts values were observed and reported corresponding to different dT concentrations (see Figure 3.3).

In chloroform, **dBCA** ring N-H proton clearly shows up two peaks for two tautomer in its ^{1}H NMR spectrum. Figure 3.4 shows changes in chemical shift of the N-H proton with concentrations of dT. The chemical shifts of both amide and ring N-H protons moved to lower fields with increase in concentrations of the titrator, implying that these protons can hydrogen bond with DNA bases. The ring N-H proton or two tautomer of a 5.15 mM **dBCA** solution was observed at 10.348 ppm and 10.220 ppm in the absence of any dT titrator. These ring N-H proton peaks were detected very clearly throughout the titration range. A known volume of dT solution was added into the NMR tube containing **dBCA** and the chemical shift was recorded after each addition. Both peaks shifted to the left with the increase of dT concentration. The observed chemical shift was 10.348 ppm and 10.220 ppm for NH(1) and NH(2) respectively when 13.1 mM dT was added with 0.260 and 0.203 ppm chemical shift change. When dT concentration was 50 times higher, the chemical shift change was 0.80 and 0.69 ppm. Similarly all the

47

chemical shifts values were shown corresponding to different dT concentrations (Figure 3.4).

dBCA = 5.15 mM Ring N-H	Concentration of dT (mM)	Chemical Shift (ppm) NH(2)	NH(1)
	192.5	11.906	11.647
	151.4	11.787	11.526
	114.3	11.634	11.373
	72.8	11.369	11.115
	50.0	11.152	10.910
	25.8	10.828	10.615
	13.1	10.608	10.423
	0.0	10.348	10.220
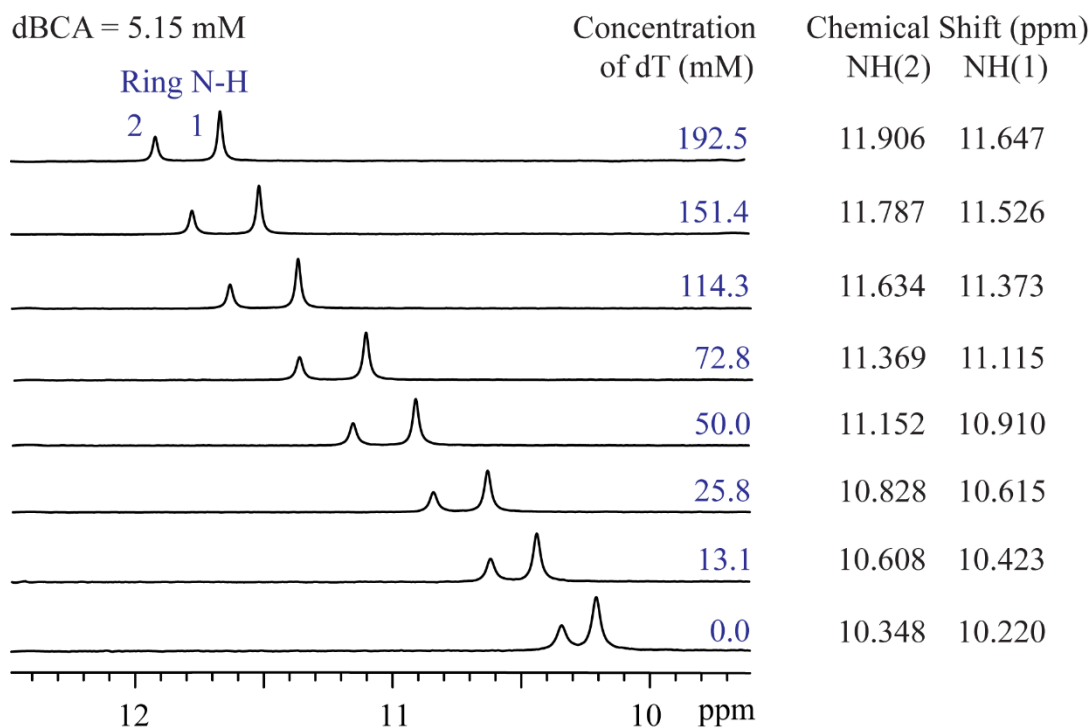
Figure 3.4. NMR titration spectra of **dBCA** ring N-H proton in different concentrations of dT (0.0 to 192.5 mM) and corresponding chemical shift values at 297 K

Similarly, when **dBCA** was titrated against dA in CDCl$_3$, the ring N-H proton of two tautomer of a 5.15 mM **dBCA** solution was observed at 10.378 ppm and 10.243 ppm in the absence of any dA titrator. These ring N-H proton peaks were detected very clearly throughout the titration range. A known volume of dA solution was added into the NMR tube containing **dBCA** and the chemical shift was recorded after each addition. Both peaks shifted to the left with the increase of dA concentration. The observed chemical shift was 10.805 ppm and 10.559 ppm for NH(2) and NH(1) respectively when 10.5 mM dA was added with 0.427 and 0.316 ppm chemical shift change. When dA concentration

was 40 mM, the chemical shift change was 1.514 and 1.225 ppm. Similarly all the chemical shifts values were shown corresponding to different dA concentrations (Figure 3.5).

dBCA = 5.15 mM	Concentration of dA (mM)	Chemical Shift (ppm)	
Ring N-H		NH(2)	NH(1)
153.9	13.102	12.826
121.1	13.048	12.615
91.4	12.797	12.346
58.2	12.340	11.886
40.0	11.892	11.468
20.6	11.225	10.893
10.5	10.805	10.559
0.0	10.378	10.243
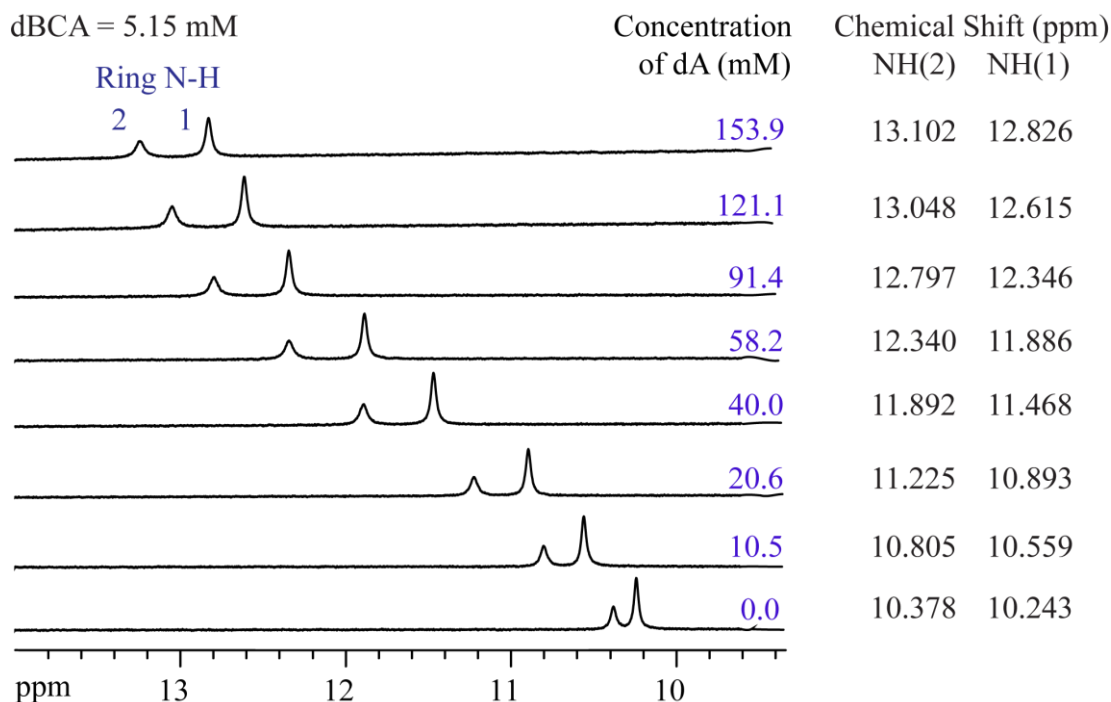
Figure 3.5. NMR titration spectra of **dBCA** ring N-H proton in different concentrations of dA (0.0 to 153.9 mM) and corresponding chemical shift values at 297 K

The titration data were analyzed using HypNMR 2008 program from which association constants deduced, expressed as $\text{Log}\beta_{11}$ for a 1:1 complex and $\text{Log}\beta_{12}$ for a 1:2 complex of substrate to titrator (Table 3.1). β_{11} is the overall or cumulative stability constant for 1:1 complex formation and β_{12} is the overall stability constant between two modified nucleosides and one reader molecule. At first glance, **dICA** and **dBCA** formed 1:1 complexes with DNA bases more stable than **dTCA** and **dPCA**. This could be explained by more loss in entropy on **dTCA** and **dPCA** interacting with DNA bases than

49

dICA and **dBCA** because **dTCA** has more tautomer forms and **dPCA** amide has more freedom to rotate. Also, these molecules form the hydrogen bonding complexes with an order of dG > dC > dT > dA. In comparison to a dA-dT base pair, **dBCA** forms more stable complexes with dA and dT, which indicates a multiplexed hydrogen bonding interaction. The NMR titration study shows that both **dBCA** and **dICA** can form 1:2 complexes with DNA bases as well. **dICA** formed 1:2 complexes with dG and dC only. Whereas **dBCA** formed 1:2 complexes with all four nucleosides dA, dG, dC and dT. Due to poor solubility of these substrates in chloroform, however, we could not determine existence of 2:1 complexes by reversed titration. Now **dBCA** showed the presence or two tautomer in deuterated chloroform solution in the experimental titration experiments. The 1:1 association constants were compared against the value obtained from origin curve fitting. It was found that the values are close to each other proving the reliability of this method. The control experiments were carried out titrating dT with dA. The titration data were then fit into the 1:1, 1:2 or 1:1/1:2 binding isotherm by nonlinear regression in the HypNMR program. It was found that the data from **dTCA** and **dPCA** were fit into the 1:1 model and those from **dBCA** were fit into the 1:1/1:2 model convergently (see Figure 3.6). Following the same procedure mentioned above, we determined the association constant of the Watson-Crick base pair dA-dT, which is close to the value reported in literature.[93] Since all these reading molecules interact with each DNA bases differently, they were used to detect the nucleobase in a nanogap by recognition tunneling.

Table 3.1. Association constants determined from curve fitting of NMR titration data in chloroform*

Substrate	Titrator	Log β_{11}	Log β_{12}	Log β_{11}'	Log β_{12}'
dICA[1]	dA	1.08 ± 0.09			
	dC	3.10 ± 0.10	5.04 ± 0.09	−	−
	dG	3.31 ± 0.02	5.49 ± 0.01		−
	dT	1.46 ± 0.05		−	−
dBCA	dA	1.53 ± 0.04	2.79 ± 0.03	1.33 ± 0.07	2.68 ± 0.10
	dC	2.54 ± 0.10	3.83 ± 0.08	2.44 ± 0.07	4.01 ± 0.05
	dG	2.70 ± 0.04	5.18 ± 0.01	2.66 ± 0.01	5.15 ± 0.05
	dT	1.81 ± 0.01	2.79 ± 0.02	1.62 ± 0.01	2.65 ± 0.04
dTCA	dA	1.15 ± 0.01	−	−	−
	dC	1.89 ± 0.04	−		
	dG	2.12 ± 0.03	−	−	−
	dT	1.20 ± 0.02	−	−	−
dPCA	dA**	0.18 ± 0.01	−	−	−
	dC**	0.73 ± 0.03	−		
	dG	2.25 ± 0.08	−	−	−
	dT**	0.32 ± 0.01	−	−	−
dA	dT	1.44 ± 0.07	−	−	−

[1] Adopted from reference.[40]

* Each number is an average of two individual experiments.

** The interactions were too weak to be accurately determined.

The dICA data was adopted from literature reference.[40]

Curve fitting by HypNMR Program

Each titration experiment was repeated at least two times and association constants were obtained by curve fitting analysis in HypNMR program. A 1:1 equilibrium model was used for **dTCA** and **dPCA** readers and 2:1 model was used for **dBCA** reader. Fitting curves were reported in Figure 3.6 below.
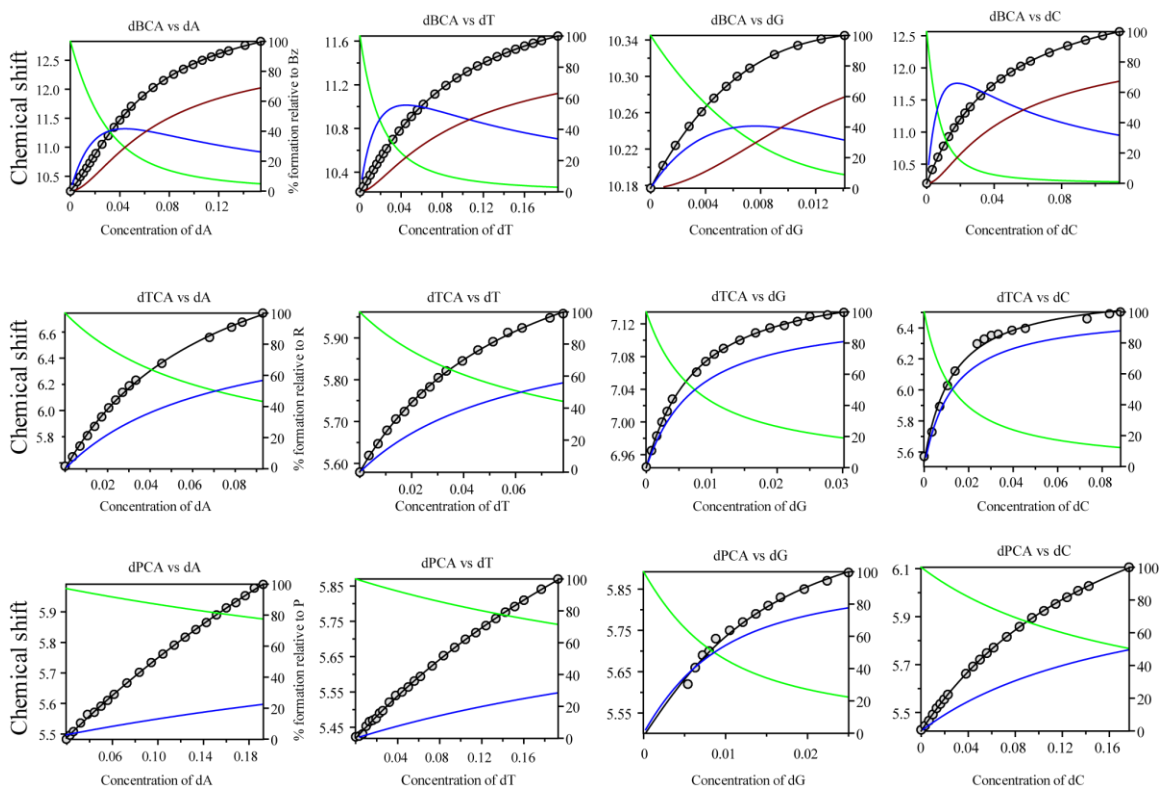
Figure 3.6. Fitting curves obtained from HypNMR program for the determination of association constants

3.3. Synthesis of Modified Universal Readers for NMR Titration Study

Universal reader **dBCA** was synthesized in three steps starting from 4-amino-3-nitrophenol as the commercially available starting material (Scheme 3.1). The hydroxyl group was protected by TBDMS to produce compound **5.1** in 95% yield. The nitro group

52

was reduced to amino group by reacting with sodium dithionite in presence of 50% aqueous ethanol to provide compound **5.2** in 94% yield. Finally cyclization reaction with 2-chloroacetamide in presence of molecular sulfur provide **dBCA** in 25% yield.
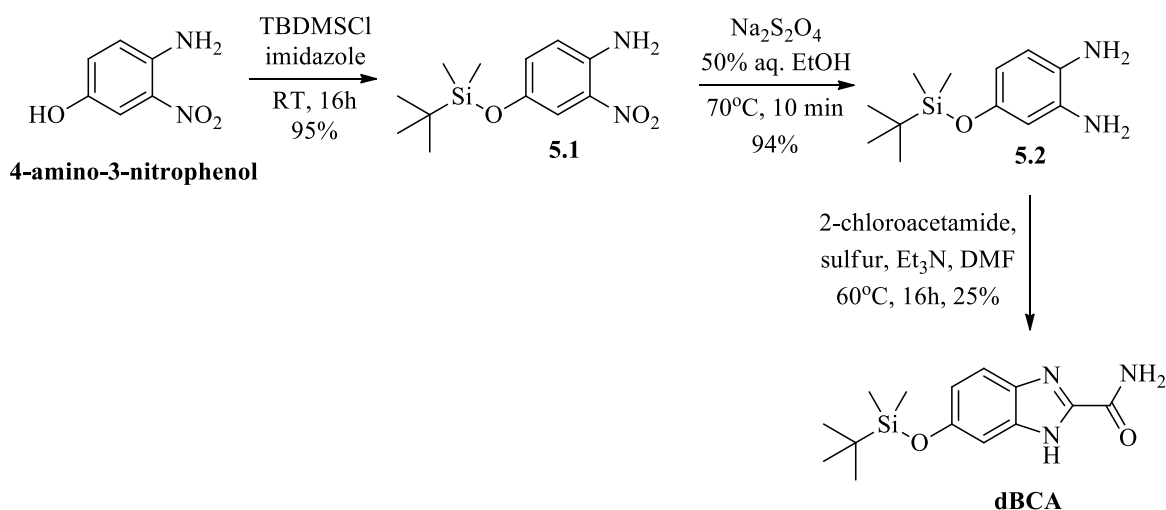
Scheme 3.1. Synthesis of 6-((*tert*-butyldimethylsilyl)oxy)-1*H*-benzo[*d*]imidazole-2-carboxamide (**dBCA**)

Universal reader **dTCA** was synthesized from 1-adamantaneethanol as the commercially available starting material (Scheme 3.2). Hydroxyl group was converted to tosyl group by tosylchloride reaction to produce compound **6.1** in 95% yield. Tosyl group was converted to cyanide by reacting with sodium cyanide to generate 6.2 in 77% yield. It was converted to salt **6.3** in 94% yield by reacting with bubbling anhydrous HCl in a solution of ethyl ether and benzyl mercaptan. Finally cyclization reaction with oxamic hydrazide in pyridine at high temperature gave **dTCA** in 60% yield.

53
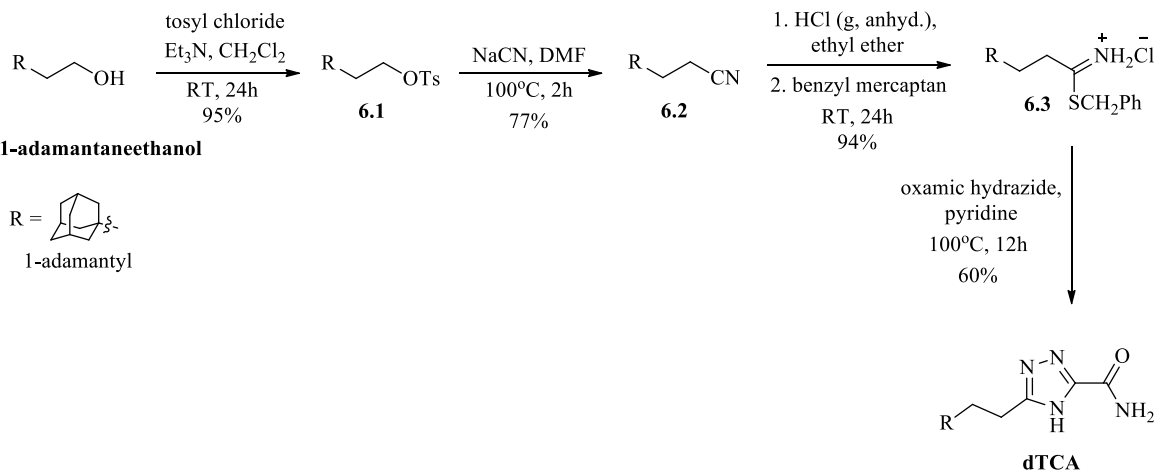
Scheme 3.2. Synthesis of 5-(2-(1-adamantyl)ethyl)-4*H*-1,2,4-triazole-3-carboxamide (**dTCA**)
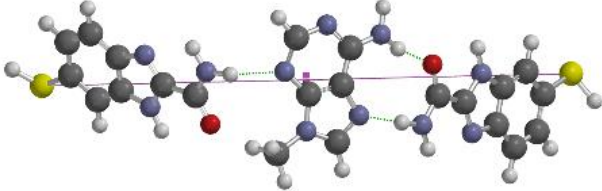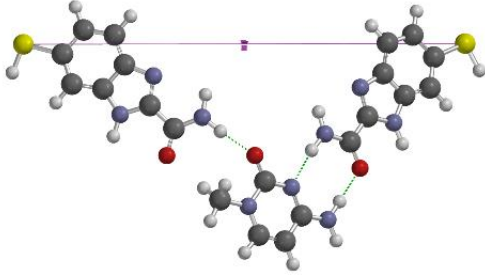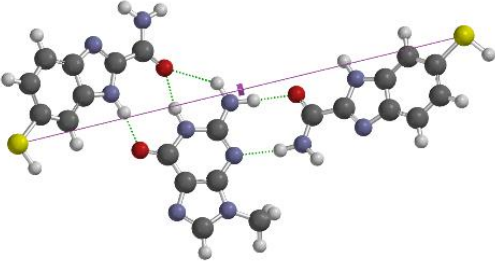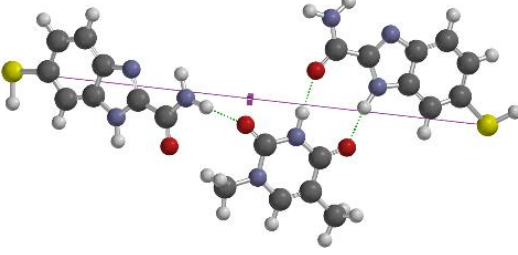
3.4. Calculation of Hydrogen Bonding Energies by DFT Calculation

Hydrogen bond strength in AT and GC base pairs (plain nucleic bases) are found

to be -12.3 kcal/mol and -25.5 kcal/mol respectively by DFT calculation using B3LYP/6-

31G** basis set (Figure 3.7).[94] We have calculated the hydrogen bonding energies of

triplet complexes between two **BCA** readers and a nucleobase. The nucleobases used

were 9-methyl purines and 1-methyl pyrimidines for simplicity of DFT calculation. Their

optimized structures are shown and H-bond energies were reported in Table 3.2. The

distances between two sulfur atoms were fixed at 2 nm for all the calculations.
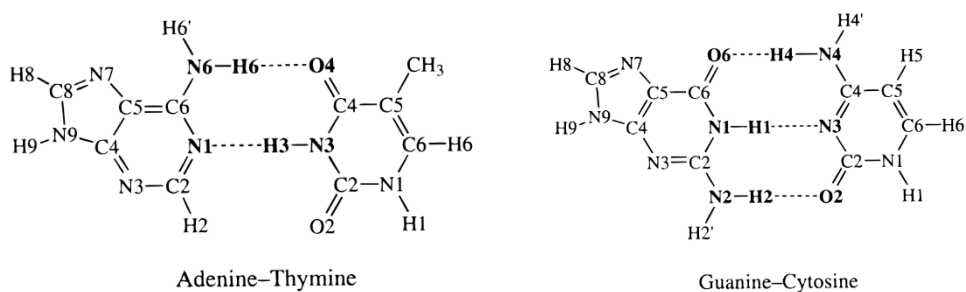
Figure 3.7. A-T and G-C base pairs showing only bases

Table 3.2. Optimized structures of Hydrogen Bonding triplet complexes with two **BCA** reader and a DNA base and their H-bonding energies obtained by DFT Calculation

DNA Base	Optimized structure (BCA-Base-BCA)	H-bond energy (Kcal/mol)
9-methyladenine (MeA)		-18.1
1-methylcytosine (MeC)		-23.4
9-methylguanine (MeG)		-29.2
1-methylthymine (MeT)		-21.2

The complex with MeG showed the lowest hydrogen bonding energy (-29.2 Kcal/mol) compared to other nucleobases implying the most stable H-bonding complex formation with G base. Similarly the formation of the triplet complexes were calculated for all four universal reader candidates with all four methylated bases and reported in Table 3.3. For all four reader candidates, MeG formed the most stable hydrogen bonding complex compared to other nucleobases.

Table 3.3. Calculated H-bonding energies (Kcal/mol) of triplet complexes in vacuum

	MeA	MeC	MeG	MeT
BCA	-18.1	-23.4	-29.2	-21.2
ICA	-17.1	-22.4	-30.9	-22.3
PCA	-12.6	-18.6	-28.3	-19.5
TCA	-19.0	-23.5	-31.2	-23.2

3.5. Experimental Procedures

Reagents and solvents were purchased from commercial suppliers (Sigma-Aldrich, Alfa Aesar, Fluka, TCI America etc. as required) and used as received unless otherwise noted. All experiments requiring anhydrous conditions were performed in flame-dried glassware under nitrogen atmosphere. Reactions were monitored by thin layer chromatography (TLC) using glass plates precoated with silica gel (EMD Chemicals Inc.). Flash chromatography was performed in an automated flash

chromatography system (CombiFlash R_f, Teledyne Isco, Inc.) with silica gel columns (60-120 mesh). ^{1}H NMR and ^{13}C NMR spectra were recorded on Varian INOVA 400 (400 MHz) and Varian INOVA 500 (500 MHz) spectrometers at 25°C at the Magnetic Resonance Research Center at Arizona State University. Chemical shifts (δ) are given in parts per million (ppm) and are referenced to the residual solvent peak (CDCl$_3$: δ_H = 7.26 ppm, CD$_3$OD: δ_H = 3.31 ppm, DMSO-d$_6$: δ_H = 2.50 ppm). Coupling constants (J) are expressed in hertz (Hz) and the values are rounded to the nearest 0.1 Hz. Splitting patterns are reported as follows: br, broad; s, singlet; d, doublet; dd, doublet of doublets; t, triplet; dt, doublet of triplets; q, quartet and m, multiplet. High resolution mass spectra (HRMS) are acquired at the Arizona State University CLAS High Resolution Mass Spectrometry Facility.

^{1}H NMR Titration Method

^{1}H NMR titration experiments were carried out using a Varian INOVA 500 spectrometer operating at 500 MHz at 25°C. Chemical shifts are reported in ppm and referenced to CDCl$_3$ residual peak (δ_H = 7.26 ppm). Deuterated chloroform (CDCl$_3$) was purchased from Spectrum Chemical MFG CORP (99.8 atom % D). It was stored over activated molecular sieve (4 Å) in a glove box (0.5 ppm moisture and 0.05 ppm oxygen). Chemicals were dried at 40°C under vacuum for two days and stored over drierite. Solutions were prepared in glove box before each NMR titration. A gas tight syringe was used for the addition of nucleoside solution into the NMR tube. The addition was done as quickly as possible to minimize moisture entering into the system. Typically 0.6 mL of ~5 mM **dBCA**, **dPCA** and ~1.0 mM **dTCA** were used in a NMR tube for the titration.

The guest solutions (modified nucleosides) were prepared to a concentration of around 800 mM (0.2 mL) for **dBCA** and **dPCA** and around 500 mM for **dTCA**. CDCl$_3$ was used as solvent to prepare these solutions. In a typical titration experiment, the titrating solution (guest) was added in 15-25 separate portions to the host solution and ^{1}H NMR was recorded of the resulting mixture after each addition. For example, modified dT nucleoside solution with a concentration of 800 mM was added to **dBCA** solution (5.15 mM) in an increasing order of 2.5, 5.0 and 10.0 µL aliquots up to a total volume of 190 µL of dA. Similarly, a modified dT solution with a concentration of 500 mM was added to a 1.0 mM solution of **dTCA** in a certain increment up to a total volume of 180 µL of dT and chemical shifts of the amide proton were recorded after each titration. Titration data were analyzed by nonlinear regression analysis using the HypNMR2008 program.

Computational Methods:

DFT calculations were performed using Spartan'14 for Windows, a commercially available software from Wavefunction, Inc. Molecules were drawn in ChemDraw Ultra 12.0 and imported to Spartan'14 to generate corresponding 3D structures as well as the hydrogen bonding complexes. Each structure was subjected to energy minimization using the built-in MMFF molecular mechanics prior to optimization calculation. The DFT calculations for hydrogen bonding complexes and individual structure optimization of reader molecules and N-methylated nucleosides were performed at their ground state equilibrium geometry conformation using B3LYP/6-31+G* basis set in vacuum.
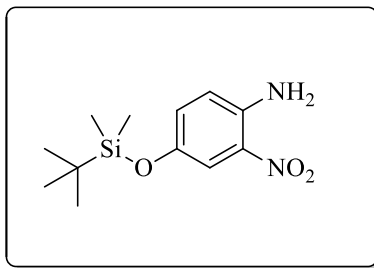
4-((*tert*-butyldimethylsilyl)oxy)-2-nitroaniline (5.1). 4-amino-3-nitrophenol (0.5g, 3.25 mmol) was added in anhydrous dichloromethane (10 mL) at RT under inert atmosphere. Imidazole (0.66g, 9.75 mmol) and catalytic amount of *N,N*-dimethyl aminopyridine (0.08g, 0.65 mmol) were added into the solution and stirred for 30 mins. Finally *tert*-butyl dimethylsilylchloride (0.58g, 3.90 mmol) was added into the reaction mixture and stirred for 16h at RT. Upon completion, dichloromethane (50 mL) was added into the reaction mixture and washed with saturated sodium bicarbonate solution (30 mL×2). The organic layer was further washed with brine (30 mL), dried over $MgSO_4$, and concentrated in rotary evaporator. The residue was purified by flash column chromatography to obtain the pure product (0.83g, 95%) as a red solid **5.1**. ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.56 (1H, d, J = 3.0 Hz, Ar<u>H</u>), 6.98 (1H, m, Ar<u>H</u>), 6.72 (1H, d, J = 9.0 Hz, Ar<u>H</u>), 5.80 (2H, broad, N<u>H</u>$_2$), 0.98 (9H, m, *tert*-butyl), 0.19 ppm (6H, m, SiC<u>H</u>$_3$); ^{13}C NMR (115 MHz, $CDCl_3$): δ = 146.7, 139.8, 130.4, 120.1, 115.2, 26.0, 18.5, -4.21 ppm.
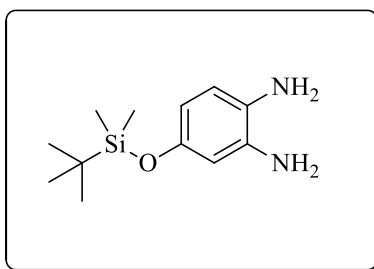
4-((*tert*-butyldimethylsilyl)oxy)benzene-1,2-diamine (5.2). Compound 5.**1** (0.4 g, 1.49

mmol) was dissolved in 50% aqueous ethanol (20 ml) and then sodium dithionite (1.56 g,

8.96 mmol) was added portion wise over a period of 20 min. The stirred solution was

gradually heated to 70°C and refluxed for about 10 min while the red solution became

colorless. Heating was stopped when the red color disappeared. It was cooled to room

temperature and the solvents were evaporated to dryness using rotary evaporator. Water

(20 mL) was added to the crude and extracted with dichloromethane (30 mL×3). The

combined organic layers were washed with brine (30 mL), dried over MgSO$_4$, and

concentrated in rotary evaporator. The residue was purified by flash column

chromatography to obtain the pure product **5.2** (0.33g, 94%) as brown solid. ^{1}H NMR

(400 MHz, CDCl$_3$): δ = 6.55 (1H, d, *J* = 8.4, Ar$\underline{H}$), 6.24 (1H, d, *J* = 2.4, Ar$\underline{H}$), 6.19 (1H,

m, Ar$\underline{H}$), 3.23 (4H, s, broad, N$\underline{H}_2$), 0.96 (9H, m, *tert*-butyl), 0.15 ppm (6H, m, SiC$\underline{H}_3$).
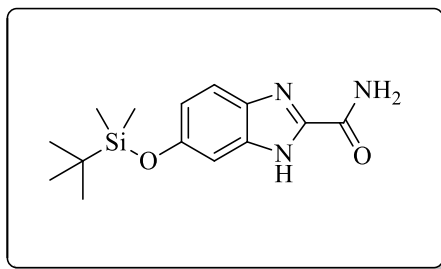
6-((*tert*-butyldimethylsilyl)oxy)-1*H*-benzo[*d*]imidazole-2-carboxamide (dBCA).

2-Chloroacetamide (0.47 g, 5.0 mmol) was added to a prepared mixture of **5.2** (1.0 g, 4.2

mmol), sulfur (0.54 g, 16.8 mmol) and triethyl amine (1.0 mL) in DMF (10 mL). The

reaction mixture was stirred at 60°C for 16 h, cooled to room temperature, diluted with

water (30 mL) and extracted with ethyl acetate (3 × 30 mL). The combined organic

extracts were washed with brine three times (each 20 mL), dried over MgSO$_4$, filtered

and concentrated by rotary evaporator. The residue was separated in a silica gel column

by flash chromatography with a gradient of methanol (0-5% in 4 h) in dichloromethane to obtain product **dBCA** (0.3 g, 25%). ^{1}H NMR (500 MHz, DMSO-d$_6$): δ = 12.90 (1H, s, N$\underline{H}$), 8.13 (1H, s, broad, N$\underline{H}_2$), 7.70 (1H, s, broad, N$\underline{H}_2$), 7.56 (1H, d, J = 9.0, Ar$\underline{H}$), 6.88 (1H, d, J = 2.0; Ar$\underline{H}$), 6.79 (1H, m, Ar$\underline{H}$),), 0.96 (9H, m, *tert*-butyl), 0.19 ppm (6H, m, SiC$\underline{H}_3$); HRMS (APCI+): *m/z* calculated for C$_{14}$H$_{21}$N$_3$O$_2$Si+H: 292.1481; measured: 292.1478.
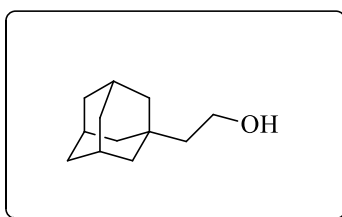
Synthesis of 2-(adamantan-1-yl)ethyl 4-methylbenzenesulfonate (6.1). 2-(adamantan-1-yl)ethanol (2.0 g, 11.11 mmol) was added into dichloromethane (25.0 mL) under inert atmosphere. Triethyl amine (7.7 mL, 55.55 mol) was added into it and stirred for 30 mins at RT. Tosyl chloride (2.53 g, 13.33 mol) was added portion wise into the solution and stirred for 24 h at RT. Upon completion the reaction mixture was washed with saturated sodium bicarbonate solution (20 mL×2) and brine (20 mL). The combined organic extracts were dried over magnesium sulfate, filtered and concentrated by rotary evaporator. The product was purified by silica gel flash column chromatography. Pure product **6.1** (3.54 g, 95%) was obtained as a colorless liquid. ^{1}H NMR (500 MHz, CDCl$_3$): δ = 7.77 (2H, d, *J* = 8.0 Hz, ArH), 7.33 (2H, d, *J* = 8.0 Hz, ArH), 4.08 (2H, t, *J* = 7.5, CH$_2$), 2.44 (3H, s, CH$_3$), 1.90 (3H, s, broad, CH of Adamantane), 1.55-1.67 (6H, m, CH$_2$ of Adamantane), 1.45 (2H, m, CH$_2$), 1.42 ppm (6H, m, CH$_2$ of Adamantane); ^{13}C

NMR (125 MHz, CDCl$_3$): δ = 144.9, 133.6, 130.1, 128.2, 67.6, 42.8, 42.6, 37.1, 32.0, 28.7, 21.9 ppm.
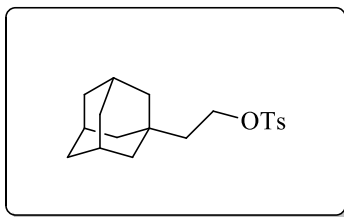
3-(adamantan-1-yl)propanenitrile (6.2). Compound **6.1** (2.0 g, 5.99 mmol) was added in anhydrous DMF (10 mL) under inert atmosphere at RT. Sodium cyanide (0.35 g, 7.18 mmol) was added into the solution and refluxed at 100 ^{0}C for 2h. The reaction mixture was cooled at RT and the solvent was evaporated to dryness by rotary evaporator to obtain pale yellow crude liquid. Water (25 mL) was added into the crude mixture and extracted with ethyl acetate (25 mL×3). The combined organic layers were washed with brine (20 mL), dried over magnesium sulfate, filtered and concentrated by rotary evaporator. The product was purified by silica gel flash column chromatography. Pure product **6.2** (0.87 g, 77%) was obtained as a colorless liquid. ^{1}H NMR (400 MHz, CDCl$_3$): δ = 2.25-2.29 (2H, m, CH$_2$), 1.98 (3H, s, broad, CH of Adamantane), 1.60-1.74 (6H, m, CH$_2$ of Adamantane), 1.50 (2H, m, CH$_2$), 1.47 ppm (6H, m, CH$_2$ of Adamantane); ^{13}C NMR (125 MHz, CDCl$_3$): δ = 121.3, 42.0, 39.8, 37.2, 32.4, 28.7, 11.3.
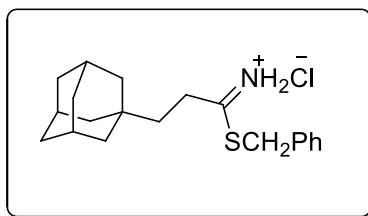
benzyl 3-(1-adamantyl)propanimidothioate hydrochloride (6.3). Hydrogen chloride was bubbled into a solution of **6.2** (0.2 g, 1.06 mmol) and benzyl mercaptan (0.2 mL,

1.71 mmol) in anhydrous ethyl ether (12 mL) in an ice bath under nitrogen for 2 h. The solution was capped and allowed to warm to room temperature, stirred for another 24 h, and then stood still for 2 h. The compound **6.3** was crystallized from the reaction solution. It was filtered through a Buchner funnel, washed with cold ethyl ether three times (each 5 mL), and dried in vacuum at 25°C overnight. It weighed 0.3 g (yield: 81%). ^{1}H NMR (400 MHz, CDCl$_3$): δ = 13.16 (s, br, 1H, N$\underline{H}_2$), 12.22 (s, br, 1H, N$\underline{H}_2$), 7.31-7.39 (m, 5H, Ar$\underline{H}$), 4.73 (s, 2H, SC$\underline{H}_2$Ph), 2.87 (m, 2H, CH$_2$C$\underline{H}_2$CS), 1.95 (s, br, 3H, adamantane), 1.49-1.68 ppm (m, 14H, adamantane + C$\underline{H}_2$CH$_2$CS); ^{13}C NMR (100 MHz, CDCl$_3$): δ = 196.0, 132.1, 129.8, 129.4, 129.0, 43.6, 42.1, 38.4, 37.1, 33.1, 32.5, 28.8 ppm; HRMS (APCI+): m/z calculated for C$_{20}$H$_{27}$NS+H: 314.1942; measured: 314.1952.
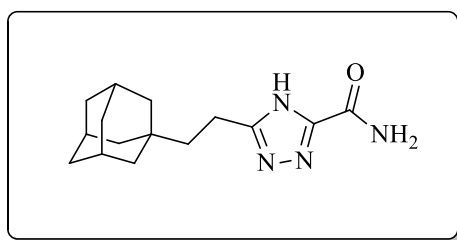
5-(2-(1-adamantyl)ethyl)-4H-1,2,4-triazole-3-carboxamide (dTCA). A solution of **6.3** (0.21 g, 0.6 mmol) and oxamic hydrazide (62 mg, 0.6 mmol) in anhydrous pyridine (2 mL) was refluxed at 100°C for 12 h. The solvent was removed by co-evaporating with toluene (1 mL × 2) using rotary evaporator. The yellow residue was separated in a silica gel column by flash chromatography with a gradient of methanol (0-5% in 3h) in dichloromethane to obtain compound **dTCA** (0.11 g, 67%). ^{1}H NMR (400 MHz, CD$_3$OD): δ = 2.75-2.79 (2H, m, CH$_2$), 1.97 (3H, s, broad, CH of adamantane), 1.67-1.78 (6H, m, CH$_2$ of adamantane), 1.45 (6H, m, CH$_2$ of adamantane), 1.42 ppm (2H, m, CH$_2$); HRMS (APCI+): m/z calculated for C$_{15}$H$_{22}$N$_4$O+H: 275.1872; measured: 275.1880.

CHAPTER 4

NONCOVALENT INTERACTION STUDY BY ELECTROSPRAY MASS

SPECTROMETRY (ESIMS) AND TANDEM MASS SPECTROMETRY (ESIMS/MS)

4.1. General Introduction

The ability of complexation in organic solvent has been studied extensively by

NMR spectroscopy. Recognition tunneling experiment was performed in phosphate

buffer solution. In order to prove the presence of non-covalent complexes in aqueous

solution, electrospray ionization mass spectrometry have been employed. Electrospray is

a soft ionization technique and proved to retain the identity of a non-covalent complex in

the gas phase after ionization. This technique has been extremely helpful and universally

applied to detect and identify molecular complexes. Electrospray has been used to study

weak noncovalent interactions,[67] for example hydrogen bonding,[95] π stacking,[96] small

molecular interactions,[97,98] supramolecular interactions,[99,100] and noncovalent

protein/DNA complexes.[101,102] It is a soft ionization method[103] and it can preserve non-

covalent complex from its native solution state into the gas phase in the form of single or

multiple charged ions.[104] It has also been used for screening carbohydrate libraries of

protein interaction,[105,106] and quantification of protein ligand complexes.[107,108]

4.2. Electrospray MS of ICA-Amino acid Solutions

In our study we measured the non-covalent interactions between **ICA** reader

molecule and amino acids to form stoichiometric complexes in aqueous solution. We

began with recording ESI mass spectra of individual amino acids and **ICA** in aqueous

solution and assigned their observed characteristic m/z peaks (Table 4.1). The ESI mass

data show that **ICA** and individual amino acids can exist in both monomer and self-associated adducts forms including dimers and trimers. Based on the measured pH of each solution, we have deduced the major form of the analyte occurring in the solution (see Table 4.1) according to its reported pKa. Under our experimental conditions, the **ICA** molecule stayed neutral.[40] Sequentially we mixed **ICA** with seven amino acids respectively in 1:1 and 2:1 ratios, and measured their ESI mass spectra. Examples of spectra from an **ICA**-Leu mixture and the corresponding **ICA** and Leu solutions are given in Figure 4.1. By comparing these spectra, we can clearly see two new m/z peaks
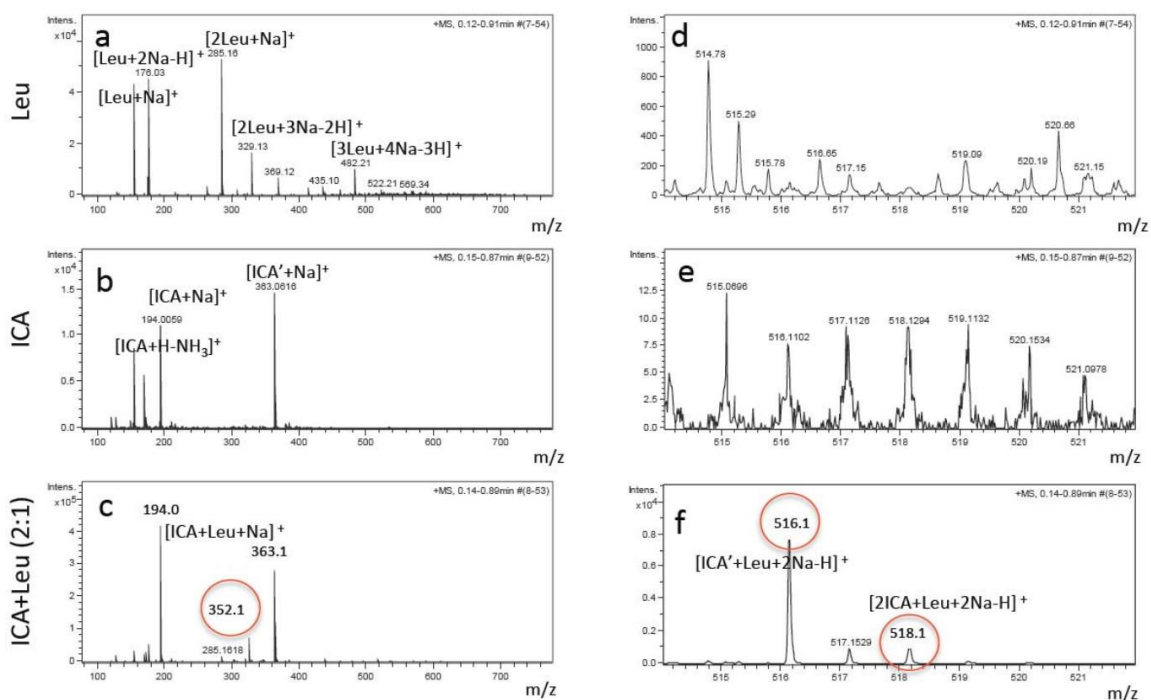
Figure 4.1. Example ES-MS spectra of pure compounds and complexes. **(a)** Leucine, **(b) ICA**, **(c) ICA** + Leucine at 2:1 ration. **(d)**, **(e)** & **(f)** show spectra at higher resolution.

appearing in the spectra of the **ICA**-Leu mixture, which correspond to their 1:1 and 2:1 adducts. We confirm the complexes by tandem mass spectrometry (ESIMS/MS,

examples are given in Figure 4.2). It should be noted that an **ICA** disulfide species

(**ICA'**) existed in all of the measured solutions due to oxidation of the thiol during the

sample handling process although we tried to avoid the oxidation using the argon

sparging. **ICA'** also formed the 1:1 adducts with amino acids. However, the **ICA'**

adducts can readily be distinguished from the **ICA** adducts by ESIMS/MS. The

ESIMS/MS data of **ICA**-amino acid mixture are given in Table 4.2. Only are those
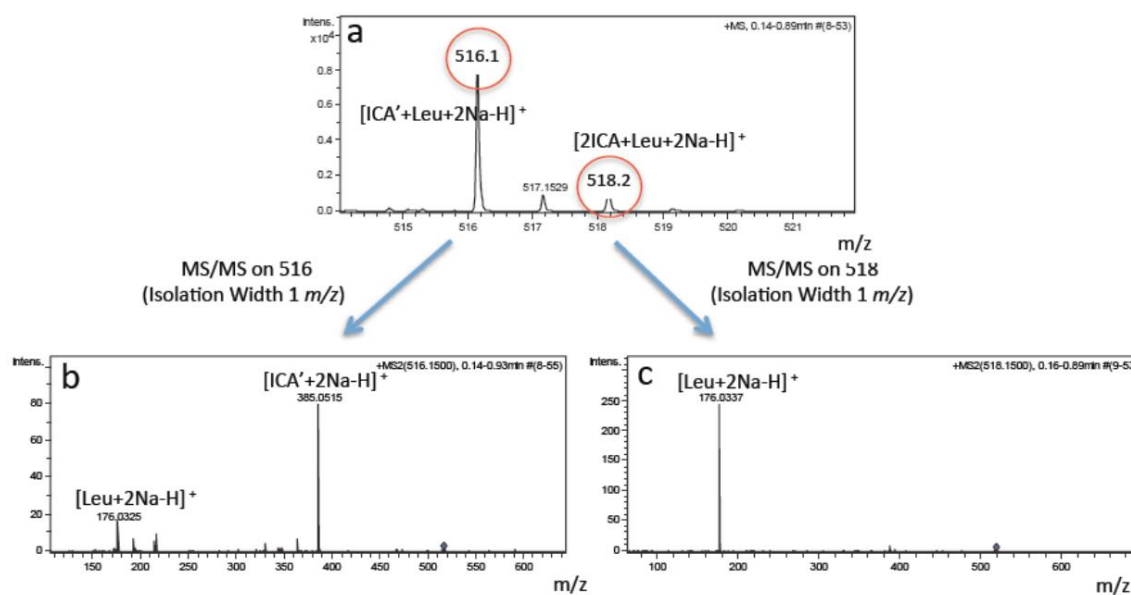
Figure 4.2. Examples of MS-MS spectra. Two peaks are found in 2:1 mixtures of **ICA** with Leucine, circled in **(a)**. MS-MS shows that the peak at 516 Daltons a complex of an oxidized **ICA** (**ICA'**) in which two **ICA** molecules are joined by a disulfide linkage **(b)**. The peak at 518 Daltons is shown **(c)** to consist of two non-oxidized ICA molecules with one Leucine.

ESIMS peaks that were confirmed by tandem mass spectrometry listed in the table.

Further analysis indicates that abundance of 1:1 and 2:1 adducts increases with **ICA**

concentrations (data not shown). In addition, we also observed the m/z peaks

corresponding to 1:3 **ICA**-amino acid adducts in ESIMS of these mixtures, which need to

be further confirmed. In summary, the ESIMS/MS studies show that **ICA** can form 1:1 and 2:1 complexes with amino acids in aqueous solutions.

4.3. Electrospray MS of ICA-Carbohydrate Solutions

ESI can preserve non-covalent complexes in going from their native solution state into the gas phase in the form of single or multiple charged ions.[104] Carbohydrates, particularly those attached to proteins, play a central role as mediators in most biological processes. Examples are protein folding,[109] cell adhesion[110] and signaling,[111] fertilization[112] and embryogenesis,[113] pathogen recognition[114] and immune responses.[115] Carbohydrate structures are complicated by isomerism. Epimers, anomers and regioisomers contribute to structural variability. For example 1.05×10^{12} structures are possible to a hexasaccharide,[116,117] a complexity that challenges even the most sophisticated analytical tools such as NMR and mass spectrometry. NMR is capable of determining carbohydrate structures, but it requires milligrams of sample, long data acquisition times (hours or even days), and cannot distinguish small amount of coexisting isomers.[117] Since many carbohydrates share a molecular weight, mass spectrometry is unable to identify them without additional chemical steps.[118] The problem has recently been addressed by combining ion-mobility spectrometry, which used collision cross-sections to separate isomers, with mass spectrometry (IM-MS),[119] but IM-MS is not inherently quantitative and cannot resolve some closely related epimers with almost identical collision cross-sections. We employed recognition tunneling approach to identify isomers of carbohydrate molecules using **ICA** as trapping reader molecule. The formation of complexes between two **ICA** readers and a carbohydrate molecule has been

studied by electrospray ionization mass spectrometry. In our present study we observed stoichiometric complexes between **ICA** and carbohydrates. By means of ESI-MS, we first obtained characteristic m/z peaks for each individual carbohydrate and **ICA** in aqueous solution, and then measured the aqueous solutions of **ICA** mixed with each carbohydrate in 2:1 ratio. Data showed that all monosaccharides and **ICA** existed as self-associated dimers in solution along with their monomeric form (Table 4.3), **ICA** formed 1:1 molecular complex ions with all carbohydrates (Table 4.4), and 2:1 molecular complex ions with most of the carbohydrates (Table 4.5). Each molecular ion complex was confirmed by tandem mass spectrometry (MS/MS) and reported as MS/MS product ion in the respective columns. We purchased those materials from commercial sources and confirmed their configurations by NOESY NMR before use (see Figure 4.3). In case of the α-anomer, a NOE cross peak between H_1 and H_2 was observed as expected (Figure 4.3, a). For the β-anomer, two NOE cross peaks were observed between H_1 and H_3 as well as between H_1 and H_5 protons (Figure 4.3, b).[120]
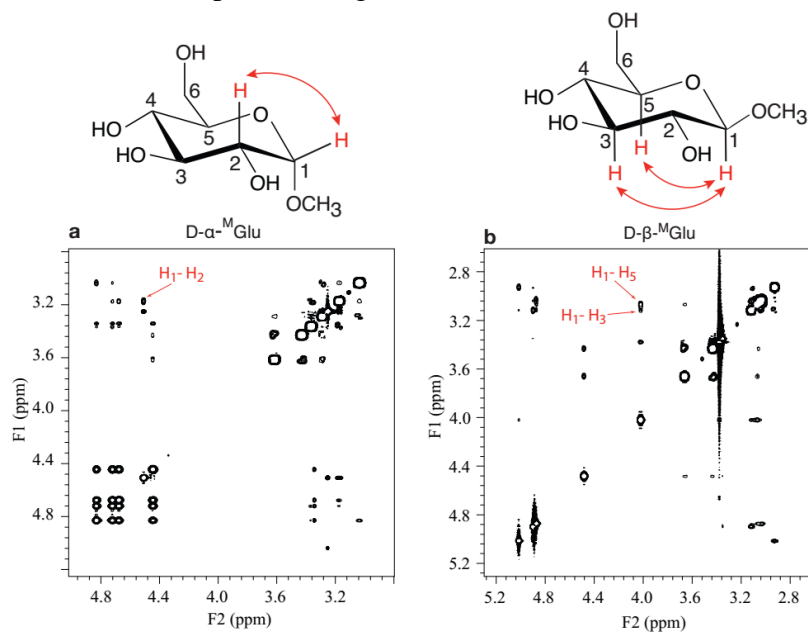
Figure 4.3. ^{1}H-^{1}H NOESY NMR spectra of (a) α-^{M}Glu and (b) β-^{M}Glu

4.4. Electrospray MS of Universal Readers-DNA Nucleotides Solutions

Single MS data was measured for the individual universal readers and DNA nucleotides in aqueous solution. A 200 µM solution of each reader molecule and a 100 µM solution of each DNA nucleotide monophosphate has been used for their individual MS spectra. Data showed that each reader and nucleotide can exist as self-associated dimers along with monomers (Table 4.6 and Table 4.7). An example spectra provided below between **BCA** and **AMP** showed the formation of 1:1 complex (Figure 4.4, a). Tandem mass spectra showed the formation of product ions corresponds to the reader and/or nucleotide ion. Tandem mass was taken for both complexes to prove their identity. Here MS/MS on 569 peak was shown (figure 4.4, b). After bombarding the 1:1 complex into their product ions, two product ion peaks were observed in the spectrum. One was monosodium form of **BCA** and disodium form of AMP (base peak).
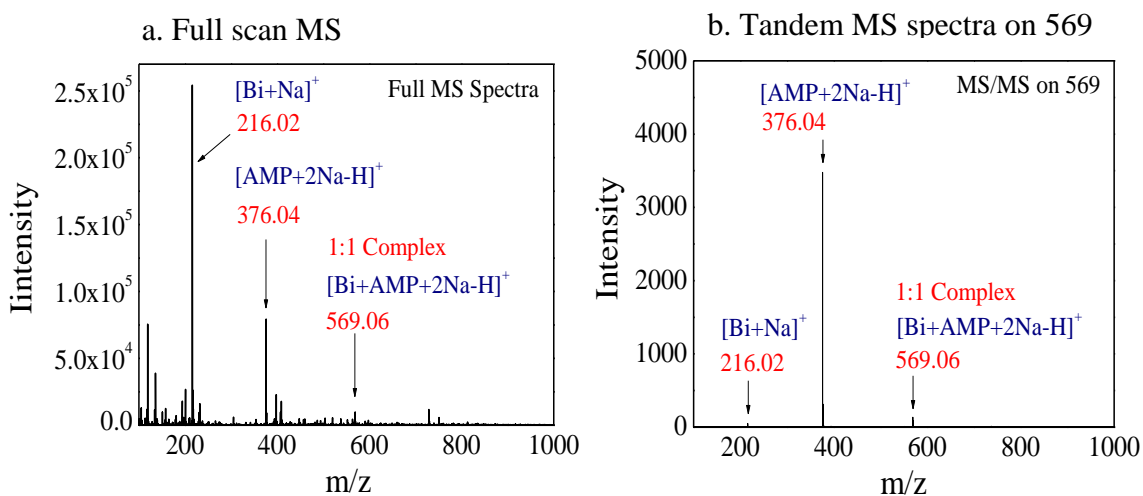
Figure 4.4. (a) A full scan single MS spectra of a solution of 2:1 mixture of **BCA** and **AMP** was shown. Individual **BCA** peak was observed at m/z 216.02 for $[BCA+Na]^+$ ion. AMP peak was observed at m/z 376.04 for $[AMP+2Na-H]^+$ ion. 1:1 complex was observed at m/z 569.06 for $[BCA+AMP+2Na-H]^+$ ion. **(b)** Tandem MS spectra of 1:1 complex ion peak showed the product ions at m/z 376.04 and 216.02 for $[BCA+Na]^+$ and $[AMP+2Na-H]^+$ ion.

69

The formation of 2:1 molecular complex ions between two **BCA** molecule and an AMP molecule were also observed (Table 4.8). A list of all characteristic MS peaks of 1:1 and 2:1 **BCA**-Nucleotide complexes and their MS/MS product ions were reported in table 4.8. The formation of complexes in aqueous solution were observed for all universal reader candidates with all four nucleotides. The observed m/z molecular complex ions and their MS/MS product ions of all four nucleotides with **ICA**, **TCA** and **PCA** were reported in Table 4.9, Table 4.10 and 4.11 respectively.

4.5. Experimental Details

      ICA (200 μM) and carbohydrate (100 μM each) solutions were respectively prepared in water and sparged with argon. Each sample solution was injected into a Bruker maXis 4G electrospray ionization quadrupole time-of-flight (ESI-Q-TOF) mass spectrometer at a 3 μL/min infusion rate via syringe pump. Tandem (MS/MS) mass spectrometry was used to observe product ion peaks from molecular complex ion peaks to confirm the composition of the molecular complex. The ESI source was equipped with a microflow nebulizer needle operated in a positive ion mode. The spray needle was held at ground and the inlet capillary set to -4500 V. The end plate offset was set to -500 V. The nebulizer gas and dry gas (N_2) were set to 1.2 Bar and 1.5 L/min, respectively, and the dry gas was heated to 220°C. In TOF-only mode the quadrupole ion energy was set to 4 eV and the collision energy was set to 1 eV. Collision gas (Ar) was set to a flow rate of 20%. In most cases MS/MS experiments were conducted with a precursor ion isolation width of 3 *m/z* units. However, if other ions were present in this range precursor ion isolation width was set to 1 *m/z* unit. Collision energy was set to 10-20 eV, which was

sufficient to fragment noncovalent complexes. Each spectrum was recorded over a time period of 0.5 to 1 min. Typically a spectrum acquired for one minute is an accumulation of 60 separate recorded mass spectra averaged across 1 min time period. Signal to noise ratio greater than three (S/N>3) was used to define the limit of detection. Due to the lack of an acid modifier in the infused solutions, most carbohydrates and molecular complexes were observed as single or multiply sodium ions [M+nNa-(n-1)H]+ rather than as protonated molecular form [M + H]+. Average mass accuracy was within 0.025 Da. The parameters were same for mass study between universal readers and DNA nucleoside monophosphates. Individual readers were prepared in 200 µM solutions and individual analytes were prepared in 100 µM solutions for their individual mass data. For preparing mixtures, 200 µM solution of a reader was mixed with a 100 µM of analyte solution (DNA, amino acid or carbohydrate) to maintain 2:1 mixing ratio. The instrumental parameters were same for DNA study, but they were little different for amino acids. Amino acids and ICA were mixed in water (specific resistance: ~18 MΩ-cm; total organic carbon: ~ 4 ppb) with a concentration of 100 µM each, sparged with argon. Buffers were not used and solution pH's are listed in the Tables. We checked that the omission of buffer slats does not affect the acquisition of RT signals, by obtaining RT signals from asparagine in pure water Each mixture was infused at 1 µL/min via syringe pump into a Bruker MicrOTOF-Q electrospray ionization quadrupole time-of-flight (ESI-Q-TOF) mass spectrometer. The ESI source was equipped with a microflow nebulizer needle operated in positive ion mode. The spray needle was held at ground and the inlet capillary set to -4100 V. The end plate offset was set to -500 V. The nebulizer gas and dry gas (N_2) were set to 0.4 Bar and 1.2 L/min, respectively, and the dry gas was heated

to 180 °C. In TOF-only mode the quadrupole ion energy was set to 8 eV and the collision energy was set to 10 eV. Collision gas (Ar) was set to a flow rate of 15%. In most cases MS/MS experiments were conducted with a precursor ion isolation width of 3 m/z units. However, if other ions were present in this range precursor ion isolation width was set to 1 m/z unit. Collision energy was set to 10 eV.

NOESY NMR of α-^{M}Glu and β-^{M}Glu.

300 mM stock solutions of β-^{M}Glu and α-^{M}Glu were prepared in DMSO-d_6 at room temperature under inert atmosphere. A volume of 0.75 mL solution from each stock solution was used for an individual NMR experiment. ^{1}H, COSY and NOESY were recorded in a Varian 500 MHz NMR at 25°C at the Arizona State University NMR facility lab. The NOE mixing time was set to 400 ms and a total number of scans were set to 16. Each spectrum was recorded for 3 h at 25°C. Data was plotted in VnmrJ 4.0 and exported to adobe illustrator and finally exported to word document.

Table 4.1. Structure information and MS data of Individual Amino Acids and ICA

Analyte	Calculated Monoisotopic Mass	Solution pH	Molecular form	[1] Observed *m/z*
L-Leu	131.0946	8.1		154.04, [M+Na]+, (82) 176.03, [M+2Na-H]+, (85) 285.16, [2M+Na]+, (100)
L-Ile	131.0946	8.0		154.04, [M+Na]+, (65) 176.03, [M+2Na-H]+, (100) 285.16, [2M+Na]+, (50)
L-Asn	132.0535	8.1		155.00, [M+Na]+, (100) 176.99, [M+2Na-H]+, (98) 287.09, [2M+Na]+, (23) 485.10, [3M+4Na-3H]+, (81)
D-Asn	132.0535	8.1		155.00, [M+Na]+, (34) 176.99, [M+2Na-H]+, (66) 287.07, [2M+Na]+, (11) 485.09, [3M+4Na-3H]+, (100)
L-Gly	75.0320	7.8		97.96, [M+Na]+, (39) 119.95, [M+2Na-H]+, (100) 173.02, [2M+Na]+, (70) 314.02, [3M+4Na-3H]+, (53)
***N*-MeGly**	89.0477	7.9		111.98, [M+Na]+, (100) 133.97, [M+2Na-H]+, (66) 201.06, [2M+Na]+, (45) 356.07, [3M+4Na-3H]+, (16)
L-Arg	174.1117	8.1		175.08, [M+H]+, (100) 197.07, [M+Na]+, (37) 219.05, [M+2Na-H]+, (23) 371.20, [2M+Na]+, (57)
ICA	171.0466	8.3		172.02, [M+H]+, (8) 194.01, [M+Na]+, (76) 365.07, [2M+Na]+, (25) 363.06, [ICA'+Na]+

1. The relative Intensity (%) value of observed ions are given in parentheses next to each complex ion. The most intense peaks in single stage MS spectra are defined as 100.

Table 4.2. Characteristic ESIMS of ICA-amino acids 1:1 & 2:1 mixtures and their MS/MS products*

Analyte			Observed *m/z*	MS/MS Product Ion
	Ratio	pH	Mass, adduct ion, (Intensity, S/N)	Mass, molecular ion, (Intensity)
ICA+L-Leu	1:1	7.8	325.12, [ICA+Leu+Na]$^+$, (15.3, 1703)	194.00, [ICA+Na]$^+$, (100)
	2:1	7.9	518.16, [2ICA+Leu+2Na-H]$^+$, (0.2, 80)	176.03, [Leu+2Na-H]$^+$, (100)
ICA+L-Ile	1:1	7.8	325.12, [ICA+Ile+Na]$^+$, (13.5, 1494)	194.00, [ICA+Na]$^+$, (100)
	2:1	7.9	496.18, [2ICA+Ile+Na]$^+$, (0.1, 42)	194.01, [ICA+Na]$^+$, (100)
			518.16, [2ICA+Ile+2Na-H]$^+$, (0.2, 60)	176.03, [Ile+2Na-H]$^+$, (100)
ICA+L-Asn	1:1	7.9	326.08, [ICA+L-Asn+Na]$^+$, (6.1, 800)	155.00, [L-Asn+Na]$^+$, (100) 194.00, [ICA+Na]$^+$, (5)
	2:1	8.0	497.13, [2ICA+L-Asn+Na]$^+$, (0.5, 60)	365.06, [2ICA+Na]$^+$, (100) 155.00, [L-Asn+Na]$^+$, (48)
			519.12, [2ICA+L-Asn+2Na-H]$^+$, (0.4, 42)	176.99, [L-Asn+2Na-H]$^+$, (100)
ICA+D-Asn	1:1	7.9	326.08, [ICA+D-Asn+Na]$^+$, (3.9, 691)	155.00, [D-Asn+Na]$^+$, (100) 194.01, [ICA+Na]$^+$, (5)
	2:1	8.0	497.13, [2ICA+D-Asn+Na]$^+$, (0.4, 67)	365.06, [2ICA+Na]$^+$, (100) 155.00, [D-Asn+Na]$^+$, (74)
ICA+L-Gly	1:1	8.0	269.05, [ICA+Gly+Na]$^+$, (0.2, 53)	194.01, [ICA+Na]$^+$, (100) 172.02, [ICA+H]$^+$, (24)
			291.03, [ICA+Gly+2Na-H]$^+$, (0.1, 30)	119.95, [Gly+2Na-H]$^+$, (100)
	2:1	8.1	462.10, [2ICA+Gly+2Na-H]$^+$, (0.1, 24)	119.95, [Gly+2Na-H]$^+$, (100)
ICA+N-MeGly	1:1	8.0	261.09, [ICA+N-MeGly+H]$^+$, (0.2, 23)	172.02, [ICA+H]$^+$, (100)
			283.07, [ICA+N-MeGly+Na]$^+$, (0.4, 35)	194.00, [ICA+Na]$^+$, (100)
	2:1	8.1	476.11, [2ICA+N-MeGly+2Na-H]$^+$, (0.1, 11)	133.97, [N-MeGly+2Na-H]$^+$, (100)
ICA+L-Arg	1:1	7.8	346.16, [ICA+Arg+H]$^+$, (0.2, 81)	175.09, [Arg+H]$^+$, (100)
	2:1	7.9	517.21, [2ICA+Arg+H]$^+$, (0.2, 59)	175.09, [Arg+H]$^+$, (100)
			539.19, [2ICA+Arg+Na]$^+$, (0.3, 92)	197.07, [Arg+Na]$^+$, (100)

The Relative Intensity (I%) and Signal to Noise Ratio (S/N) values are given in parentheses next to each complex ion in observed m/z column. I% values are reported in parentheses next to each complex ion in MS/MS product ion column. The most intense peak is considered as 100.

Table 4.3. MS data of individual Carbohydrates and ICA

Analytes	Calculated Monoisotopic Mass	[1]Observed m/z	Analytes	Calculated Monoisotopic Mass	Observed m/z
Galactose ($C_6H_{12}O_6$)	180.0634	203.05, [M+Na]$^+$, (74) 225.03, [M+2Na-H]$^+$, (2) 383.12, [2M+Na]$^+$, (100) 405.10, [2M+2Na-H]$^+$, (2)	α-^{M}Glu ($C_7H_{14}O_6$)	194.0790	217.07, [M+Na]$^+$, (67) 239.05, [M+2Na-H]$^+$, (2) 411.15, [2M+Na]$^+$, (100) 433.13, [2M+2Na-H]$^+$, (1)
Glucose ($C_6H_{12}O_6$)	180.0634	203.05, [M+Na]$^+$, (87) 225.03, [M+2Na-H]$^+$, (2) 383.12, [2M+Na]$^+$, (100) 405.10, [2M+2Na-H]$^+$, (3)	β-^{M}Glu ($C_7H_{14}O_6$)	194.0790	217.07, [M+Na]$^+$, (68) 239.05, [M+2Na-H]$^+$, (2) 411.15, [2M+Na]$^+$, (100) 433.13, [2M+2Na-H]$^+$, (1)
Galactosamine ($C_6H_{13}NO_5$)	179.0794	180.09, [M+H]$^+$, (100) 202.07, [M+Na]$^+$, (16) 359.17, [2M+H]$^+$, (64) 381.15, [2M+Na]$^+$, (19) 162.08, [M-H$_2$O+H]$^+$, (46)	Maltose ($C_{12}H_{22}O_{11}$)	342.1162	365.11, [M+Na]$^+$, (100) 685.24, [2M+H]$^+$, (0.2) 707.22, [2M+Na]$^+$, (36) 729.21, [2M+2Na-H]$^+$, (0.1)
Glucosamine ($C_6H_{13}NO_5$)	179.0794	180.09, [M+H]$^+$, (100) 202.07, [M+Na]$^+$, (18) 359.17, [2M+H]$^+$, (49) 381.15, [2M+Na]$^+$, (23) 162.08, [M-H$_2$O+H]$^+$, (33)	Cellobiose ($C_{12}H_{22}O_{11}$)	342.1162	343.13, [M+H]$^+$, (0.3) 365.11, [M+Na]$^+$, (100) 685.24, [2M+H]$^+$, (2.2) 707.22, [2M+Na]$^+$, (38)
N-Acetyl-Galactosamine ($C_8H_{15}NO_6$)	221.0899	222.10, [M+H]$^+$, (0.01) 244.08, [M+Na]$^+$, (71) 266.06, [M+2Na-H]$^+$, (1) 465.17, [2M+Na]$^+$, (100) 487.15, [2M+2Na-H]$^+$, (1)	Xylose ($C_5H_{10}O_5$)	150.0528	173.04, [M+Na]$^+$, (100) 195.02, [M+2Na-H]$^+$, (2) 323.09, [2M+Na]$^+$, (81) 345.07, [2M+2Na-H]$^+$, (3) 367.06, [2M+3Na-2H]$^+$, (0.1)
N-Acetyl-Glucosamine ($C_8H_{15}NO_6$)	221.0899	222.10, [M+H]$^+$, (1) 244.08, [M+Na]$^+$, (68) 266.06, [M+2Na-H]$^+$, (2) 465.17, [2M+Na]$^+$, (100) 487.15, [2M+2Na-H]$^+$, (1)	Mannose ($C_6H_{12}O_6$)	180.0634	203.05, [M+Na]$^+$, (98) 225.03, [M+2Na-H]$^+$, (1.4) 383.11, [2M+Na]$^+$, (100) 405.10, [2M+2Na-H]$^+$, (4) 427.08, [2M+3Na-2H]$^+$, (0.2)
N-Acetyl-Neuraminic acid ($C_{11}H_{19}NO_9$)	309.1060	332.09, [M+Na]$^+$, (16) 354.08, [M+2Na-H]$^+$, (100) 376.06, [M+3Na-2H]$^+$, (0.5) 641.20, [2M+Na]$^+$, (0.4) 663.18, [2M+2Na-H]$^+$, (2) 685.16, [2M+3Na-2H]$^+$, (28)	Fucose ($C_6H_{12}O_5$)	164.0685	187.06, [M+Na]$^+$, (69) 209.04, [M+2Na-H]$^+$, (1) 351.13, [2M+Na]$^+$, (100)
Glucuronic acid ($C_6H_{10}O_7$)	194.0427	194.04, [M+H]$^+$, (2) 217.03, [M+Na]$^+$, (6) 239.01, [M+2Na-H]$^+$, (100) 433.06, [2M+2Na-H]$^+$, (1) 455.04, [2M+3Na-2H]$^+$, (50)	D0A4	481.0502	482.05, [M+H]$^+$, (1) 504.03, [M+Na]$^+$, (51) 526.02, [M+2Na-H]$^+$, (100)
ICA ($C_6H_9N_3OS$)	171.0466	172.05, [M+H]$^+$, (10) 194.04, [M+Na]$^+$, (100) 216.02, [M+2Na-H]$^+$, (2) 365.08, [2M+Na]$^+$, (33)	D0A6	481.0502	482.06, [M+H]$^+$, (2) 504.04, [M+Na]$^+$, (45) 526.02, [M+2Na-H]$^+$, (100)

1. The relative Intensity (%) value of observed ions are given in parentheses next to each complex ion. The most intense peak in single state MS spectra are defined as 100.

Table 4.4. Characteristic MS Peaks of 1:1 ICA-Carbohydrate complexes and their MS/MS

(M denotes the corresponding carbohydrate molecule)

Analytes	Observed m/z	MS/MS Product Ion	Analytes	Observed m/z	MS/MS Product Ion
ICA + carbohydrate	Mass of adduct ion (Intensity, S/N)	Mass of molecular ion (Intensity)	ICA + carbohydrate	Mass of adduct ion (Intensity, S/N)	Mass of molecular ion (Intensity)
Galactose	374.10, $[ICA+M+Na]^+$, (41, 15417)	194.04, $[ICA+Na]^+$, (100) 203.05, $[M+Na]^+$, (87)	α-^{M}Glu	388.12, $[ICA+M+Na]^+$, (2.1, 1494)	194.04, $[ICA+Na]^+$, (100) 217.07, $[M+Na]^+$, (15)
Glucose	374.10, $[ICA+M+Na]^+$, (0.9, 804)	194.04, $[ICA+Na]^+$, (100) 203.05, $[M+Na]^+$, (8)	β-^{M}Glu	366.08, $[ICA+M+H]^+$, (6, 1935)	172.05, $[ICA+H]^+$, (3) 194.05, $[ICA+Na]^+$, (100) 195.04, $[M+H]^+$, (72)
	396.08, $[ICA+M+2Na-H]^+$, (0.2, 195)	216.02, $[ICA+2Na-H]^+$, (51) 225.03, $[M+2Na-H]^+$, (100)		388.12, $[ICA+M+Na]^+$, (65, 21034)	194.04, $[ICA+Na]^+$, (100) 217.07, $[M+Na]^+$, (14)
Galactosamine	351.13, $[ICA+M+H]^+$, (55, 33497)	162.08, $[M-H_2O+H]^+$, (13) 172.05, $[ICA+H]^+$, (19) 180.09, $[M+H]^+$, (100)	Maltose	514.17, $[ICA+M+H]^+$, (1.5, 451)	172.06, $[ICA+H]^+$, (100)
	373.12, $[ICA+M+Na]^+$, (16, 10165)	194.04, $[ICA+Na]^+$, (46) 202.07, $[M+Na]^+$, (100)		536.15, $[ICA+M+Na]^+$, (9.5, 3093)	194.04, $[ICA+Na]^+$, (2) 365.11, $[M+Na]^+$, (100)
Glucosamine	351.13, $[ICA+M+H]^+$, (43, 18890)	162.08, $[M-H_2O+H]^+$, (10) 172.05, $[ICA+H]^+$, (53) 180.09, $[M+H]^+$, (100)	Cellobiose	514.17, $[ICA+M+H]^+$, (1.2, 838)	172.06, $[ICA+H]^+$, (100)
				536.15, $[ICA+M+Na]^+$, (1.1, 846)	365.11, $[M+Na]^+$, (100)
N-Acetyl-Galactosamine	393.14, $[ICA+M+H]^+$, (0.3, 63)	172.05, $[ICA+H]^+$, (68) 222.10, $[M+H]^+$, (100)	Xylose	344.09, $[ICA+M+Na]^+$, (9.0, 5451)	173.04, $[M+Na]^+$, (10) 194.03, $[ICA+Na]^+$, (100) 195.04, $[M+2Na-H]^+$, (5)
	415.12, $[ICA+M+Na]^+$, (15, 3030)	194.03, $[ICA+Na]^+$, (0.2) 244.08, $[M+Na]^+$, (100)		366.08, $[ICA+M+2Na-H]^+$, (2.6, 1553)	194.03, $[ICA+Na]^+$, (100) 195.04, $[M+2Na-H]^+$, (12)
	437.10, $[ICA+M+2Na-H]^+$, (0.5, 107)	244.08, $[M+Na]^+$, (47) 266.06, $[M+2Na-H]^+$, (100)		388.06, $[ICA+M+3Na-2H]^+$, (0.1, 65)	194.03, $[ICA+Na]^+$, (13) 216.02, $[ICA+2Na-H]^+$, (100)
N-Acetyl-Glucosamine	415.13, $[ICA+M+Na]^+$, (53, 13720)	194.04, $[ICA+Na]^+$, (12) 244.08, $[M+Na]^+$, (100)	Mannose	374.10, $[ICA+M+Na]^+$, (20, 10478)	194.03, $[ICA+Na]^+$, (78) 203.05, $[M+Na]^+$, (87)
				396.08, $[ICA+M+2Na-H]^+$, (0.5, 291)	216.02, $[ICA+2Na-H]^+$, (22) 225.03, $[M+2Na-H]^+$, (100)
N-Acetyl-Neuraminic acid	503.12, $[ICA+M+Na]^+$, (0.03, 15)	194.03, $[ICA+Na]^+$, (42) 332.08, $[M+Na]^+$, (39)	Fucose	358.10, $[ICA+M+Na]^+$, (20, 9978)	187.06, $[M+Na]^+$, (4) 194.03, $[ICA+Na]^+$, (100)
	525.12, $[ICA+M+2Na-H]^+$, (3.1, 1569)	354.08, $[M+2Na-H]^+$, (100)		380.08, $[ICA+M+2Na-H]^+$, (0.6, 290)	194.03, $[ICA+Na]^+$, (9) 209.04, $[M+2Na-H]^+$, (31) 216.02, $[ICA+2Na-H]^+$, (100)
	547.10, $[ICA+M+3Na-2H]^+$, (0.03, 17)	332.09, $[M+Na]^+$, (8) 354.08, $[M+2Na-H]^+$, (100)			
Glucuronic acid	366.08, $[ICA+M+H]^+$, (7, 2658)	194.04, $[ICA+Na]^+$, (100) 195.04, $[M+H]^+$, (73)	D0A4	675.09, $[ICA+M+Na]^+$, (0.01, 3)	194.01, $[ICA+Na]^+$, (53) 504.03, $[M+Na]^+$, (100)
	388.08, $[ICA+M+Na]^+$, (0.1, 43)	194.04, $[ICA+Na]^+$, (100)		697.07, $[ICA+M+2Na-H]^+$, (0.1, 42)	194.01, $[ICA+Na]^+$, (14) 526.02, $[M+2Na-H]^+$, (100)
			D0A6	675.09, $[ICA+M+Na]^+$, (0.01, 11)	194.04, $[ICA+Na]^+$, (14) 504.04, $[M+Na]^+$, (100)
				697.07, $[ICA+M+2Na-H]^+$, (0.02, 22)	194.04, $[ICA+Na]^+$, (16) 526.02, $[M+2Na-H]^+$, (100)

The Relative Intensity (I%) and Signal to Noise Ratio (S/N) values are given in parentheses next to each complex ion in observed m/z column. I% values are reported in parentheses next to each complex ion in MS/MS product ion column. The most intense peak is considered as 100.

Table 4.5. Characteristic MS Peaks of 2:1 ICA-Carbohydrate complexes and their MS/MS (M denotes the corresponding carbohydrate

Analyte	Observed m/z	MS/MS Product Ion	Analyte	Observed m/z	MS/MS Product Ion
ICA + carbohydrate	Mass of adduct ion, (Intensity, S/N)	Mass of molecular ion, (Intensity)	ICA + carbohydrate	Mass of adduct ion, (Intensity, S/N)	Mass of molecular ion,(Intensity)
Galactose	545.13, [2ICA+M+Na]$^+$, (0.1, 35)	365.06, [2ICA+Na]$^+$, (87)	α-^{M}Glu	559.08, [2ICA+M+Na]$^+$, (0.02, 14)	Not measured
Glucose	Not measured		β-^{M}Glu	559.14, [2ICA+M+Na]$^+$, (0.04, 12)	365.06, [2ICA+Na]$^+$, (87)
Galactosamine	522.16, [2ICA+M+H]$^+$, (0.1, 60)	180.09, [M+H]$^+$, (5)	Maltose	Not measured	
Glucosamine	Not measured		Cellobiose	Not measured	
N-Acetyl-Galactosamine	586.15, [2ICA+M+Na]$^+$, (1.3, 264)	244.08, [M+Na]$^+$, (12)	Xylose	537.11, [2ICA+M+2Na-H]$^+$, (0.01, 4)	194.03, [ICA+Na]$^+$, (52)
	608.13, [2ICA+M+2Na-H]$^+$, (0.1, 11)	216.02, [ICA+2Na-H]$^+$, (100) 244.08, [M+Na]$^+$, (97)		559.09, [2ICA+M+3Na-2H]$^+$, (0.03, 19)	194.03, [ICA+Na]$^+$, (9) 216.02, [ICA+2Na-H]$^+$, (15)
N-Acetyl-Glucosamine	Not measured		Mannose	567.11, [2ICA+M+2Na-H]$^+$, (0.03, 18)	216.02, [ICA+2Na-H]$^+$, (100)
N-Acetyl-Neuraminic acid	674.17, [2ICA+M+Na]$^+$, (0.05, 24)	332.08, [M+Na]$^+$, (100) 354.08, [M+2Na-H]$^+$, (50)	Fucose	551.11, [2ICA+M+2Na-H]$^+$, (0.02, 12)	194.04, [ICA+Na]$^+$, (33) 209.04, [M+2Na-H]$^+$, (16) 216.02, [ICA+2Na-H]$^+$, (100)
	696.15, [2ICA+M+2Na-H]$^+$, (0.1, 53)	354.08, [M+2Na-H]$^+$, (100)		573.11, [2ICA+M+3Na-2H]$^+$, (0.01, 9)	216.02, [ICA+2Na-H]$^+$, (100)
Glucuronic acid	559.11, [2ICA+M+Na]$^+$, (0.02, 6)	194.03, [ICA+Na]$^+$, (15)	D0A4	868.13, [2ICA+M+2Na-H]$^+$, (0.02, 23)	194.01, [ICA+Na]$^+$, (48) 526.02, [M+2Na-H]$^+$, (100)
			D0A6	868.12, [2ICA+M+2Na-H]$^+$, (0.01, 13)	194.04, [ICA+Na]$^+$, (43) 526.02, [M+2Na-H]$^+$, (100)

The Relative Intensity (I%) and Signal to Noise Ratio (S/N) values are given in parentheses next to each complex ion in observed m/z column. I% values are reported in parentheses next to each complex ion in MS/MS product ion column. The most intense peak is considered as 100.

Table 4.6. MS data of individual Universal Reader candidates

Analyte (M)	Calculated Monoisotopic Mass	[1]Observed *m/z*
BCA (C$_8$H$_7$N$_3$OS)	193.0310	194.04, [M+H]$^+$, (0.8) 216.02, [M+Na]$^+$, (100) 238.00, [M+2Na-H]$^+$, (4) 409.05, [2M+Na]$^+$, (14) 431.03, [2M+2Na-H]$^+$, (1.5) 453.01, [2M+3Na-2H]$^+$, (0.4)
ICA (C$_6$H$_9$N$_3$OS)	171.0466	172.05, [M+H]$^+$, (21) 194.04, [M+Na]$^+$, (100) 216.02, [M+2Na-H]$^+$, (0.4) 343.10, [2M+H]$^+$, (1.5) 365.08, [2M+Na]$^+$, (37) 387.06, [2M+2Na-H]$^+$, (0.2)

TCA ($C_5H_8N_4OS$)	172.0419	173.05, $[M+H]^+$, (2) 195.03, $[M+Na]^+$, (99) 217.01, $[M+2Na-H]^+$, (0.8) 345.09, $[2M+H]^+$, (0.5) 367.07, $[2M+Na]^+$, (100) 389.05, $[2M+2Na-H]^+$, (1)
PCA ($C_7H_{10}N_2OS$)	170.0514	171.06, $[M+H]^+$, (10) 193.04, $[M+Na]^+$, (22) 341.11, $[2M+H]^+$, (1.2) 363.09, $[2M+Na]^+$, (20)

1. The relative Intensity (%) value of observed ions are given in parentheses next to each complex ion. The most intense peak in single state MS spectra are defined as 100.

Table 4.7. MS data of individual DNA nucleotide monophosphates

Analyte (M)	Calculated Monoisotopic Mass	[1]Observed *m/z*
AMP ($C_{10}H_{14}N_5O_6P$)	331.0682	332.07, $[M+H]^+$, (32) 354.06, $[M+Na]^+$, (23) 376.04, $[M+2Na-H]^+$, (100) 398.02, $[M+3Na-2H]^+$, (6) 707.10, $[2M+2Na-H]^+$, (2) 729.09, $[2M+3Na-2H]^+$, (20)
GMP ($C_{10}H_{14}N_5O_7P$)	347.0631	370.05, $[M+Na]^+$, (4) 392.03, $[M+2Na-H]^+$, (100) 414.01, $[M+3Na-2H]^+$, (53) 761.08, $[2M+3Na-2H]^+$, (1.3)
CMP ($C_9H_{14}N_3O_7P$)	307.0569	308.06, $[M+H]^+$, (2) 330.04, $[M+Na]^+$, (49) 352.03, $[M+2Na-H]^+$, (100) 374.01, $[M+3Na-2H]^+$, (3) 637.10, $[2M+Na]^+$, (1) 659.08, $[2M+2Na-H]^+$, (4) 681.06, $[2M+3Na-2H]^+$, (10)
TMP ($C_{10}H_{13}N_2Na_2O_8P$)	366.0205	367.03, $[M+H]^+$, (100) 389.01, $[M+Na]^+$, (68) 410.99, $[M+2Na-H]^+$, (4) 733.04, $[2M+H]^+$, (4) 755.03, $[2M+Na]^+$, (1) 777.00, $[2M+2Na-H]^+$, (0.4) 798.99, $[2M+3Na-2H]^+$, (0.4)

1. The relative Intensity (%) value of observed ions are given in parentheses next to each complex ion. The most intense peak in single state MS spectra are defined as 100.

Table 4.8. Characteristic MS Peaks of 1:1 & 2:1 BCA-Nucleotide complexes and their MS/MS products

Analyte		Observed *m/z*	MS/MS Product Ion
	Complex	Mass, Complex ion, (Intensity, S/N)	Mass, Complex ion, (Intensity)
BCA + AMP	1:1	569.07, [BCA+AMP+2Na-H]$^+$, (3.9, 292)	216.02, [BCA+Na]$^+$, (2) 376.04, [AMP+2Na-H]$^+$, (100)
		591.05, [BCA+AMP+3Na-2H]$^+$, (1.2, 84)	216.02, [BCA+Na]$^+$, (9) 398.02, [AMP+3Na-2H]$^+$, (100)
	2:1	740.07, [2BCA+AMP+Na]$^+$, (0.6, 34)	216.02, [BCA+Na]$^+$, (10) 354.05, [AMP+Na]$^+$, (12) 376.04, [AMP+2Na-H]$^+$, (100)
		762.06, [2BCA+AMP+2Na-H]$^+$, (0.3, 13)	216.02, [BCA+Na]$^+$, (6) 376.04, [AMP+2Na-H]$^+$, (100)
		784.07, [2BCA+AMP+3Na-2H]$^+$, (0.4, 17)	216.02, [BCA+Na]$^+$, (8) 376.03, [AMP+2Na-H]$^+$, (12) 398.02, [AMP+3Na-2H]$^+$, (100)
	1:2	944.09, [BCA+2AMP+4Na-3H]$^+$, (0.3, 11)	376.04, [AMP+2Na-H]$^+$, (99) 398.02, [AMP+3Na-2H]$^+$, (100)
BCA + GMP	1:1	585.06, [BCA+GMP+2Na-H]$^+$, (0.6, 17)	216.02, [BCA+Na]$^+$, (9) 392.03, [GMP+2Na-H]$^+$, (100)
		607.04, [BCA+GMP+3Na-2H]$^+$, (0.6, 15)	216.02, [BCA+Na]$^+$, (7) 414.01, [GMP+3Na-2H]$^+$, (100)
	2:1	778.10, [2BCA+GMP+2Na-H]$^+$, (0.1, 1)	392.03, [GMP+2Na-H]$^+$, (100)
		800.09, [2BCA+GMP+3Na-2H]$^+$, (0.1, 1)	414.01, [GMP+3Na-2H]$^+$, (100)
BCA + CMP	1:1	545.05, [BCA+CMP+2Na-H]$^+$, (1.2, 21)	352.02, [CMP+2Na-H]$^+$, (100)
		567.03, [BCA+CMP+3Na-2H]$^+$, (0.7, 12)	216.02, [BCA+Na]$^+$, (6) 352.02, [CMP+2Na-H]$^+$, (21) 374.01, [CMP+3Na-2H]$^+$, (100)
	2:1	716.18, [2BCA+CMP+Na]$^+$, (0.1, 1)	330.05, [CMP+Na]$^+$, (15) 352.02, [CMP+2Na-H]$^+$, (100) 374.01, [CMP+3Na-2H]$^+$, (31)
		738.14, [2BCA+CMP+2Na-H]$^+$, (0.1, 1)	352.02, [CMP+2Na-H]$^+$, (100)
		760.01, [2BCA+CMP+3Na-2H]$^+$, (0.4, 4)	352.02, [CMP+2Na-H]$^+$, (100) 374.00, [CMP+3Na-2H]$^+$, (68)
BCA + TMP	1:1	560.05, [BCA+TMP+H]$^+$, (0.5, 5)	367.02, [TMP+H]$^+$, (100)
		582.03, [BCA+TMP+Na]$^+$, (0.7, 7)	216.02, [BCA+Na]$^+$, (16)

		Observed m/z	MS/MS Product Ion
			367.02, [TMP+H]$^+$, (13)
			389.01, [TMP+Na]$^+$, (100)
		604.01, [BCA+TMP+2Na-H]$^+$, (0.6, 6)	216.02, [BCA+Na]$^+$, (27)
			367.02, [TMP+H]$^+$, (21)
			389.01, [TMP+Na]$^+$, (21)
			410.99, [TMP+2Na-H]$^+$, (100)
		626.07, [BCA+TMP+3Na-2H]$^+$, (0.2, 2)	389.00, [TMP+Na]$^+$, (100)
			410.99, [TMP+2Na-H]$^+$, (20)
	2:1	775.14, [2BCA+TMP+Na]$^+$, (0.2, 1)	216.02, [BCA+Na]$^+$, (69)
			389.00, [TMP+Na]$^+$, (100)

The Relative Intensity (I%) and Signal to Noise Ratio (S/N) values are given in parentheses next to each complex ion in observed m/z column. I% values are reported in parentheses next to each complex ion in MS/MS product ion column. The most intense peak is considered as 100.

Table 4.9. Characteristic MS Peaks of 1:1 & 2:1 ICA-Nucleotide complexes and their MS/MS products

Analyte		Observed m/z	MS/MS Product Ion
	Complex	Mass, Complex ion, (Intensity, S/N)	Mass, Complex ion, (Intensity)
ICA + AMP	1:1	525.10, [ICA+AMP+Na]$^+$, (0.2, 105)	354.06, [AMP+Na]$^+$, (100)
		547.08, [ICA+AMP+2Na-H]$^+$, (2, 1145)	376.04, [AMP+2Na-H]$^+$, (100)
		569.06, [ICA+AMP+3Na-2H]$^+$, (0.2, 98)	398.02, [AMP+3Na-2H]$^+$, (100)
	2:1	718.13, [2ICA+AMP+2Na-H]$^+$, (0.3, 149)	376.04, [AMP+2Na-H]$^+$, (100)
		740.11, [2ICA+AMP+3Na-2H]$^+$, (0.3, 192)	194.04, [ICA+Na]$^+$, (10)
			398.02, [AMP+3Na-2H]$^+$, (100)
	1:2	900.13, [ICA+2AMP+3Na-2H]$^+$, (0.2, 96)	376.04, [AMP+2Na-H]$^+$, (100)
		922.11, [ICA+2AMP+4Na-3H]$^+$, (0.3, 162)	398.02, [AMP+3Na-2H]$^+$, (100)
ICA + GMP	1:1	563.08, [ICA+GMP+2Na-H]$^+$, (1.2, 512)	392.03, [GMP+2Na-H]$^+$, (100)
		585.06, [ICA+GMP+3Na-2H]$^+$, (1.3, 536)	414.01, [GMP+3Na-2H]$^+$, (100)
	2:1	734.11, [2ICA+GMP+2Na-H]$^+$, (0.04, 15)	392.03, [GMP+2Na-H]$^+$, (100)
		756.10, [2ICA+GMP+3Na-2H]$^+$, (0.2, 94)	194.04, [ICA+Na]$^+$, (100)
			414.01, [GMP+3Na-2H]$^+$, (60)
ICA + CMP	1:1	479.15, [ICA+CMP+H]$^+$, (0.1, 21)	308.10, [CMP+H]$^+$, (100)
		501.09, [ICA+CMP+Na]$^+$, (0.1, 31)	194.03, [ICA+Na]$^+$, (6)
			330.04, [CMP+Na]$^+$, (100)
		523.07, [ICA+CMP+2Na-H]$^+$, (2.2, 917)	352.03, [CMP+2Na-H]$^+$, (100)
		545.05, [ICA+CMP+3Na-2H]$^+$, (0.6, 223)	374.01, [CMP+3Na-2H]$^+$, (100)
	2:1	694.11, [2ICA+CMP+2Na-H]$^+$, (0.2, 56)	352.03, [CMP+2Na-H]$^+$, (100)

		716.10, [2ICA+CMP+3Na-2H]+, (0.1, 43)	194.04, [ICA+Na]+, (31)
			374.01, [CMP+3Na-2H]+, (100)
ICA + TMP	1:1	538.07, [ICA+TMP+H]+, (0.9, 531)	367.03, [TMP+H]+, (100)
		560.05, [ICA+TMP+Na]+, (1, 522)	389.01, [TMP+Na]+, (100)
		582.04, [ICA+TMP+2Na-H]+, (0.1, 28)	410.99, [TMP+2Na-H]+, (100)
	2:1	731.10, [2ICA+TMP+Na]+, (0.6, 278)	194.04, [ICA+Na]+, (26)
			389.01, [TMP+Na]+, (100)

The Relative Intensity (I%) and Signal to Noise Ratio (S/N) values are given in parentheses next to each complex ion in observed m/z column. I% values are reported in parentheses next to each complex ion in MS/MS product ion column. The most intense peak is considered as 100.

Table 4.10. Characteristic MS Peaks of 1:1 & 2:1 TCA-Nucleotide complexes and their MS/MS products

Analyte		Observed *m/z*	MS/MS Product Ion
	Complex	Mass, Complex ion, (Intensity, S/N)	Mass, Complex ion, (Intensity)
TCA+AMP	1:1	526.10, [TCA+AMP+Na]+, (0.3, 120)	354.06, [AMP+Na]+, (100)
		548.08, [TCA+AMP+2Na-H]+, (1.1, 367)	376.04, [AMP+2Na-H]+, (100)
		570.06, [TCA+AMP+3Na-2H]+, (0.1, 44)	398.02, [AMP+3Na-2H]+, (100)
	2:1	720.12, [2TCA+AMP+2Na-H]+, (0.6, 220)	376.04, [AMP+2Na-H]+, (100)
		742.10, [2TCA+AMP+3Na-2H]+, (0.2, 74)	195.03, [TCA+Na]+, (5)
			398.02, [AMP+3Na-2H]+, (100)
	1:2	901.13, [TCA+2AMP+3Na-2H]+, (0.1, 60)	195.03, [TCA+Na]+, (20)
			376.04, [AMP+2Na-H]+, (63)
TCA+GMP	1:1	564.07, [TCA+GMP+2Na-H]+, (0.2, 80)	195.03, [TCA+Na]+, (6)
			217.01, [TCA+2Na-H]+, (17)
			392.03, [GMP+2Na-H]+, (100)
		586.06, [TCA+GMP+3Na-2H]+, (0.4, 147)	195.03, [TCA+Na]+, (4)
			414.02, [GMP+3Na-2H]+, (100)
	2:1	736.12, [2TCA+GMP+2Na-H]+, (0.2, 70)	392.03, [GMP+2Na-H]+, (100)
		758.10, [2TCA+GMP+3Na-2H]+, (0.4, 139)	414.02, [GMP+3Na-2H]+, (100)
TCA+CMP	1:1	502.09, [TCA+CMP+Na]+, (0.1, 47)	195.03, [TCA+Na]+, (13)
			330.05, [CMP+Na]+, (100)
		524.07, [TCA+CMP+2Na-H]+, (0.9, 291)	195.03, [TCA+Na]+, (10)
			352.03, [CMP+2Na-H]+, (100)
		546.05, [TCA+CMP+3Na-2H]+, (0.4, 141)	195.03, [TCA+Na]+, (6)
			374.01, [CMP+3Na-2H]+, (100)

	2:1	696.11, [2TCA+CMP+2Na-H]$^+$, (0.6, 170)	352.03, [CMP+2Na-H]$^+$, (100)
		718.09, [2TCA+CMP+3Na-2H]$^+$, (0.4, 123)	195.03, [TCA+Na]$^+$, (9) 374.01, [CMP+3Na-2H]$^+$, (100)
	1:2	853.11, [TCA+2AMP+3Na-2H]$^+$, (0.1, 43)	195.03, [TCA+Na]$^+$, (41) 352.03, [CMP+2Na-H]$^+$, (100)
TCA+TMP	1:1	539.07, [TCA+TMP+H]$^+$, (0.3, 117)	195.03, [TCA+Na]$^+$, (100) 367.03, [TMP+H]$^+$, (61)
		561.05, [TCA+TMP+Na]$^+$, (0.4, 171)	195.03, [TCA+Na]$^+$, (11) 217.01, [TCA+2Na-H]$^+$, (78) 389.01, [TMP+Na]$^+$, (100)
	2:1	733.09, [2TCA+TMP+Na]$^+$, (0.5, 185)	195.03, [TCA+Na]$^+$, (12) 389.01, [TMP+Na]$^+$, (100)
	1:2	905.13, [TCA+2AMP+H]$^+$, (0.2, 97)	195.03, [TCA+Na]$^+$, (23) 389.01, [TMP+Na]$^+$, (100)

The Relative Intensity (I%) and Signal to Noise Ratio (S/N) values are given in parentheses next to each complex ion in observed m/z column. I% values are reported in parentheses next to each complex ion in MS/MS product ion column. The most intense peak is considered as 100.

Table 4.11. Characteristic MS Peaks of 1:1 & 2:1 PCA-Nucleotide complexes and their MS/MS products

Analyte		Observed *m/z*	MS/MS Product Ion
	Complex	Mass, Complex ion, (Intensity, S/N)	Mass, Complex ion, (Intensity)
PCA + AMP	1:1	524.11, [PCA+AMP+Na]$^+$, (0.7, 290)	354.06, [AMP+Na]$^+$, (100)
		546.09, [PCA+AMP+2Na-H]$^+$, (2.4, 1033)	376.04, [AMP+2Na-H]$^+$, (100)
	2:1	716.14, [2PCA+AMP+2Na-H]$^+$, (1.6, 482)	376.04, [AMP+2Na-H]$^+$, (100)
		738.12, [2PCA+AMP+3Na-2H]$^+$, (0.3, 83)	193.04, [PCA+Na]$^+$, (4) 398.02, [AMP+3Na-2H]$^+$, (100)
PCA + GMP	1:1	562.08, [PCA+GMP+2Na-H]$^+$, (1.4, 473)	392.03, [GMP+2Na-H]$^+$, (100)
		584.07, [PCA+GMP+3Na-2H]$^+$, (0.5, 171)	414.02, [GMP+3Na-2H]$^+$, (100)
	2:1	732.14, [2PCA+GMP+2Na-H]$^+$, (0.2, 58)	193.04, [PCA+Na]$^+$, (9) 392.03, [GMP+2Na-H]$^+$, (100)
		754.12, [2PCA+GMP+3Na-2H]$^+$, (0.2, 59)	193.04, [PCA+Na]$^+$, (9) 414.02, [GMP+3Na-2H]$^+$, (100)
PCA + CMP	1:1	500.10, [PCA+CMP+Na]$^+$, (0.5, 200)	193.04, [PCA+Na]$^+$, (1) 330.05, [CMP+Na]$^+$, (100)
		522.08, [PCA+CMP+2Na-H]$^+$, (2.3, 865)	193.04, [PCA+Na]$^+$, (1) 352.03, [CMP+2Na-H]$^+$, (100)
		544.06, [PCA+CMP+3Na-2H]$^+$, (0.4, 143)	374.01, [CMP+3Na-2H]$^+$, (100)

	2:1	692.13, [2PCA+CMP+2Na-H]$^+$, (0.7, 199)	193.04, [PCA+Na]$^+$, (6) 352.03, [CMP+2Na-H]$^+$, (100)
		714.11, [2PCA+CMP+3Na-2H]$^+$, (0.3, 74)	193.04, [PCA+Na]$^+$, (11) 374.01, [CMP+3Na-2H]$^+$, (100)
	1:2	851.12, [PCA+2CMP+3Na-2H]$^+$, (0.1, 30)	193.04, [PCA+Na]$^+$, (100) 352.03, [CMP+2Na-H]$^+$, (86)
		873.10, [PCA+2CMP+4Na-3H]$^+$, (0.1, 33)	193.04, [PCA+Na]$^+$, (100) 352.03, [CMP+2Na-H]$^+$, (77)
PCA + TMP	1:1	559.06, [PCA+TMP+Na]$^+$, (0.9, 356)	193.04, [PCA+Na]$^+$, (2) 389.01, [TMP+Na]$^+$, (100)
		581.04, [PCA+TMP+2Na-H]$^+$, (0.1, 20)	193.04, [PCA+Na]$^+$, 11 410.99, [TMP+2Na-H]$^+$, (100)
	2:1	707.13, [2PCA+TMP+H]$^+$, (0.3, 105)	367.03, [TMP+H]$^+$, (100)
		729.11, [2PCA+TMP+Na]$^+$, (0.6, 185)	193.04, [PCA+Na]$^+$, (11) 389.01, [TMP+Na]$^+$, (100)

The Relative Intensity (I%) and Signal to Noise Ratio (S/N) values are given in parentheses next to each complex ion in observed m/z column. I% values are reported in parentheses next to each complex ion in MS/MS product ion column. The most intense peak is considered as 100.

REFERENCES

(1)   Gill, P.; Jeffreys, A. J.; Werrett, D. J. *Nature* **1985**, *318*, 577.

(2)   Wilson, M. R.; DiZinno, J. A.; Polanskey, D.; Replogle, J.; Budowle, B. *Int J Legal Med* **1995**, *108*, 68.

(3)   Edgington, S. M. *Biotechnology (N Y)* **1993**, *11*, 39.

(4)   Franck, A.; Guilley, H.; Jonard, G.; Richards, K.; Hirth, L. *Cell (Cambridge, Mass.)* **1980**, *21*, 285.

(5)   Telford, E. A. R.; Watson, M. S.; McBride, K.; Davison, A. J. *Virology* **1992**, *189*, 304.

(6)   Choi, M.; Scholl, U. I.; Ji, W.; Liu, T.; Tikhonova, I. R.; Zumbo, P.; Nayir, A.; Bakkaloğlu, A.; Özen, S.; Sanjad, S.; Nelson-Williams, C.; Farhi, A.; Mane, S.; Lifton, R. P. *Proceedings of the National Academy of Sciences* **2009**, *106*, 19096.

(7)   Kogan , S. C.; Doherty , M.; Gitschier , J. *New England Journal of Medicine* **1987**, *317*, 985.

(8)   McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; DePristo, M. A. *Genome Res.* **2010**, *20*, 1297.

(9)   Adams, M. D.; Kelley, J. M.; Gocayne, J. D.; Dubnick, M.; Polymeropoulos, M. H.; Xiao, H.; Merril, C. R.; Wu, A.; Olde, B.; et, a. *Science (Washington, D. C., 1883-)* **1991**, *252*, 1651.

(10)   Hattori, M.; Fujiyama, A.; Taylor, T. D.; Watanabe, H.; Yada, T.; Park, H. S.; Toyoda, A.; Ishii, K.; Totoki, Y.; Choi, D. K.; Soeda, E.; Ohki, M.; Takagi, T.; Sakaki, Y.; Taudien, S.; Blechschmidt, K.; Polley, A.; Menzel, U.; Delabar, J.; Kumpf, K.; Lehmann, R.; Patterson, D.; Reichwald, K.; Rump, A.; Schillhabel, M.; Schudy, A.; Zimmermann, W.; Rosenthal, A.; Kudoh, J.; Shibuya, K.; Kawasaki, K.; Asakawa, S.; Shintani, A.; Sasaki, T.; Nagamine, K.; Mitsuyama, S.; Antonarakis, S. E.; Minoshima, S.; Shimizu, N.; Nordsiek, G.; Hornischer, K.; Brandt, P.; Scharfe, M.; Schon, O.; Desario, A.; Reichelt, J.; Kauer, G.; Blocker, H.; Ramser, J.; Beck, A.; Klages, S.; Hennig, S.; Riesselmann, L.; Dagand, E.; Haaf, T.; Wehrmeyer, S.; Borzym, K.; Gardiner, K.; Nizetic, D.; Francis, F.; Lehrach, H.; Reinhardt, R.; Yaspo, M. L. *Nature* **2000**, *405*, 311.

(11)   Wheeler, D. A.; Srinivasan, M.; Egholm, M.; Shen, Y.; Chen, L.; McGuire, A.; He, W.; Chen, Y.-J.; Makhijani, V.; Roth, G. T.; Gomes, X.; Tartaro, K.; Niazi, F.; Turcotte, C. L.; Irzyk, G. P.; Lupski, J. R.; Chinault, C.; Song, X.-z.; Liu, Y.;

Yuan, Y.; Nazareth, L.; Qin, X.; Muzny, D. M.; Margulies, M.; Weinstock, G. M.; Gibbs, R. A.; Rothberg, J. M. *Nature* **2008**, *452*, 872.

(12)   Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

(13)   https://en.wikipedia.org/wiki/DNA_sequencing#cite_note-olsvik1993-1.

(14)   Hall, N. *J. Exp. Biol.* **2007**, *210*, 1518.

(15)   Church, G. M. *Sci. Am.* **2005**, *294*, 46.

(16)   Schuster, S. C. *Nat. Methods* **2008**, *5*, 16.

(17)   https://www.dnalc.org/view/16036-Fred-Sanger-1975.html.

(18)   Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5463.

(19)   http://www.zo.utexas.edu/faculty/sjasper/images/20.12.gif.

(20)   Sanger, F.; Air, G. M.; Barrell, B. G.; Brown, N. L.; Coulson, A. R.; Fiddes, J. C.; Hutchison, C. A., III; Slocombe, P. M.; Smith, M. *Nature (London)* **1977**, *265*, 687.

(21)   Sanger, F. *Annu. Rev. Biochem.* **1988**, *57*, 1.

(22)   McCarthy, J. J.; McLeod, H. L.; Ginsburg, G. S. *Science Translational Medicine* **2013**, *5*, 189sr4.

(23)   Quail, M. A.; Smith, M.; Coupland, P.; Otto, T. D.; Harris, S. R.; Connor, T. R.; Bertoni, A.; Swerdlow, H. P.; Gu, Y. *BMC genomics* **2012**, *13*, 341.

(24)   Bentley, D. R.; Balasubramanian, S.; Swerdlow, H. P.; Smith, G. P.; Milton, J.; Brown, C. G.; Hall, K. P.; Evers, D. J.; Barnes, C. L.; Bignell, H. R.; Boutell, J. M.; Bryant, J.; Carter, R. J.; Keira Cheetham, R.; Cox, A. J.; Ellis, D. J.; Flatbush, M. R.; Gormley, N. A.; Humphray, S. J.; Irving, L. J.; Karbelashvili, M. S.; Kirk, S. M.; Li, H.; Liu, X.; Maisinger, K. S.; Murray, L. J.; Obradovic, B.; Ost, T.; Parkinson, M. L.; Pratt, M. R.; Rasolonjatovo, I. M.; Reed, M. T.; Rigatti, R.; Rodighiero, C.; Ross, M. T.; Sabot, A.; Sankar, S. V.; Scally, A.; Schroth, G. P.; Smith, M. E.; Smith, V. P.; Spiridou, A.; Torrance, P. E.; Tzonev, S. S.; Vermaas, E. H.; Walter, K.; Wu, X.; Zhang, L.; Alam, M. D.; Anastasi, C.; Aniebo, I. C.; Bailey, D. M.; Bancarz, I. R.; Banerjee, S.; Barbour, S. G.; Baybayan, P. A.; Benoit, V. A.; Benson, K. F.; Bevis, C.; Black, P. J.; Boodhun, A.; Brennan, J. S.; Bridgham, J. A.; Brown, R. C.; Brown, A. A.; Buermann, D. H.; Bundu, A. A.; Burrows, J. C.; Carter, N. P.; Castillo, N.; Chiara, E. C. M.; Chang, S.; Neil Cooley, R.; Crake, N. R.; Dada, O. O.; Diakoumakos, K. D.; Dominguez-Fernandez, B.; Earnshaw, D. J.; Egbujor, U. C.; Elmore, D. W.; Etchin, S. S.; Ewan, M. R.; Fedurco, M.; Fraser, L. J.; Fuentes Fajardo, K. V.; Scott Furey, W.; George, D.; Gietzen, K. J.;

Goddard, C. P.; Golda, G. S.; Granieri, P. A.; Green, D. E.; Gustafson, D. L.; Hansen, N. F.; Harnish, K.; Haudenschild, C. D.; Heyer, N. I.; Hims, M. M.; Ho, J. T.; Horgan, A. M. *Nature* **2008**, *456*, 53.

(25)   Rothberg, J. M.; Hinz, W.; Rearick, T. M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J. H.; Johnson, K.; Milgrew, M. J.; Edwards, M.; Hoon, J.; Simons, J. F.; Marran, D.; Myers, J. W.; Davidson, J. F.; Branting, A.; Nobile, J. R.; Puc, B. P.; Light, D.; Clark, T. A.; Huber, M.; Branciforte, J. T.; Stoner, I. B.; Cawley, S. E.; Lyons, M.; Fu, Y.; Homer, N.; Sedova, M.; Miao, X.; Reed, B.; Sabina, J.; Feierstein, E.; Schorn, M.; Alanjary, M.; Dimalanta, E.; Dressman, D.; Kasinskas, R.; Sokolsky, T.; Fidanza, J. A.; Namsaraev, E.; McKernan, K. J.; Williams, A.; Roth, G. T.; Bustillo, J. *Nature* **2011**, *475*, 348.

(26)   Hayden, R. C. *Nature* **2014**, *507*, 294.

(27)   Metzker, M. L. *Genome research* **2005**, *15*, 1767.

(28)   Chan, E. Y. In *Single Nucleotide Polymorphism. Method in Molecular Biology*; Komar, A. A., Ed. 2009; Vol. 578, p 95.

(29)   Sims, D.; Sudbery, I.; Ilott, N. E.; Heger, A.; Ponting, C. P. *Nature reviews. Genetics* **2014**, *15*, 121.

(30)   Puckelwartz, M. J.; Pesce, L. L.; Nelakuditi, V.; Dellefave-Castillo, L.; Golbus, J. R.; Day, S. M.; Cappola, T. P.; Dorn, G. W., 2nd; Foster, I. T.; McNally, E. M. *Bioinformatics* **2014**, *30*, 1508.

(31)   Treangen, T. J.; Salzberg, S. L. *Nature reviews. Genetics* **2012**, *13*, 36.

(32)   Zwolak, M.; Di Ventra, M. *Reviews of Modern Physics* **2008**, *80*, 141.

(33)   Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. *Nat Biotech* **2008**, *26*, 1146.

(34)   KASIANOWICZ, J. J.; BRANDIN, E.; BRANTON, D.; DEAMER, D. W. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13770.

(35)   Jain, M.; Fiddes, I. T.; Miga, K. H.; Olsen, H. E.; Paten, B.; Akeson, M. *Nature methods* **2015**, *12*, 351.

(36)   Manrao, E. A.; Derrington, I. M.; Laszlo, A. H.; Langford, K. W.; Hopper, M. K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. *Nature biotechnology* **2012**, *30*, 349.

(37)   Laszlo, A. H.; Derrington, I. M.; Ross, B. C.; Brinkerhoff, H.; Adey, A.; Nova, I. C.; Craig, J. M.; Langford, K. W.; Samson, J. M.; Daza, R.; Doering, K.; Shendure, J.; Gundlach, J. H. *Nature biotechnology* **2014**, *32*, 829.

(38)   Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. *Nature biotechnology* **2008**, *26*, 1146.

(39)   Ivanov, A. P.; Instuli, E.; McGilvery, C. M.; Baldwin, G.; McComb, D. W.; Albrecht, T.; Edel, J. B. *Nano letters* **2011**, *11*, 279.

(40)   Liang, F.; Li, S.; Lindsay, S.; Zhang, P. *Chemistry* **2012**, *18*, 5998.

(41)   Huang, S.; He, J.; Chang, S.; Zhang, P.; Liang, F.; Li, S.; Tuchband, M.; Fuhrmann, A.; Ros, R.; Lindsay, S. *Nat Nano* **2010**, *5*, 868.

(42)   Richaud, A.; Barba-Behrens, N.; Méndez, F. *Organic letters* **2011**, *13*, 972.

(43)   Thordarson, P. *Chem. Soc. Rev.* **2011**, *40*, 1305.

(44)   Macomber, R. S. *J. Chem. Educ.* **1992**, *69*, 375.

(45)   Blight, B. A.; Hunter, C. A.; Leigh, D. A.; McNab, H.; Thomson, P. I. T. *Nat. Chem.* **2011**, *3*, 244.

(46)   Etxebarria, J.; Vidal-Ferran, A.; Ballester, P. *Chem. Commun. (Cambridge, U. K.)* **2008**, 5939.

(47)   Parker, C. A.; Rees, W. T. *Analyst* **1962**, *87*, 83.

(48)   Dubertret, B.; Calame, M.; Libchaber, A. J. *Nature biotechnology* **2001**, *19*, 365.

(49)   Freire, E.; Mayorga, O. L.; Straume, M. *Analytical chemistry* **1990**, *62*, 950A.

(50)   Lewis, E. A.; Murphy, K. P. In *Protein-Ligand Interactions: Methods and Applications*; Ulrich Nienhaus, G., Ed.; Humana Press: Totowa, NJ, 2005, p 1.

(51)   Mustapha, S. M. F. D. S.; Phillips, T. N. *Journal of Physics D: Applied Physics* **2000**, *33*, 1219.

(52)   Kimberly, M. M.; Goldstein, J. H. *Analytical chemistry* **1981**, *53*, 789.

(53)   http://www.hyperquad.co.uk/hypnmr.htm.

(54)   Vacca, A.; Nativi, C.; Cacciarini, M.; Pergoli, R.; Roelens, S. *J. Am. Chem. Soc.* **2004**, *126*, 16456.

(55)   Frassineti, C.; Ghelli, S.; Gans, P.; Sabatini, A.; Moruzzi, M. S.; Vacca, A. *Anal. Biochem.* **1995**, *231*, 374.

(56)   Frassineti, C.; Alderighi, L.; Gans, P.; Sabatini, A.; Vacca, A.; Ghelli, S. *Analytical and bioanalytical chemistry* **2003**, *376*, 1041.

(57)   Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L. *Clin Biochem Rev* **2003**, *24*, 3.

(58)   Banerjee, S.; Mazumdar, S. *Int. J. Anal. Chem.* **2012**, 282574.

(59)   Pitt, J. J. *Clin Biochem Rev* **2009**, *30*, 19.

(60)   Likhtarovich, A.; Luhin, V.; Sovastei, O.; Zukowski, P.; Dado, M. *Acta Phys. Pol. A* **2015**, *128*, 901.

(61)   Hofstadler, S. A.; Sannes-Lowery, K. A. *Nat. Rev. Drug Discovery* **2006**, *5*, 585.

(62)   Roepstorff, P. *Curr. Opin. Biotechnol.* **1997**, *8*, 6.

(63)   Yates, J. R., III *J. Mass Spectrom.* **1998**, *33*, 1.

(64)   Mayr, B. M.; Kohlbacher, O.; Reinert, K.; Sturm, M.; Groepl, C.; Lange, E.; Klein, C.; Huber, C. G. *J. Proteome Res.* **2006**, *5*, 414.

(65)   Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6591.

(66)   Domon, B.; Aebersold, R. *Science (Washington, DC, U. S.)* **2006**, *312*, 212.

(67)   Loo, J. A. *Int. J. Mass Spectrom.* **2000**, *200*, 175.

(68)   Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4451.

(69)   Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science (Washington, D. C., 1883-)* **1989**, *246*, 64.

(70)   Fenn, J. B. *J Biomol Tech* **2002**, *13*, 101.

(71)   Olumee, Z.; Callahan, J. H.; Vertes, A. *J. Phys. Chem. A* **1998**, *102*, 9154.

(72)   Mora, J. F. d. l. *Annual Review of Fluid Mechanics* **2007**, *39*, 217.

(73)   Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287.

(74)   Nguyen, S.; Fenn, J. B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1111.

(75)   Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* **1968**, *49*, 2240.

(76)   https://en.wikipedia.org/wiki/Tandem_mass_spectrometry.

(77)   Huang, S.; He, J.; Chang, S.; Zhang, P.; Liang, F.; Li, S.; Tuchband, M.; Fuhrmann, A.; Ros, R.; Lindsay, S. *Nature Nanotechnology* **2010**, *5*, 868.

(78)   Chang, S.; He, J.; Zhang, P.; Gyarfas, B.; Lindsay, S. *Journal of the American Chemical Society* **2011**, *133*, 14267.

(79)   Biswas, S.; Sen, S.; Im, J.; Biswas, S.; Krstic, P.; Ashcroft, B.; Borges, C.; Lindsay, S.; Zhang, P. *Hydrogen-Bonding Universal Readers for Identification of DNA Bases in Nanogaps by Recognition Tunneling with High Accuracy (Manuscript in Preparation)* **2016**.

(80)   Chang, S.; Huang, S.; Liu, H.; Zhang, P.; Liang, F.; Akahori, R.; Li, S.; Gyarfas, B.; Shumway, J.; Ashcroft, B.; He, J.; Lindsay, S. *Nanotechnology* **2012**, *23*, 235101.

(81)   Zhao, Y.; Ashcroft, B.; Zhang, P.; Liu, H.; Sen, S.; Song, W.; Im, J.; Gyarfas, B.; Manna, S.; Biswas, S.; Borges, C.; Lindsay, S. *Nature nanotechnology* **2014**, *9*, 466.

(82)   Sen, S.; Biswas, S.; Im, J.; Biswas, S.; Krstic, P.; Ashcroft, B.; Lindsay, S.; Zhang, P. *The π-π Stacking as an Alternative to Hydrogen Bonding for Identification of DNA Bases in Nanogaps by Recognition Tunneling (manuscript ready to be submitted)* **2016**.

(83)   Loader, C. E.; Anderson, H. J. *Can. J. Chem.* **1981**, *59*, 2673.

(84)   Moorthy, J. N.; Singhal, N. *J. Org. Chem.* **2005**, *70*, 1926.

(85)   Schmidt, A.; Shilabin, A. G.; Nieger, M. *Organic & biomolecular chemistry* **2003**, *1*, 4342.

(86)   Zavarzin, I. V.; Yarovenko, V. N.; Shirokov, A. V.; Smirnova, N. G.; Es'kov, A. A.; Krayushkin, M. M. *ARKIVOC*, *2003*, 205.

(87)   Abuzar, S.; Sharma, S. *Indian Journal of Chemistry* **1985**, *24B*, 338.

(88)   Lunt, E.; Newton, C. G.; Smith, C.; Stevens, G. P.; Stevens, M. F. G.; Straw, C. G.; Walsh, R. J. A.; Warren, P. J.; Fizames, C.; Lavelle, F.; Langdon, S. P.; Vickers, L. M. *J. Med. Chem.* **1987**, *30*, 357.

(89)   Huynh Dinh, T.; Igolen, J.; Bisagni, E.; Marquet, J. P.; Civier, A. *Journal of the Chemical Society, Perkin Transactions 1: Organic and Bio-Organic Chemistry* **1977**, 761.

(90)   Morrison, J. J.; Botting, N. P. *J. Labelled Compd. Radiopharm.* **2005**, *48*, 897.

(91)   Sahali, Y.; Skipper, P. L.; Tannenbaum, S. R. *J. Org. Chem.* **1990**, *55*, 2918.

(92)   Chang, S.; Huang, S.; He, J.; Liang, F.; Zhang, P.; Li, S.; Chen, X.; Sankey, O.; Lindsay, S. *Nano letters* **2010**, *10*, 1070.

(93)   Sartorius, J.; Schhneider, H.-J. *Chem. - Eur. J.* **1996**, *2*, 1446.

(94)   Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117.

(95)   Taubitz, J.; Luning, U.; Grotemeyer, J. *Chemical communications* **2004**, 2400.

(96)   Sherman, C. L.; Brodbelt, J. S.; Marchand, A. P.; Poola, B. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1162.

(97)   Qiu, B.; Liu, J.; Qin, Z.; Wang, G.; Luo, H. *Chemical communications* **2009**, 2863.

(98)   Kitova, E. N.; Soya, N.; Klassen, J. S. *Anal. Chem. (Washington, DC, U. S.)* **2011**, *83*, 5160.

(99)   Kwon, S.; Lee, W.; Shin, H.-J.; Yoon, S.-i.; Kim, Y.-t.; Kim, Y.-J.; Lee, K.; Lee, S. *Journal of Molecular Structure* **2009**, *938*, 192.

(100)   Brivio, M.; Oosterbroek, R. E.; Verboom, W.; van den Berg, A.; Reinhoudt, D. N. *Lab on a chip* **2005**, *5*, 1111.

(101)   Veenstra, T. D. *Biochemical and Biophysical Research Communications* **1999**, *257*, 1.

(102)   Loo, J. A. *Mass Spectrom. Rev.* **1997**, *16*, 1.

(103)   Smith, R. D.; Light-Wahl, K. J.; Winger, B. E.; Loo, J. A. *Org. Mass Spectrom.* **1992**, *27*, 811.

(104)   Pramanik, B. N.; Bartner, P. L.; Mirza, U. A.; Liu, Y.-H.; Ganguly, A. K. *J. Mass Spectrom.* **1998**, *33*, 911.

(105)   Kitova, E. N.; El-Hawiet, A.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1908.

(106)   Han, L.; Kitova, E. N.; Tan, M.; Jiang, X.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 111.

(107)   El-Hawiet, A.; Kitova, E. N.; Arutyunov, D.; Simpson, D. J.; Szymanski, C. M.; Klassen, J. S. *Anal. Chem. (Washington, DC, U. S.)* **2012**, *84*, 3867.

(108)   Cubrilovic, D.; Biela, A.; Sielaff, F.; Steinmetzer, T.; Klebe, G.; Zenobi, R. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1768.

(109)   Parodi, A. J. *Annu. Rev. Biochem.* **2000**, *69*, 69.

(110)   Zhao, Y.-Y.; Takahashi, M.; Gu, J.-G.; Miyoshi, E.; Matsumoto, A.; Kitazume, S.; Taniguchi, N. *Cancer Sci.* **2008**, *99*, 1304.

(111)   Ohtsubo, K.; Marth, J. D. *Cell (Cambridge, MA, U. S.)* **2006**, *126*, 855.

(112)   Rosati, F.; Capone, A.; Della Giovampaola, C.; Brettoni, C.; Focarelli, R. *Int. J. Dev. Biol.* **2000**, *44*, 609.

(113)   Lowe, J. B.; Marth, J. D. *Annu. Rev. Biochem.* **2003**, *72*, 643.

(114)   Kawai, T.; Akira, S. *Int. Immunol.* **2009**, *21*, 317.

(115)   Zhang, X.-L. *Curr. Med. Chem.* **2006**, *13*, 1141.

(116)   Laine, R. A. *Glycobiology* **1994**, *4*, 759.

(117)   Han, X.; Zheng, Y.; Munro, C. J.; Ji, Y.; Braunschweig, A. B. *Curr. Opin. Biotechnol.* **2015**, *34*, 41.

(118)   Nagy, G.; Pohl, N. L. B. *Anal. Chem. (Washington, DC, U. S.)* **2015**, *87*, 4566.

(119)   Hofmann, J.; Hahm, H. S.; Seeberger, P. H.; Pagel, K. *Nature (London, U. K.)* **2015**, *526*, 241.

(120)   Im, J.; Biswas, S.; Liu, H.; Zhao, Y.; Sen, S.; Biswas, S.; Ashcroft, B.; Borges, C.; Wang, X.; Lindsay, S.; Zhang, P. *arXiv:1601.04221 [physics.chem-ph]* **2016**.

APPENDIX A

COPYRIGHT PERMISSIONS

ChemSocRev

RSC Publishing

Title:	Determining association constants from titration experiments in supramolecular chemistry
Author:	Pall Thordarson
Publication:	Chemical Society Reviews
Publisher:	Royal Society of Chemistry
Date:	Dec 1, 2010

Copyright © 2010, Royal Society of Chemistry

Order Completed

Thank you for your order.

This Agreement between Sovan Biswas ("You") and Royal Society of Chemistry ("Royal Society of Chemistry") consists of your license details and the terms and conditions provided by Royal Society of Chemistry and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Get the printable license.

License Number	3853340364485
License date	Apr 20, 2016
Licensed Content Publisher	Royal Society of Chemistry
Licensed Content Publication	Chemical Society Reviews
Licensed Content Title	Determining association constants from titration experiments in supramolecular chemistry
Licensed Content Author	Pall Thordarson
Licensed Content Date	Dec 1, 2010
Licensed Content Volume	40
Licensed Content Issue	3
Type of Use	Thesis/Dissertation
Requestor type	academic/educational
Portion	figures/tables/images
Number of figures/tables/images	3
Distribution quantity	100
Format	print and electronic
Will you be translating?	no
Order reference number	None
Title of the thesis/dissertation	Design, Synthesis and Association Study of Universal Readers for Recognition Tunneling
Expected completion date	Apr 2016
Estimated size	110
Requestor Location	Sovan Biswas 1015 E University Drive Apt 101 TEMPE, AZ 85281

npg

nature publishing group

Title: Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling

Author: Yanan Zhao, Brian Ashcroft, Peiming Zhang, Hao Liu, Suman Sen, Weisi Song

Publication: Nature Nanotechnology

Publisher: Nature Publishing Group

Date: Apr 6, 2014

Copyright © 2014, Rights Managed by Nature Publishing Group

Author Request

If you are the author of this content (or his/her designated agent) please read the following. If you are not the author of this content, please click the Back button and select an alternative Requestor Type to obtain a quick price or to place an order.

Ownership of copyright in the article remains with the Authors, and provided that, when reproducing the Contribution or extracts from it, the Authors acknowledge first and reference publication in the Journal, the Authors retain the following non-exclusive rights:

a) To reproduce the Contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).

b) They and any academic institution where they work at the time may reproduce the Contribution for the purpose of course teaching.

c) To reuse figures or tables created by them and contained in the Contribution in other works created by them.

d) To post a copy of the Contribution as accepted for publication after peer review (in Word or Text format) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the Journal article on NPG's web site (eg through the DOI).

NPG encourages the self-archiving of the accepted version of your manuscript in your funding agency's or institution's repository, six months after publication. This policy complements the recently announced policies of the US National Institutes of Health, Wellcome Trust and other research funding bodies around the world. NPG recognises the efforts of funding bodies to increase access to the research they fund, and we strongly encourage authors to participate in such efforts.

Authors wishing to use the published version of their article for promotional use or on a web site must request in the normal way.

If you require further assistance please read NPG's online author reuse guidelines.

For full paper portion: Authors of original research papers published by NPG are encouraged to submit the author's version of the accepted, peer-reviewed manuscript to their relevant funding body's archive, for release six months after publication. In addition, authors are encouraged to archive their version of the manuscript in their institution's repositories (as well as their personal Web sites), also six months after original publication.

v2.0

95