DNA Chemistry

Contents

**Topics Included:** DNA building blocks, DNA structures, chemical stability/susceptibility, relevant enzymatic processes, DNA modifications/bioconjugation, Leveraging DNA: base pairing, chemical diversity, and stereochemical environment.

**Topics Briefly Discussed:** oligonucleotide drugs, chemical etiology, non-natural DNA

**Topics Not Included:** RNA, nucleoside chemistry (O’Hara, 2012; Gianatassio, 2013), nucleotide synthesis, oligonucleotide synthesis, sequencing, nucleobase synthesis, prebiotic chemistry

Selected History

1869 - Friedrich Miescher identifies “nuclein” from white blood cells
1919 - Phoebus Levene identifies components as base, nucleoside and nucleotide
1935 - Klein and Thannhauser isolate single nucleotides from degraded DNA
1937 - First X-ray diffraction showing DNA as a regular structure
1944 - Avery-MacLeod-McCarty experiment published – DNA as ‘transforming principle’
1950 - Chargaff published on species specificity and G/C and A/T ratios (Chargaff’s Rule)
1952 - Hershy-Chase experiment confirms DNA’s role as genetic material
1952 - Franklin X-ray photographs are taken (A- and B-DNA)
1953 - Watson and Crick elucidated structure of DNA
1977 - Frederick Sanger develops dideoxy method of sequencing
1985 - PCR is invented
1994 - First gene therapy is approved by FDA
2003 - Human Genome project declared complete

Fun Facts!!

- A person has roughly $3 \times 10^9$ bases pairs!
- All of your DNA would go across the solar system…twice
- ~ 3% of your DNA is gene encoding
- ~ 8% of your DNA is viral DNA
- You’re DNA is 99.9% similar to everyone else
- There is a Potato Genome Project
- DNA is a flame retardant (phosphoric acid and ammonia)
- You shouldn’t combine dinosaur and frog DNA

WHAT DOES DNA STAND FOR?!

- 2’-deoxy is important for stability
- first found in the center of cells
- $pK_a$ is about 1.5

Fun Facts!!

- A person has roughly $3 \times 10^9$ bases pairs!
- All of your DNA would go across the solar system…twice
- ~ 3% of your DNA is gene encoding
- ~ 8% of your DNA is viral DNA
- You’re DNA is 99.9% similar to everyone else
- There is a Potato Genome Project
- DNA is a flame retardant (phosphoric acid and ammonia)
- You shouldn’t combine dinosaur and frog DNA
DNA Chemistry

Watson-Crick Pairing
- 99% of DNA base pairs
- N-N/N-O distances: 0.28 to 0.30 nm

[Diagram showing Watson-Crick base pairs]

Hoogsteen Pairing
- 1% of DNA base pairing
- disclosed 10 years later
- relevant in G-quadruplex

[Diagram showing Hoogsteen base pairs]

Conformational Considerations

Sugar Puckers
- C2'-endo
- C3'-endo
- C2'-exo

Propeller Twist
- results in greater base-base overlap

Inclination
- A-DNA: observed when dehydrated
- natural in bacterial endospores and some viral capsids
- B-DNA: majority of natural DNA
- first in artificial GC repeats
- Forced by 5mC
- Implicated in gene regulation

Structural Parameters

<table>
<thead>
<tr>
<th></th>
<th>A-type DNA</th>
<th>B-type DNA</th>
<th>Z-type DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation Sense</td>
<td>left</td>
<td>left</td>
<td>right</td>
</tr>
<tr>
<td>Helix Diameter</td>
<td>25.5 Å</td>
<td>23.7 Å</td>
<td>17.4 Å</td>
</tr>
<tr>
<td>Rise/Base</td>
<td>2.3 Å</td>
<td>3.32 Å</td>
<td>3.8 Å</td>
</tr>
<tr>
<td>Rotation/Base</td>
<td>33.6°</td>
<td>35.9°</td>
<td>-30°</td>
</tr>
<tr>
<td>Bases/Turn</td>
<td>~11</td>
<td>~10</td>
<td>12</td>
</tr>
<tr>
<td>Base Pair Tilt</td>
<td>+19°</td>
<td>-1.2°</td>
<td>-9°</td>
</tr>
<tr>
<td>Propeller Twist</td>
<td>+18°</td>
<td>+16°</td>
<td>~0°</td>
</tr>
<tr>
<td>Helix Axis Location</td>
<td>major groove</td>
<td>through base pairs</td>
<td>minor groove</td>
</tr>
<tr>
<td>Sugar Conformation</td>
<td>C3'-endo</td>
<td>C2'-endo</td>
<td>C:2'-endo, G:2'-exo</td>
</tr>
<tr>
<td>Glycosyl-Base Bond</td>
<td>anti</td>
<td>anti</td>
<td>anti (C), syn (G)</td>
</tr>
</tbody>
</table>

DNA Chemistry

**Other Structures**
- **G-Quadruplexes**
  - discovered in 1962 by X-ray diffraction
  - important structure in aptamer technology
  - thought to play important role in gene regulation
  - G-rich telomeres form these – important in cancer biology
  - can be from 1, 2, 3, or 4 strands (examples above)
  - propensity to form/type dependent on sequence and metal ions available

- **Bent B-DNA**
  - A•T propellor twist 5-7º greater than G•C
  - properly spaced Poly-A's cause kink/bend in helix
  - implication in seq. recognition

- **Solvation**
  - must be 30% w/w water to maintain B-DNA
  - many spheres of solvation – 3 defined in some sequences
  - solvation ordered in minor groove (H-bonding bases)
  - “spine of hydration” (left; blue = 1st sphere, yellow = 2nd)
  - solvation less ordered in major groove – disperse negative charge along phosphates
  - ester oxygens don’t usually participate in H-bonding
  - ring oxygens occasionally participate in H-bonding
  - A•T promotes minor groove H-bonding – entropic driving force for protein binding

- **Base Hydration/Water Networks**

- **Metal Binding in Solvation**
  - Na+ and K+ lay in minor groove replacing waters (exact position disputed)
  - exactly locating metals is difficult – Tl+ helps
  - Mg2+ plays a major role in solvation of major groove – serves as counterion to phosphates
  - Mg2+ exists as hexahydrate – associated waters H-bond

- **DNA’s Happy Place: Stability**
  - overall stability is a combination of solvent, pH, and ionic strength
  - base pair H-bonding contributes very little (-3 kcal/mol/bp)
  - equally as many H-bonds available w/ water in ssDNA
  - pi-stacking contributes majority of stabilization (-16 to -51 kcal/mol/bp)
  - colder = more stable
  - pH around neutral is best (pH<11 = denatured)
  - Denatured ≠ Decomposition
  - Tm dependent on above cond. (and G/C content)

- **DNA vs. RNA**
  - RNA is more hydrolytically labile
  - hydroly. results in 2’ or 3’ phosphate
  - Thymine in DNA; Uracil in RNA
  - deamination of C is U (that’s a problem…)
  - base pairing and larger structure is the same

- **Holliday Junction**
  - 4 stranded, 4 armed structure
  - intermediate in chromosomal recombination and double-strand break repair
  - symmetrical sequences can slide at juncton
  - unsemmetrical sequenced cannot – uses in nanotechnology/DNA Origami

Niedle, 2008. (see above)
**DNA Chemistry**

**DNA’s Unhappy Place: Degradation**

**General**
- main degradation is depurination (pH 4)
- faster under acidic conditions and higher heat
- abasic sites strand cleave faster under basic conditions
- thermal denaturation is sequence dependent
- chemical denaturants: ureas, DMSO, formamide, propylene glycol
- denatured DNA is more susceptible
- pH 1 = phosphodiester cleavage

**Metals**
Rule of Thumb: *If it’s not Mg²⁺, Ca²⁺, Na⁺, or K⁺ its going to do something bad.*
Hud, Nucleic Acid-Metal Ion Interactions, 2009.

**Interstrand Binding**
- cisplatnin targets G,G repeats
- binds to N7 of G
- large conformational and extrahelical bases
- signals for apoptosis

**Intrastrand Binding**
- Metal ion invades helix or binds during denaturation
- flips bases (syn) or makes helix “breathe”
- Pt²⁺, Ag⁺, Hg²⁺

**Stabilizing Other Conformations**
- metal binding can stabilize non-B-DNA structures
- can inhibit recognition and/or make susceptible to other chemical modification
- [Co(NH₃)₆]³⁺ and Zn²⁺ polyamine complexes stabilize Z-DNA w/ phosphate binding
- Pt²⁺, Ag⁺, Hg²⁺ all stabilize quadruplexes and triplexes by bridging non-cannonical base pairing

**Depurination**
- nucleotides are naturally susceptible to this at low pH
- Metals have high affinity for N7 of purine bases
- Metal binding retards depurination at low pH & accelerates it at high pH
- ex: Pt⁴⁺(dien) depurination of G pH<4 (slower) pH>4 (faster)
- Pt²⁺, Ag⁺, Zn²⁺, Co²⁺, Ni²⁺, Cu²⁺, Pd²⁺

**Miscellaneous**
- Pb²⁺, Zn²⁺, Cd²⁺ hydroxides cleave phosphodiester
- Pt²⁺ facilitates deamination of C (MeC→T)
- OsO₄ and MnO₄⁻ oxidize 5,6 bond of pyrimidines

**Radicals**
- radical reactions generally result in cleavage of glycosidic bond
- can cause phosphodiester cleavage, abasic sites
- electron rich nucleobases are most sensitive (G)
- fenton chemistry most implicated in biological systems
- ribose H-atom abstraction
- most radicals can participate
- Cu⁺, Fe³⁺, Mn³⁺-oxo, [Ru(bipy)₃]²⁺/²⁺, ROOH

**Photochemistry**
- well documented intrastrand pyrimidine dimers (254 nm)
- guanine oxidation at 266 nm
- photosensitizers (S) lead to a variety of reactivity

**Alkylation**
- any strong alkyl electrophile (most carcinogens are alkylators)
- act mainly at phosphate, guanine-N7, adenine-N3
- generally leads depurination
- mutation results from mispairing or unfixed sites replicating
- generally leads depurination
- act mainly at phosphate, guanine-N7, adenine-N3
- dimethylsulfate, nitrogen mustards, acrylates, most things in the lab

**Intercalating Agents**
- usually flat, cationic, fused aromatic tricycles
- lengthening (~3.4 Å) and unwinding (~26°) of helix
- interactions in major/minor groove increase affinity
- sequence insensitive
- easy entry = dynamic DNA
- increases chemical liability & modulates gene regulation
**DNA Chemistry**

**ENZYMATIC PROCESSES**

**Polymerases**
- Extending Strand
  - Base
  - 3’
  - Template Strand
  - Extending Strand
  - Base
  - 5’
  - Free 3’-OH

**Replication**
- Complex process utilizing many proteins and complexes
- Chemically dangerous process
- Error rate: 1/10^7 bases
- Fidelity is controlled by kinetics
- Many polymerases in each organism
- DNA and RNA polymerase acts similarly

**Transcription**
- RNA polymerases select for NTPs over dNTPs
- Fidelity is controlled by kinetics
- RNA plays many roles (stay tuned!)

**Nucleases**
- Break nucleic acid polymers
- DNase (left) is selective for DNA, and RNase (not pictured) for RNA
- Some have Zn^2+ sites
- Exonucleases cleave at ends of chain
- Endo nucleases cleave mid-chain
- RNase is EVERYWHERE!
- Restriction nucleases (bacterial) used extensively in molecular biology
- Usually guided or sequence specific
- Play roles in gene regulation and repair

**DNA Repair**
- Major repair pathway for many types of damage/mispairs
- Can also involve invasive polymerase
- Repairs: 8-oxo-G, Fapy-G, 3-Me-A, 7-Me-G, xanthine, hypoxanthine, U (T-derived), etc.
- Implicated in some cancers

**Base Excision Repair**
- Methyl guanosine methyl transferase
- Common damage from smoking
- Stoichiometric in MGMT – extremely energy expensive

**Biochemistry 2004, 14317.**

E + DNA \[\rightarrow\] E-DNA \[\rightarrow\] E-DNA-dNTP \[\rightarrow\] E*-DNA-dNTP \[\rightarrow\] E*-DNA+dPP \[\rightarrow\] E-DNA+dPP

- Correct Pair: step 3 is rate limiting
- Incorrect Pair: step 4 is rate limiting

**Chem. Rev. 2003, 2203.**

For mRNA:
- Template strand is “anti-sense”
- Coding strand/RNA transcript are “sense”
DNA Chemistry

Enzymatic Labeling

Random Primed Labeling
- denature/ anneal primer
- polymerase/ labeled dNTPs
- repeat/ denature

PCR Labeling
- denature/ anneal primer
- polymerase/ labeled dNTPs
- labeled ssDNA

Terminal Transferase Labeling
- TdT (enzyme)
- ssDNA, Co²⁺

Biotin Labeling
- intercalation based
- pairing disruptive

Chemical Labeling

Bisulfite Conjugation
- amine and hydrazide nucleophiles common
- water competes
- disrupts pairing (3% labeling ideal)

Bromine Activation
- G: C8
- C: C5
- A: hard

Carbodiimide 5’ Phosphate Conjugation
- reacts via: phosphorimidazolide

Post Incorporation Modifications
- one pot bisfunctionalization

Controlled Radical Polymerization

End Labeled DNA

Labeled ssDNA

- also: dual enzyme Nick Translation Labeling
- control labeled [dNTP] for RPL and NTL
- no site selectivity in RPL and NTL
- PCR labeling possible from RNA (MLLV reverse transcriptase)
- end labeling doesn’t affect hybridization

DNA Chemistry

**Biosensors** (base pairing)

*Ideally*: fast, highly specific detection of analytes at very low concentrations

- Utilizes high specificity of base pairing
- FRET or fluorophore/quencher pair
- Microarray bound = parallel detection
- Point-of-care diagnostics
- Metals, DNA, RNA, proteins

**Deoxyribozymes** (diversity)

- No known natural DNAzymes; lots of ribozymes known
- Nearly all catalyze bioorganic reactions
- In vitro selection necessary; cannot predict 1st→2nd→Function
- Libraries synthesized; hits amplified
- Selection = 10^14 tested simultaneously; Combinatorial = 10^3
- “Coverage” decreases with length but longer sequences yield better hits
- Rate accelerations beyond “effective concentration” effects (10^6 to 10^9)

**DNA-Templated Asymmetric Catalysis** (stereoselective)

- A chiral metal catalyst associated/bound to double helix
- Mostly use undefined DNA sequences (st-DNA)
- Reaction scope extremely limited currently
- Scale limitations

**Intercalative Binding**

- Nucleopeptide linking
- DNA depurination
- DNA capping (adenylation)
- DNA phosphorylation
- DNA ligation
- RNA ligation (3' or 2'-branched)
- DNA cleavage (oxidative)
- RNA cleavage
- DNA cleavage (oxidative)
- RNA ligation (3’ or 2’-branched)
- DNA ligation
- DNA phosphorilation
- DNA capping (adenylation)
- DNA depurination
- Nucleotide binding

**Covalent Binding**

- Reaction Examples:
  - Pd/Ir allylic amination
  - Chiral sulfoxide formation
  - Pd/Ir allylic amination
  - Chiral sulfoxide formation

**Solid-Support**

- Recycle 10x
- Flow application
- Scaled to 1 mmol
- SiO2 & cellulose

**Ligandless**

- Utilizes ion binding G-quadruplexes

**Reactions Catalyzed**

- RNA cleavage
- DNA cleavage (oxidative)
- RNA ligation (3’ or 2’-branched)
- DNA ligation
- DNA phosphorilation
- DNA capping (adenylation)
- DNA depurination
- Nucleotide linking

**Other Reactions**

- Pd/Ir allylic amination
- Chiral sulfoxide formation
- Epoxide resolution
- L-DNA = other antipode
**Oligonucleotide Therapeutics** (base pairing)
- Utilizes base recognition specificity to target specific genes
- Challenges: entry/uptake, RNase resistance, toxicity, affinity, high specificity
- Area for chemical innovation to overcome these challenges

**Oligonucleotide Therapeutic MOAs**
1. **Aptamers**
   - specificity for single protein
   - modulate activity/binding of target
2. **Anti miRNA**
   - decoy of mRNA targeted by native miRNA
   - UPREGULATION
3. **RISC Recruiting**
   - ds-oligo mimics siRNA
   - recruits RISC machinery to cleave target mRNA
4. **Splice Switching**
   - blocks splicing sites during pre-mRNA modification
5. **miRNA Mimic**
   - blocks translation by forming duplex (miRNA-like)
6. **Gapmers**
   - DNA capped with stable variants (see left) at 3' and 5'
   - DNA/RNA duplex recruits RNase H causing cleavage

**Key Points On This Topic**
- also known as “ASOs” or “antisense drugs”
- claimed diaphorone and pharmacophore separation – has not been true
- biggest issue to address has been toxicity: specificity and ADME
- complexity has risen as a result of poor efficacy

**Stereochemistry**
- phosphate changes introduce stereocenter
- $S_p$: more resistant to nuclease
- $R_p$: higher binding affinity
- stereorandom > stereopure (only one)
- selective orders may be beneficial
- # of isomers = $2^{n-1}$

**Conjugation**
- 18-mer = $2^{17} = 131072$ isomers!!

**Good Reviews/Perspectives:**
DNA Origami (base pairing)

Related Topics Worthy of Their Own Meeting:
- Nucleic Acid Chemical Etiology (concept, design, synthesis)
- Oligonucleotides
- Nucleobases

Chemical Etiology

Asks the question: Why are nucleic acids the way they are and not otherwise?

Experimental Set-Up:
1. Use chemical reasoning to propose alternatives to natural nucleic acid components
2. Synthesize those alternatives
3. Study their properties with respect to nucleic acid functions as we know them
4. Compare them to natural nucleic acids (NA)

Assumptions Made:
- NAs originated from a “combinatorial” process with respect to functional and informational storage abilities
- Selection took place in the domain of sugar-based oligonucleotides
- Reasoned component constitutions must be similar to those of natural NAs
- The reasoned components were available
- Must contain a nucleobase (or derivative) and have capacity for pairing in any fashion