In Silico DNA-Based Aptamer Selection and Construction

Pete Gannett

June 18, 2014

HPC

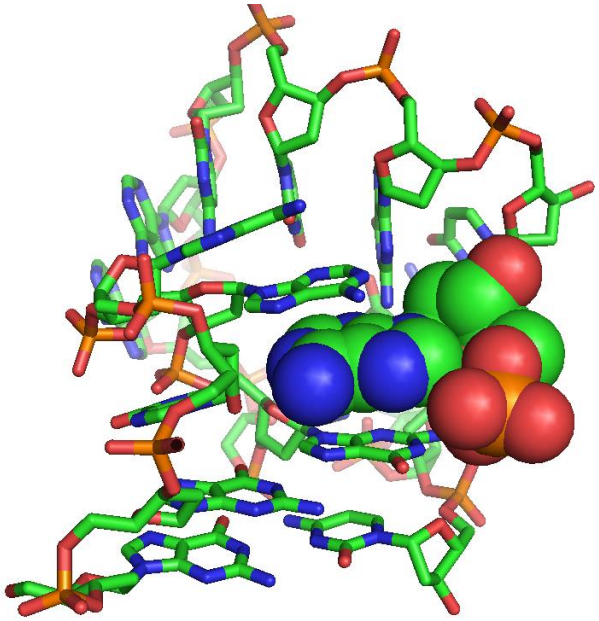
pgannett@hsc.wvu.edu
304.293.1480 (ofc)

Outline

- Introduction to Aptamers
- 1-D aptamer selection
- 2-D and 3-D aptamer structure
- Docking ligand to aptamer
- HTS of aptamer data base
- What's next
- Conclusion

What are Aptamers

- DNA or RNA based
- Typically single-stranded
- 20-30 bases long
- Selected for specific target
- Bind target selectively
- Reversible



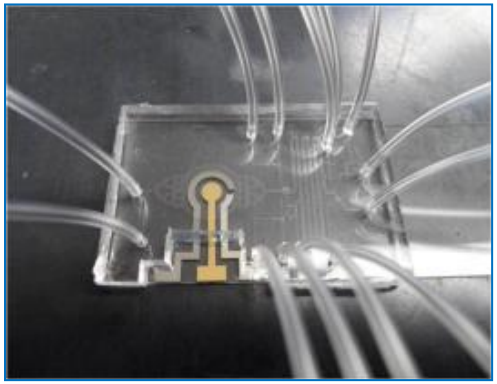
AMP (balls) bound to aptamer (sticks)
Binding is via intercalation

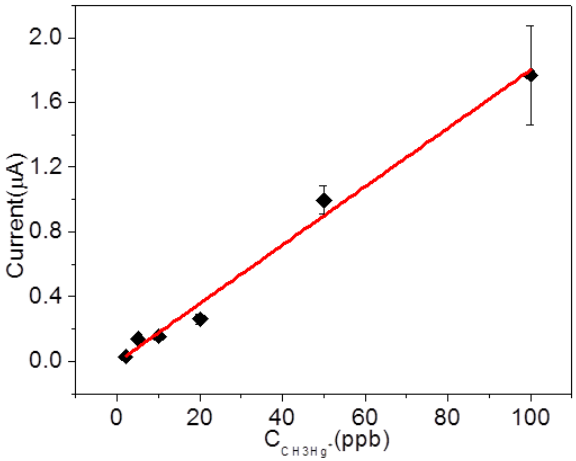
Why DNA-Based

- Technology available to make
- DNA is robust
- RNA sensitive to RNase
- Reversible binding

Application

- Many but targeting Lab-on-a-chip
- Point-of-care devices (e.g., glucose meter)
- Identify aptamers against environmental toxins (Hg, Atrazine)





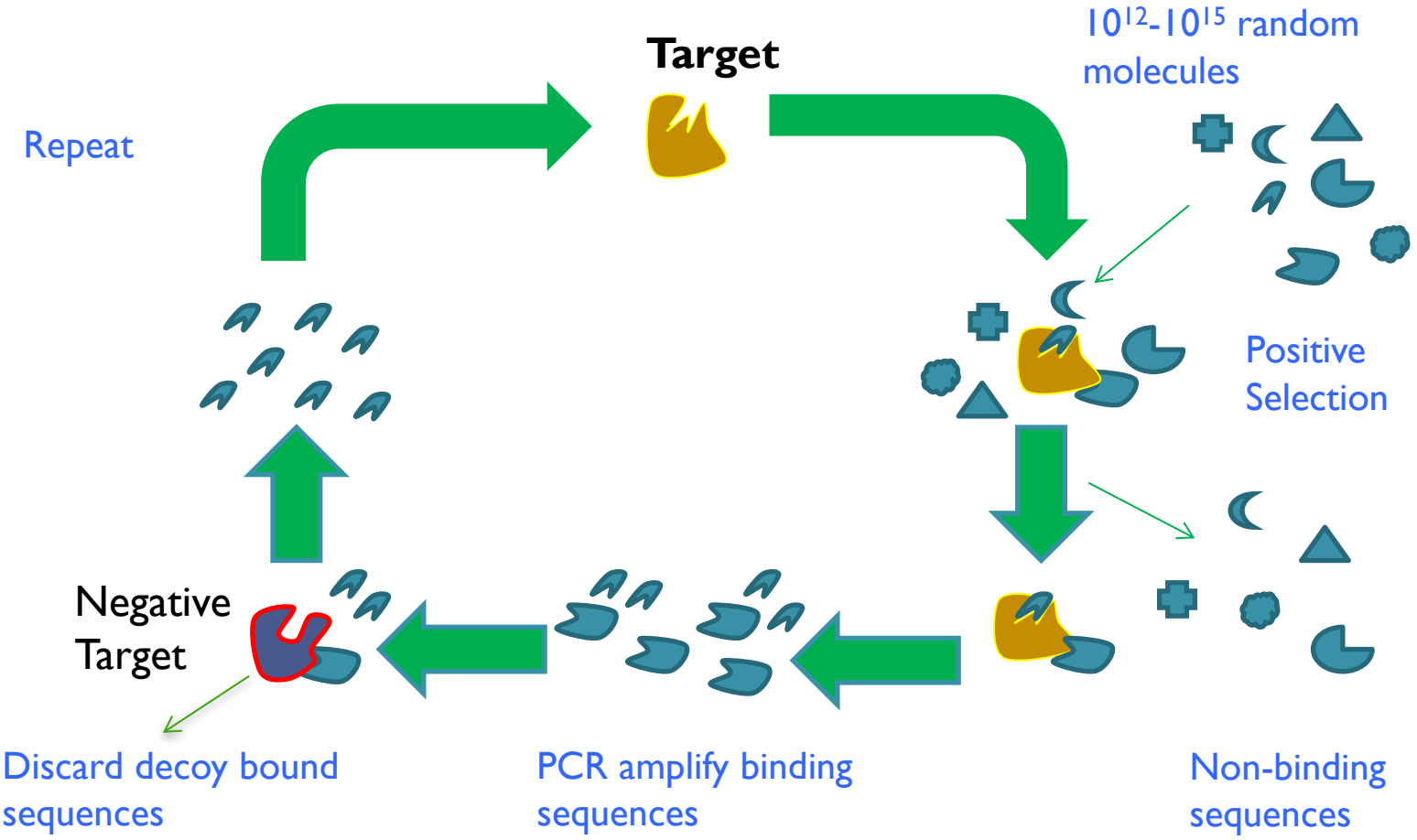
Lab-on-a-chip device for Hg (top right)
Output of chip, linear 40-100ppb

Goal

- Long-range – de novo aptamer creation
- Short range – build structure from sequence
  - Use 1-D known aptameric sequences
  - Predict 2-D/3-D structure
  - Predict Binding of substrate
  - Create candidate aptamer database for HTS

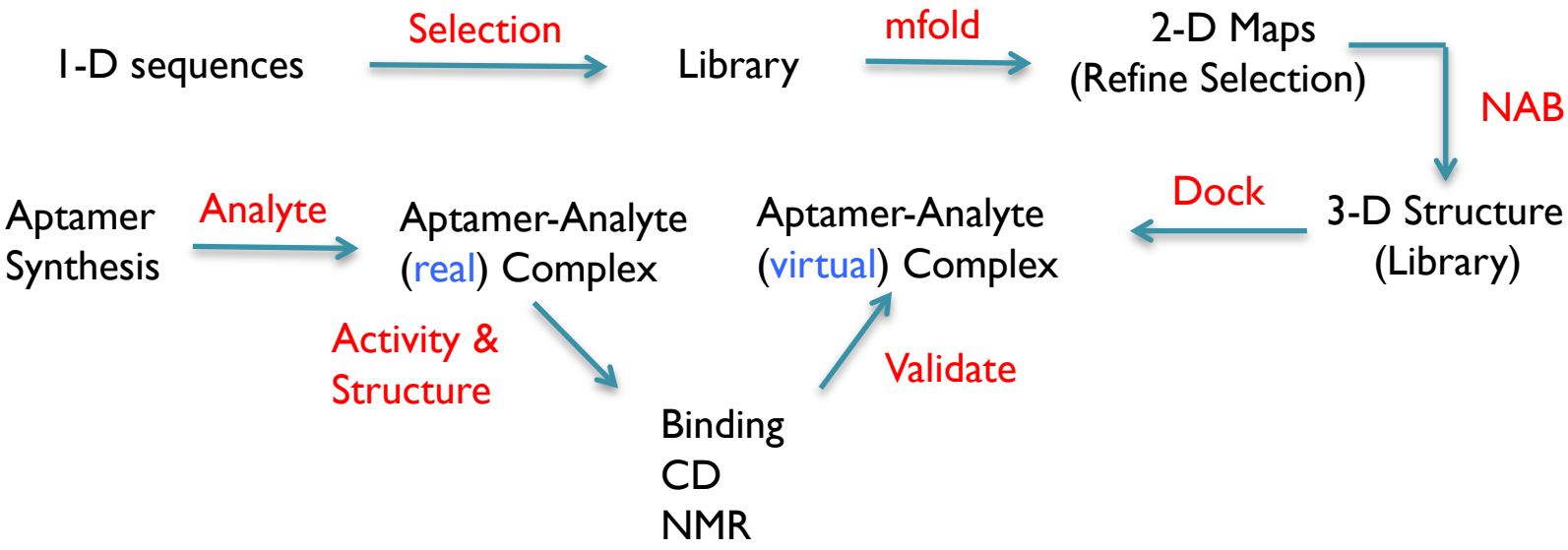
Aptamer selection

- SELEX (Systematic Evolution of Ligands by Exponential Enrichment)
- Tedious, labor intensive , time consuming but it works



Approach

- Informatics approach
  - Selection from 1-D sequence data
  - Convert to 2D with mfold (sequence and structure)
- Structural approach
  - Convert 2D->3D with NAB
  - Relax with MD
- Build database of candidate aptamers
  - Screen candidates with substrate using DOCK
- Experimental validation



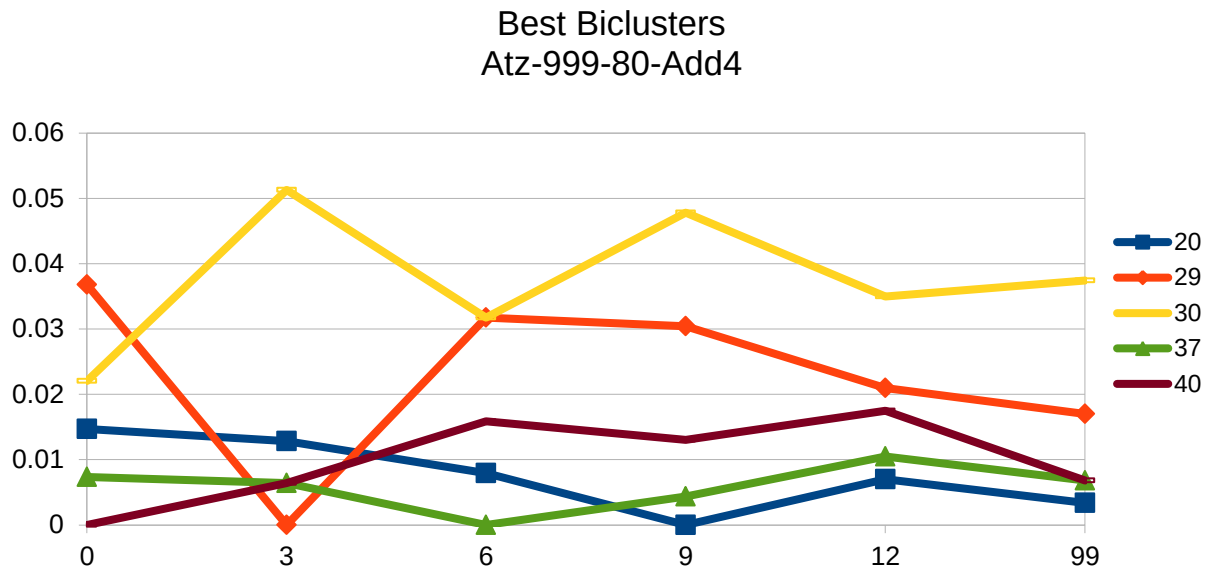
Informatics

- Aptamer sequence information - +1000 known
- Assumptions
  - Aptamers bind through usual interactions
  - Binding involves 4 or fewer bases
  - Rest of aptamer supplies infrastructure
  - May be common repeats
- Can sequences be analyzed to identify aptamers

Biclustering and Selex

- Each 3 rounds of selex aptamer candidates sequenced
  - Examine for what is being enriched
  - But only 50-100 sequences
- Examined each pool with biclustering
  - Shopping Basket Theorem
  - Monitored negative and positive selection
- Acknowledge
  - Only have random pool of 50-100 sequences
  - Ignoring any structural information

Biclustering Results



- Bicluster 40 trends up with substrate (atrazine) – during positive selection (0-12)
- Bicluster 40 trends down with decoy (simizine) – negative selection (12-99)

Biclustering Results

- Of 50 aptamers, biclustering reduced sample to 14

```
R12_1    TGTACCGTCTGAGCGATTCGTACGTTGGGCTTTAGATATGGGTGTCCGCTCAGCTCCAGCCAGTCAGTGTTAAGGAGTGC
R12_8    TGTACCGTCTGAGCGATTCGTACTTGTAGTCTGGCGGTCATTGGAATGAGGCTCTTTAGCCAGTCAGTGTTAAGGAGTGC
R12_19   TGTACCGTCTGAGCGATTCGTACTCGTCTACATTTACTGGGTTGATGATGAGTATTCAGCCAGTCAGTGTTAAGGAGTGC
R12_20   TGTACCGTCTGAGCGATTCGTACAGGAATCGTGGTCGGAATCCCCCGTCGCGCAGCAAGCCAGTCAGTGTTAAGGAGTGC
R12_21   TGTACCGTCTGAGCGATTCGTACTAGCGAGGTGGAATCGGTAAGTGTACCTAGGTTTAGCCAGTCAGTGTTAAGGAGTGC
R12_22     TACCGTCTGAGCGATTCGTACATCGTTAGATGCCGCGTCCATGACTCTACTTCTAAGCCAGTCAGTGTTAAGGAGTGC
R12_23   TGTACCGTCTGAGCGATTCGTACGAACGGCTTTGTACTGTTTGCACTGGCGGATTTAGCCAGTCAGTGTTAAGGAGTGC
R12_24   TGTACCGTCTGAGCGATTCGTACAAGAGACTCGGCTTTTGTAATCTTGCGGTTTTGGAGCCATTCATTGTTAAGGATTGC
R12_25     TACCGTCTGAGCGATTCGTACTAAGTGATGCTGAGATGATGCTCTTGGTACAGTGAGCCAGTCAGTGTTAAGGAGTGC
R12_26   TGTACCGTCTGAGCGATTCGTACGGCTAGAGTTGTTATGTTTCGATGGTCATCTGCAAGCCAGTCAGTGTTAAGGAGTGC
R12_32   TGTACCGTCTGAGCGATTCGTACCTAGCTTGTTTCTCCTCTGCCGCTGTGCGTGAGGAGCCAGTCAGTGTTAAGGAGTGC
R12_33   TGTACCGTCTGAGCGATTCGTACTAAGCGACAGAGCACTGTTGCTGTTACAGTATCCAGCCAGTCAGTGTTAAGGAGTGC
R12_34   TGTACCGTCTGAGCGATTCGTACCGGTTCTTGAGCGGCTGAATAGTATTTTTCTTGCAGCCAGTCAGTGTTAAGGAGTGC
R12_36   TGTACCGTCTGAGCGATTCGTACTGCTCTCGATACGGGTCCTTTAGGGTGAGATTTTAGCCAGTCAGTGTTAAGGAGTGC
R12_37   TGTACCGTCTGAGCGATTCGTACGTTGTAAAAATACTGGAGGGTAGTGAGTTTCCGCAGCCAGTCAGTGTTAAGGAGTGC
R12_38     GCACGATTCGTACGTTTGTACTNGTTAGGAAATAAACGTGNTGAGCGAGCCAGTCAGTGTTAAGGAGTGC
R12_40   TGTACCGTCTGAGCGATTCGTACCGCTATATGGACAACCCCCTGTGTCACATGGTTTAGCCAGTCAGTGTTAAGGAGTGC
R12_41   TGTACCGTCTGAGCGATTCGTACAGCTGACAAGTCGTGTGTCGCCGAAGGCTAGGTTAGCCAGTCAGTGTTAAGGAGTGC
R12_42   TGTACCGTCTGAGCGATTCGTACGATTATCCCGTGATAGTATGTTATACTTGGGGTGAGCCAGTCAGTGTTAAGGAGTGC
R12_44   TGTACCGTCTGAGCGATTCGTACCACTAGTCCTGGGATCAGGTTAGTGTACCCACGTAGCCAGTCAGTGTTAAGGAGTGC
```

- Selected aptamer (R12_23) was in the pool
- Need to increase size of sampling to improve this approach

Structural Approach to Aptamer Selection

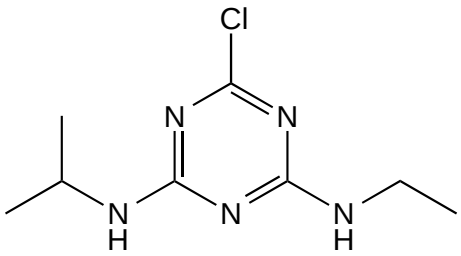
- Sequences come from Ellington Database
  - 1000 Aptamers, RNA and DNA based
  - For ~ 10 there are known structures

- In house – Dr. Letha Sooter
  - Identifies aptamers from pool of ~10^{12}
  - Targets small molecules – Atrazine
  - And large - exotoxin

Aptameric Sequence - Atrazine

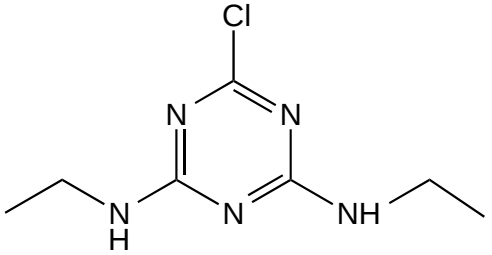
- End game – still get 50-100 sequences. Which to pick?
- Informatics reduced it to 14
- For Atrazine, Sequence selected was - TGTACCGTCTGAGCGATTCGTACGAACGGCTTTGTACTGTTTGCACTG GCGGATTTAGCCAGTCAGTGTTAAGGAGTGC
  - Red – primers used to amplify by PCR
  - Black –random sequence
  - Underline – Truncated sequence
  - Rationale – 2D structures and experience
- TACTGTTTGCACTGGCGGATTTAGCCAGTCAGTG

Positive Selection



Atrazine

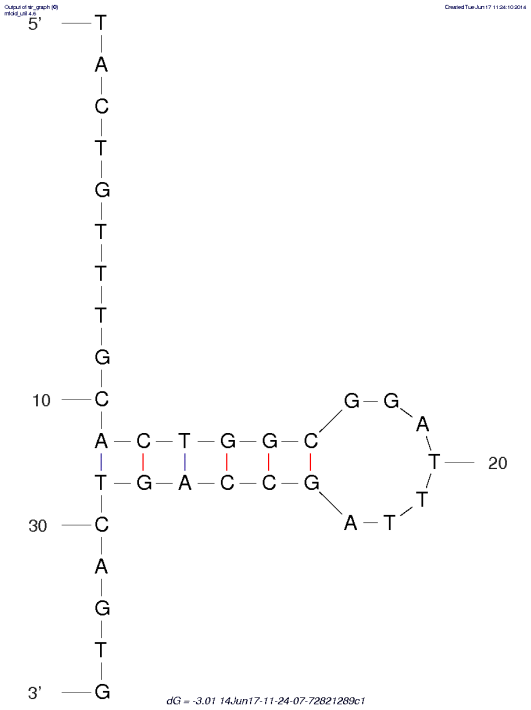
Negative Selection (Decoy)



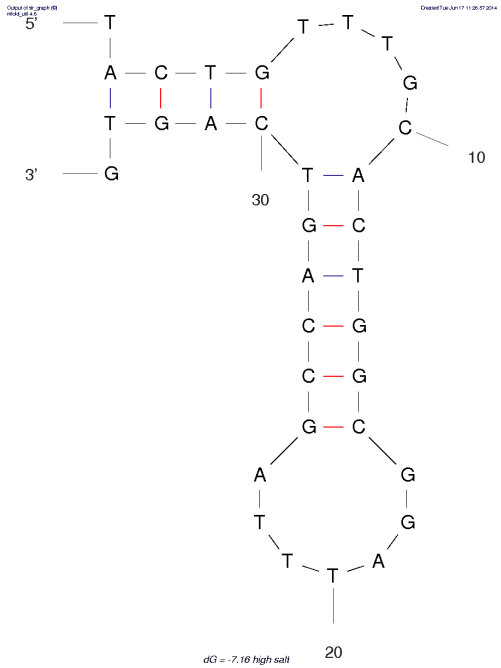
Simizine

Robyns stuff

- 1D sequence – mfold results
  TACTGTTTGCACTGGCGGATTTAGCCAGTCAGTG



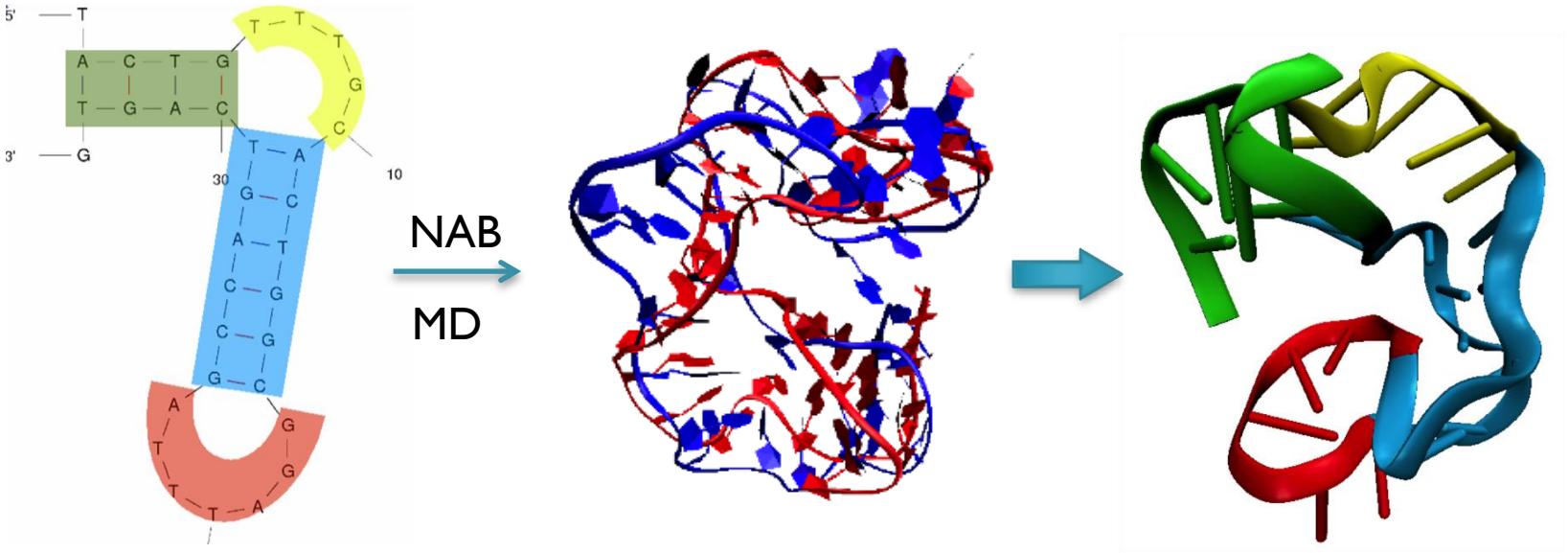
Low Salt (25 mM NaCl)



High Salt (1 M NaCl)+2mM $MgCl_2$

3D structure from 2D

- High salt mfold structure used
- Similar to Selex conditions (~3 kcal/mole more stable)

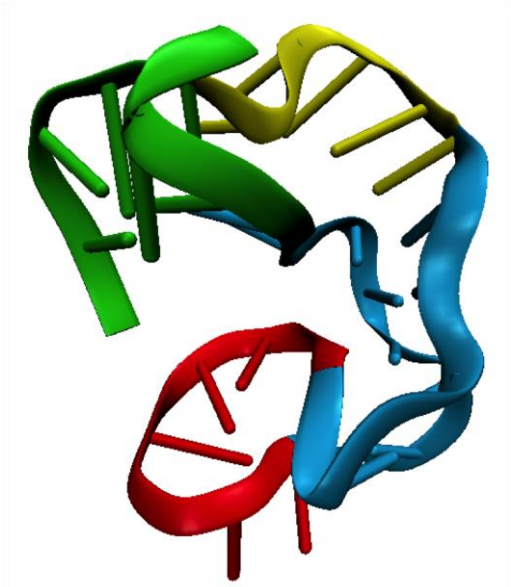


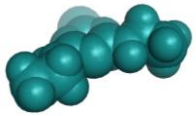
NAB – Yields structure in blue. Molecular Dynamics 'relaxes' to structure in red.

Docking With AutoDock

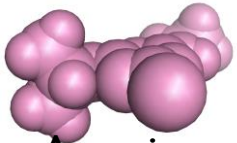
- AutoDock - computational tool that docks analyte molecules to potential binding sites
- Predicts:
  - Binding sites
  - Analyte 3D conformations.
  - Gibbs free energy values.
- Gibbs free energy (“delta G”).
  - Predict if binding is likely
  - Indicates preferred site if more than one
- Illuminates detailed 3D binding descriptors

Docking Aptamer and Substrate

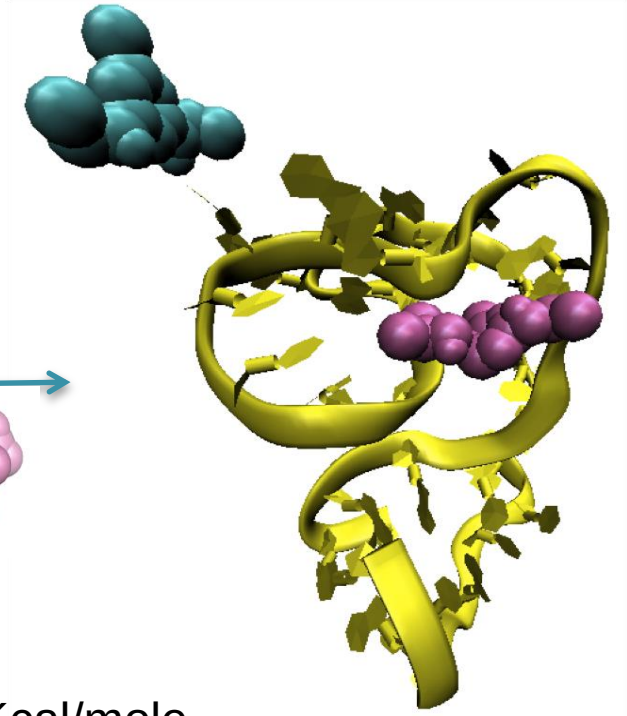




Simazine



Atrazine



$\Delta G_{ATRAZINE}$ = -9.52 Kcal/mole

$\Delta G_{SIMAZINE}$ = -1.15 Kcal/mole

Validation of 3D Structures and Docking

- Determining the Structure of Atrazine Aptamer-Atrazine complex
- Have docked ligands in their known aptamer targets with similar results
- Docked database of 3D structures to substrate and correctly identified aptamer

Challenges

- 2D -> 3D conversion via NAB still has problems
- Quad structures are problematic
- Normalize docking energies from HTS
- Identification of what contributes to a good aptamer

Solutions

- NAB issues
  - Loops are problematic
  - Adding code to resolve
- Quad structures
  - QGRS
  - Screens for quad forming sequences
  - Build with NAB
- ID aptamer properties used for binding
  - Aptamer-Ligand and aptamer structure
  - Utilize DOCK files
- Normalize docking energies
  - Implement published procedure
  - Utilize RNA data base (~500,000 aptamers)

Contact Information

- Pete Gannett
- WVU School of Pharmacy
- Pharmaceutical Sciences
- P.O. Box 9530
- Morgantown, WV 26506

- 304.293.1480 (ofc)
- 304.216.6430 (cell)
- 304.293.2576 (fax)
- pgannett@hsc.wvu.edu

In Silico vs In Vivo selection

Selex – begins with a pool of targets and one substrate (Analyte)

Targets		Target-Substrate
Pool of nucleic acids (NA) ($\sim 10^9 - 10^{13}$ NA)	Substrate →	Aptamer-Substrate (NA-Substrate)

Like high throughput screening (HTS) for drugs but 'reversed'
HTS - Have a pool of substrates and a known target

Pool of substrates ($\sim 10^6$) Potential Drugs	Target → e.g. Protein, DNA, RNA	Target-Substrate

Goal: Build a library of aptamers and use in HTS mode to discover aptamers for desired targets.