Nucleic Acid Triplexes and Quadruplexes
part of “interactions of RNAs and proteins”

Sonja Prohaska

Computational EvoDevo
University Leipzig

June 23, 2014
Nucleic acid triple helices (triplexes)

- oligonucleotide complexes made of three strands
- a DNA duplex and an RNA strand (RNA:DNA-DNA) or
- a DNA duplex and a single DNA strand (DNA:DNA-DNA)
- interaction of nucleic acids without requiring unwinding
- third strand binds major groove with sequence specificity
- forming Hoogsteen or reverse Hoogsteen hydrogen bonds
- with the purine-rich strand of the duplex
the nucleotides

Purines
- Adenine
- Guanine

Pyrimidines
- Cytosine
- Uracil
- Thymine
One Moment please...

the Hoogsteen edge

Hoogsteen

reverse Hoogsteen
Rules for the sequence specific interaction

double-helix (duplex-DNA) + third strand ↔ triple-helix

nucleotide triads

C:GC
T:AT
G:GC
A:AT

motifs

pyrimidine motif (Hoogsteen)
Purine motif (reverse Hoogsteen)
Purine-pyrimidine motif (either)

5’-Y-3’
5’-R-3’
5’-Y-3’
3’-R-5’
3’-Y-5’
3’-GT-5’
5’-GT-3’

Purine of the duplex
Pyrimidine of the duplex
Third strand nucleotide

Sonja Prohaska  Nucleic Acid Triplexes and Quadruplexes
Triplexator: Predicting triple helices

- find putative triplex target sites (TTSs)
  - $n$ and $m$ – minimum (19nt) and maximum triplex length
  - $\omega$ – maximum number of consecutive noncanonical triads
  - $\epsilon$ – maximum proportion (10%) of noncanonical triads
  - $g$ – minimum fraction (50%) of guanine in purine-rich strand
  - find unique genomic loci

- find putative triplex forming strand (TFO)
  - search transcriptome or whole genome (80% is transcribed anyway)
  - TFO length at least 19nt
  - $g = 50\%$ and $\epsilon = 10\%$
  - filter out low complexity regions (e.g. poly-A tails)

- find putative triplexes (TFO-TSS pairs)
Triplexes *in vivo*?

**Triplexator results:**

Are there Triplexes *in vivo*?

- triple-helix formation induces recombination in repair
- potentially regulating gene expression
- replication stalling at \((\text{GAA})_n\) repeats (Friedreich’s Ataxia)
- rather a tool: addressing system e.g. for direct genome modification
Quadruplex DNA

- Hoogsteen hydrogen-bonded guanine (G)-tetrad (G-quartet)
- central metal ion ($K^+$, $Na^+$)
- at least two contiguous G-tetrades
- many stacked G-tetrades form a right handed helix
Different Forms of Quatruplexes

- **number of molecules:**
  - one strand: unimolecular quadruplex
  - two strands: bimolecular quadruplex
  - four strands: tetramolecular quadruplex

- **strand direction:** parallel or anti-parallel

- **loops:** lateral, diagonal, propeller

- **loop size:** ranges from 1 to 15nt
A **unimolecular** G-quatruplex-forming sequence can be described as follows:

\[ G_m X_n G_m X_o G_m X_p G_m \]  

where \( G \) ist the nucleotide guanine, \( m \) is the number of stacked G-triads, \( X_n, X_o \) and \( X_p \) are the loop sequences where \( X \) is any nucleotide and \( n, o \) and \( p \) are their lengths.

**Example:** thrombin-binding sequence \( GGTTGGTGTGGTTGG \), an aptamer
A unimolecular quadruplex in Na\(^+\) form with a diagonal and two lateral loops forming from the human telomere sequence is \(AGGG(\text{TTAGGG})_3\). The TTA sequences are found in the loops (PDB 143D).
G-quartet at *Oxytricha nova* Telomeres

A bimolecular quadruplex with four G-quartets and two diagonal $T_4$ loops. The telomere sequence is $G_4T_4G_4$ (PDB 1JPQ).
(G₄T₃G₄): bimolecular, four G-quartets, two lateral loops each with the sequence TTT oriented head to tail (PDB 2AVH).
$G_4 T_3 G_4$: bimolecular, four G-quartets, two lateral loops each with the sequence 'TTT' oriented head to head (PDB 2AVJ).
Alternative Human Telomere quadruplex (metal matters)

AGGG(TTAGGG)$_3$: unimolecular, three G-quartets, $K^+$ form, one propeller loop and two lateral loops (loops have been obtained by molecular dynamics simulations).
NMR-derived topology of the *c-myc* quadruplex

unimolecular, three G-quartets and three propeller loops (PDB 1XAV). Located in the promoter region of *c-myc*. 
Busk FA, Bauer DC, Mattick JS and Bailey TL. (2012) *Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data*; Genome Research 22:1372-1381