

The rigid-body model of nucleic acids, with a personal bias for RNA

Mauricio Esguerra
Lynn Kamerlin Lab.
Uppsala Universitet

mesguerra.org

March 28, 2014

Dagens Meny

|3:00 - |4:10 — Theory background of rigid-block model and geometrical interpretations of collected structural data of nucleic acids.

|4:10 - |4:30 — Fika

|4:30 - |6:00 — Practical exercise on generating nucleic acid structures and playing around with pymol tricks and tips to produce publication quality, “REAL”, nucleic acid cartoons.

Usual ways to classify nucleic acid conformations and a still rather unexplored one -- the rigid-body perspective --.

- The atom-based perspective (A)

- Comparison of backbone atom positions (e.g. RMSD between).
- Comparison from set of backbone, sugar, base, atom positions.

- The bond-based perspective

- Covalent bonds – backbone + glycosidic torsions. (B)
- Pseudo-bonds η and θ . (C)
- Hydrogen bonds forming in edge boundaries (Watson-Crick, Sugar, and Hoogsten).

- The rigid-body-based perspective

- Base-pair parameters.
- Base-pair (base) step parameters.

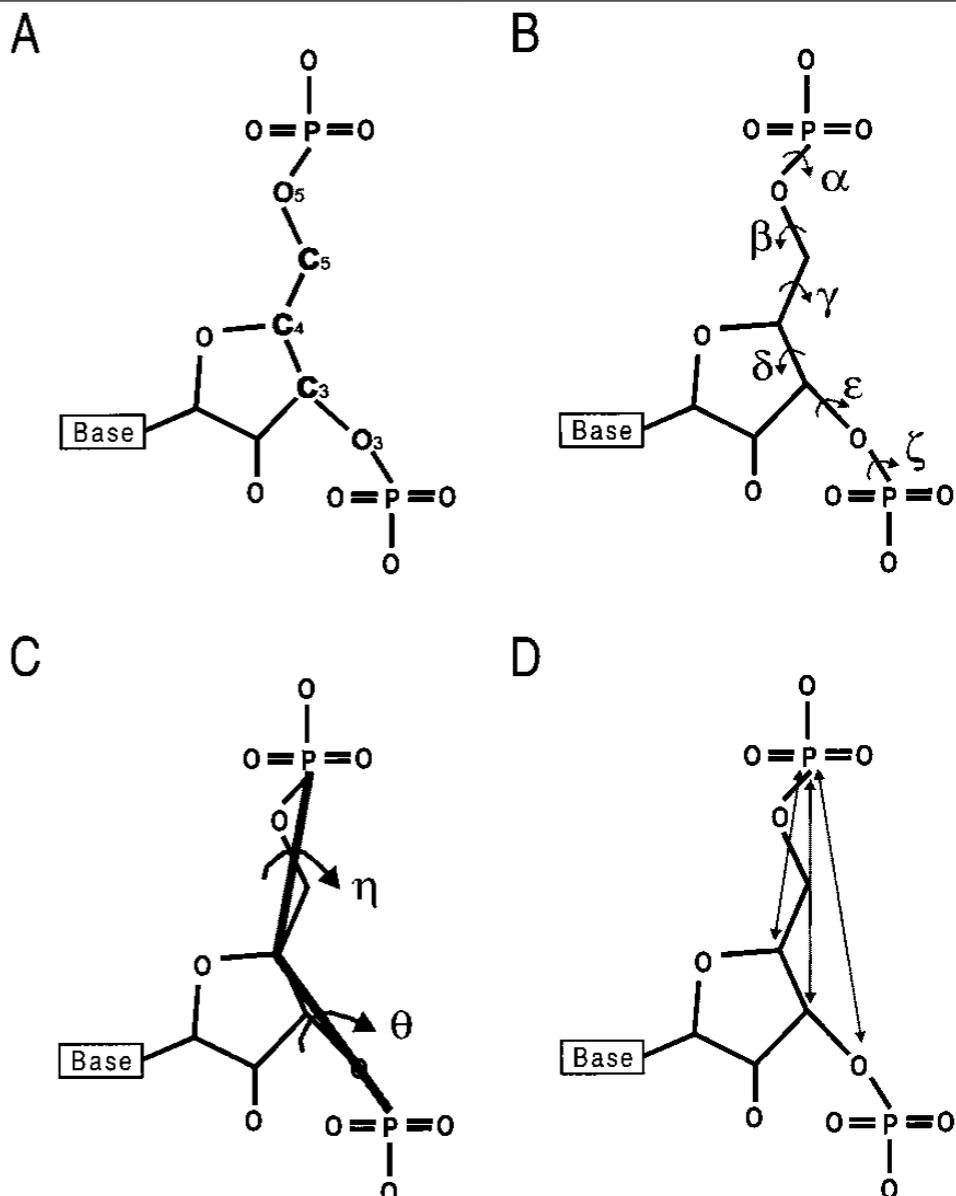


Figure 2. Overview of the four different representations used to fix the molecular structure information of RNA trinucleotides. A: Cartesian coordinates representation, B: torsion angles representation, C: pseudotorsion angles representation, D: distances representation.

Reijmers, T. H.; Wehrens, R. & Buydens, L. M. C.
The Influence of Different Structure Representations on the Clustering of an RNA Nucleotides Data Set
J. Chem. Inf. Comp. Sci., **2001**, 41, 1388-1394

The atom perspective used by Sykes and Levitt to describe the space of conformations of RNA's.

doi:10.1016/j.jmb.2005.06.024

J. Mol. Biol. (2005) 351, 26–38

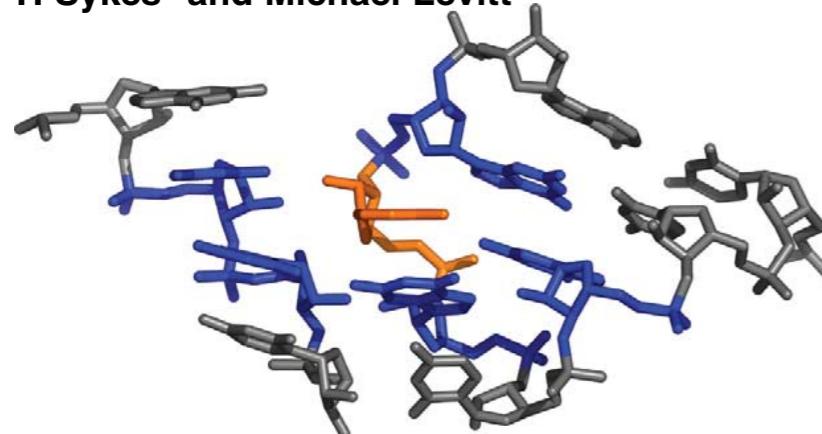
JMB

Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

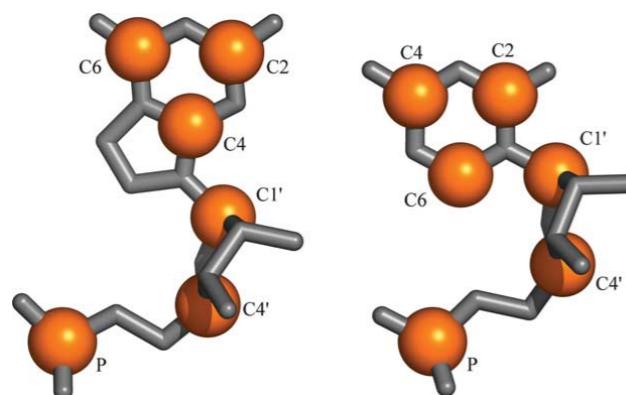


Describing RNA Structure by Libraries of Clustered Nucleotide Doublets

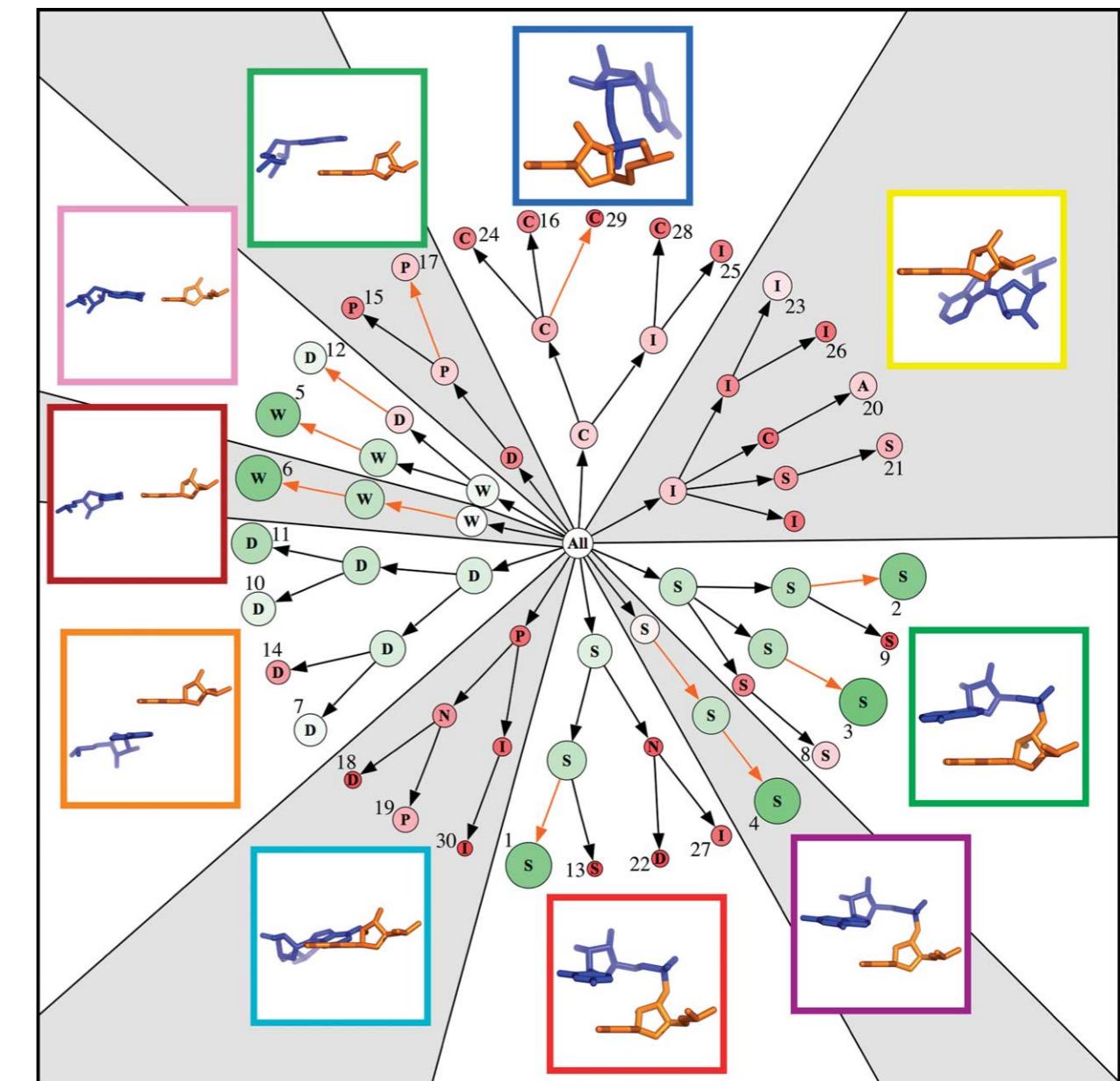
Michael T. Sykes* and Michael Levitt



Five doublets between orange nucleotide and blue neighbors. Doublets are any two nucleotides in a structure that have any heavy atoms within 4 Å of each other



Left guanine, right cytosine. RMSD is computed between atoms highlighted in orange and then clustered using k-means.



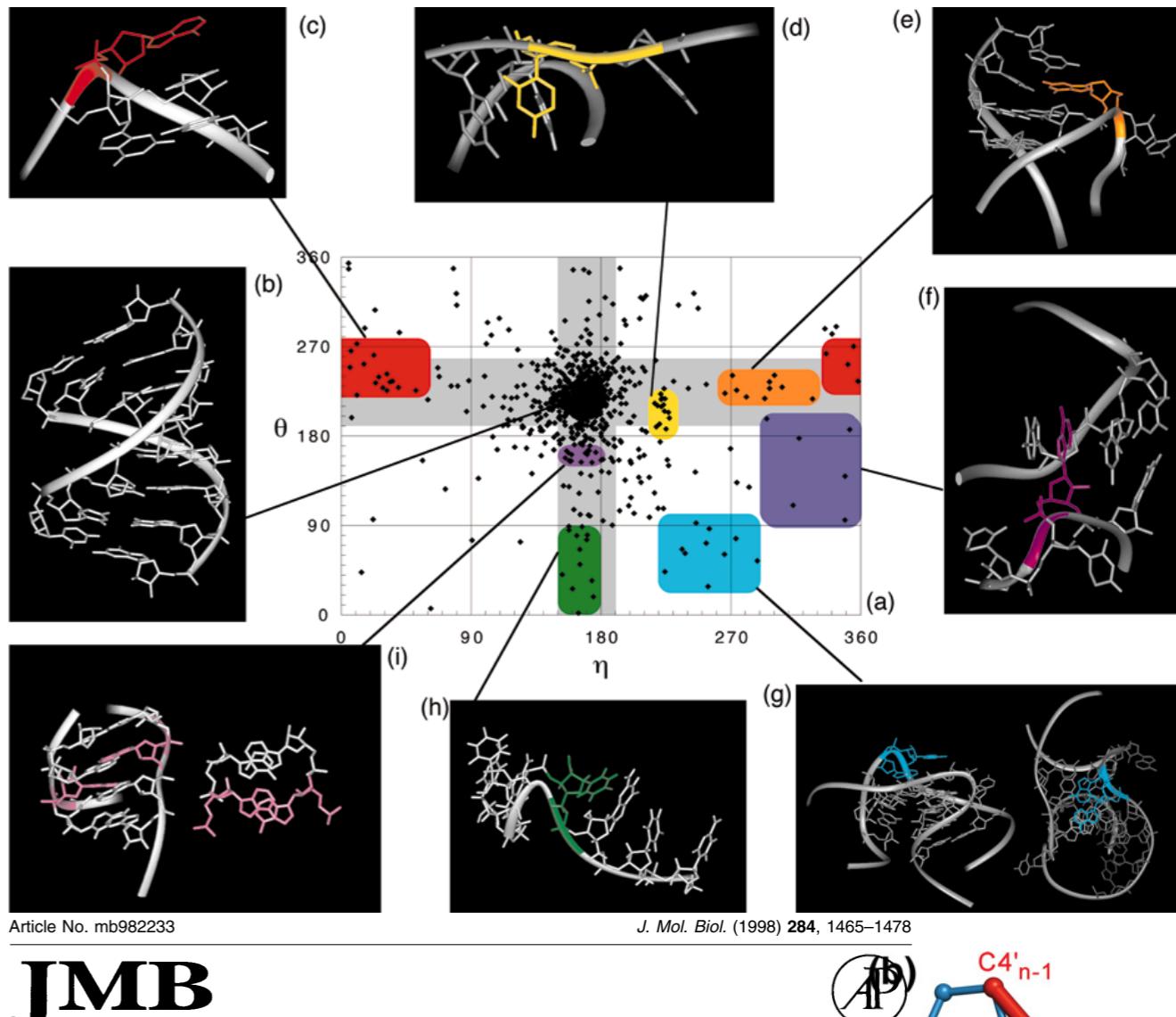
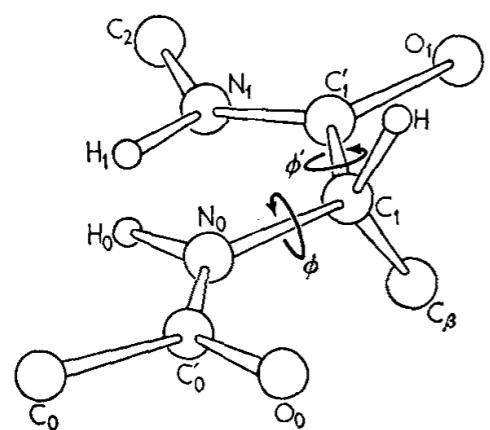
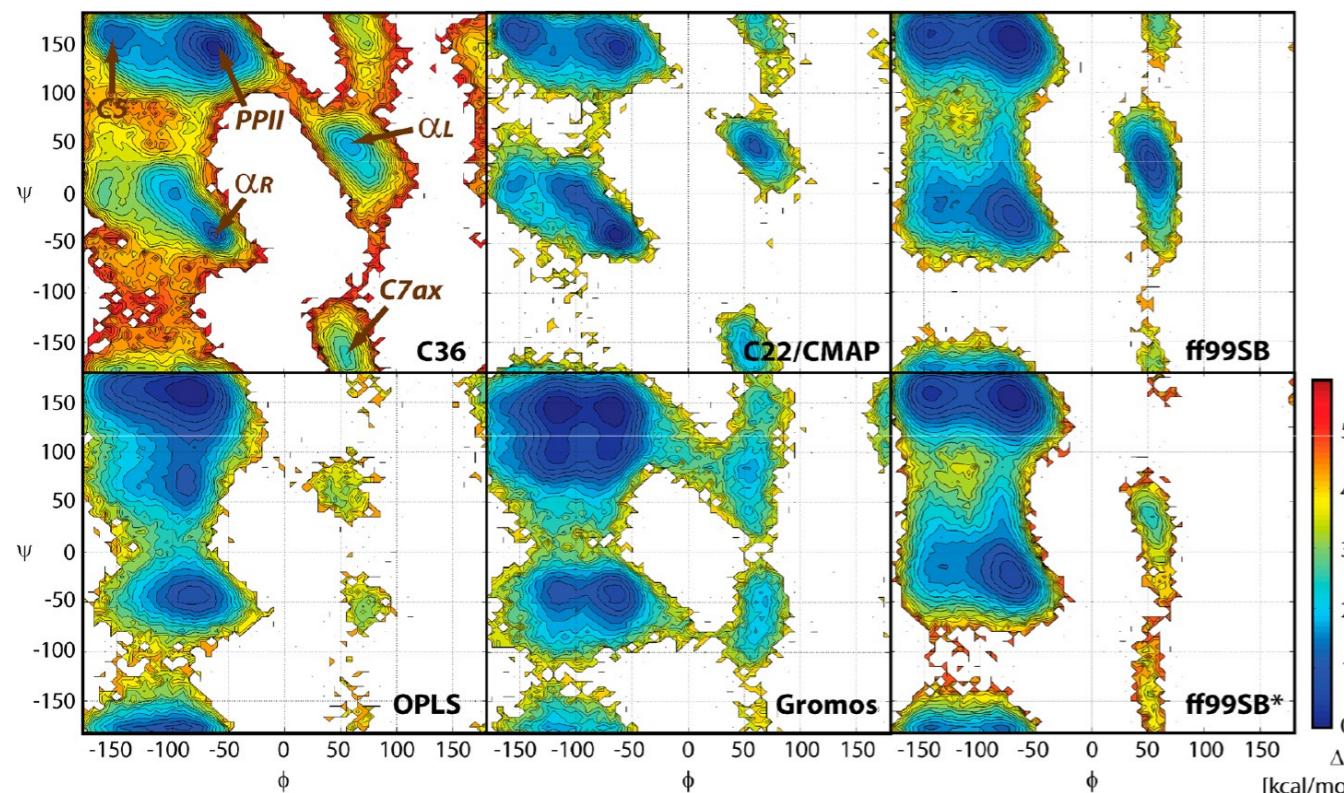
Simulated annealing k-means clustering method of RMSD's between doublets..

The η , θ plot first proposed by Malathi & Yathindra and independently by Olson in analogy to Ramachandran.

Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles

Robert B. Best,^{†,‡} Xiao Zhu,^{‡,§} Jihyun Shim,[‡] Pedro E. M. Lopes,[‡] Jeetain Mittal,[§] Michael Feig,^{||} and Alexander D. MacKerell, Jr.^{*,‡}

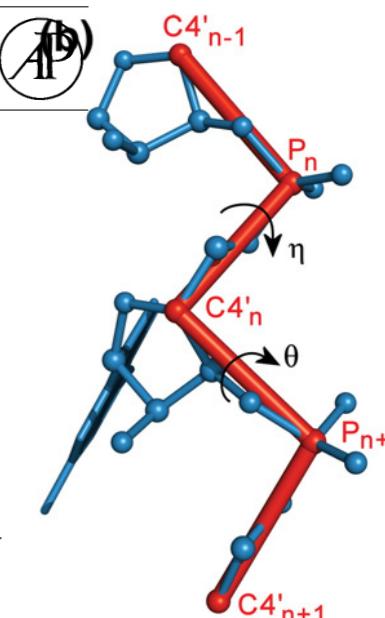
J. Chem. Theory Comput. 2012, 8, 3257–3273



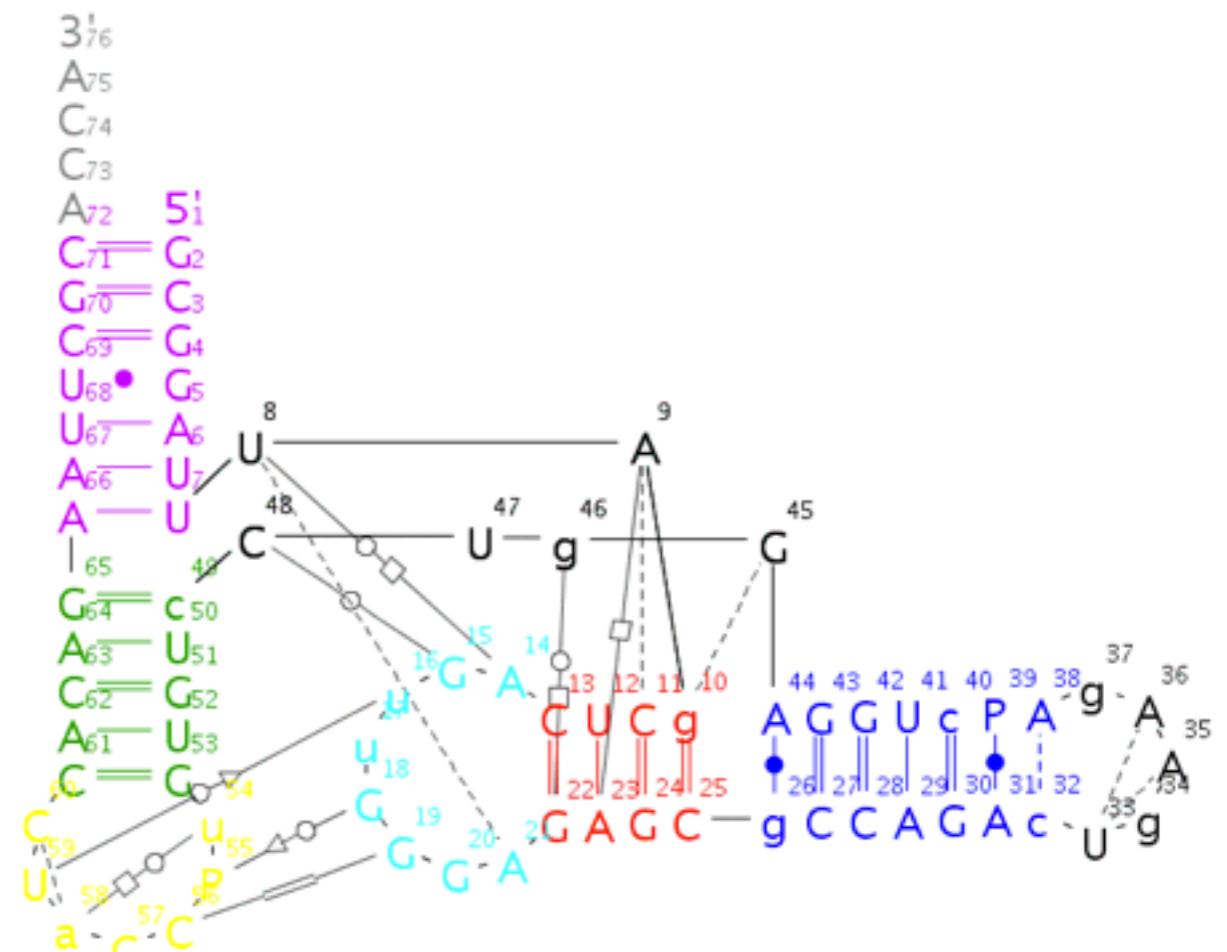
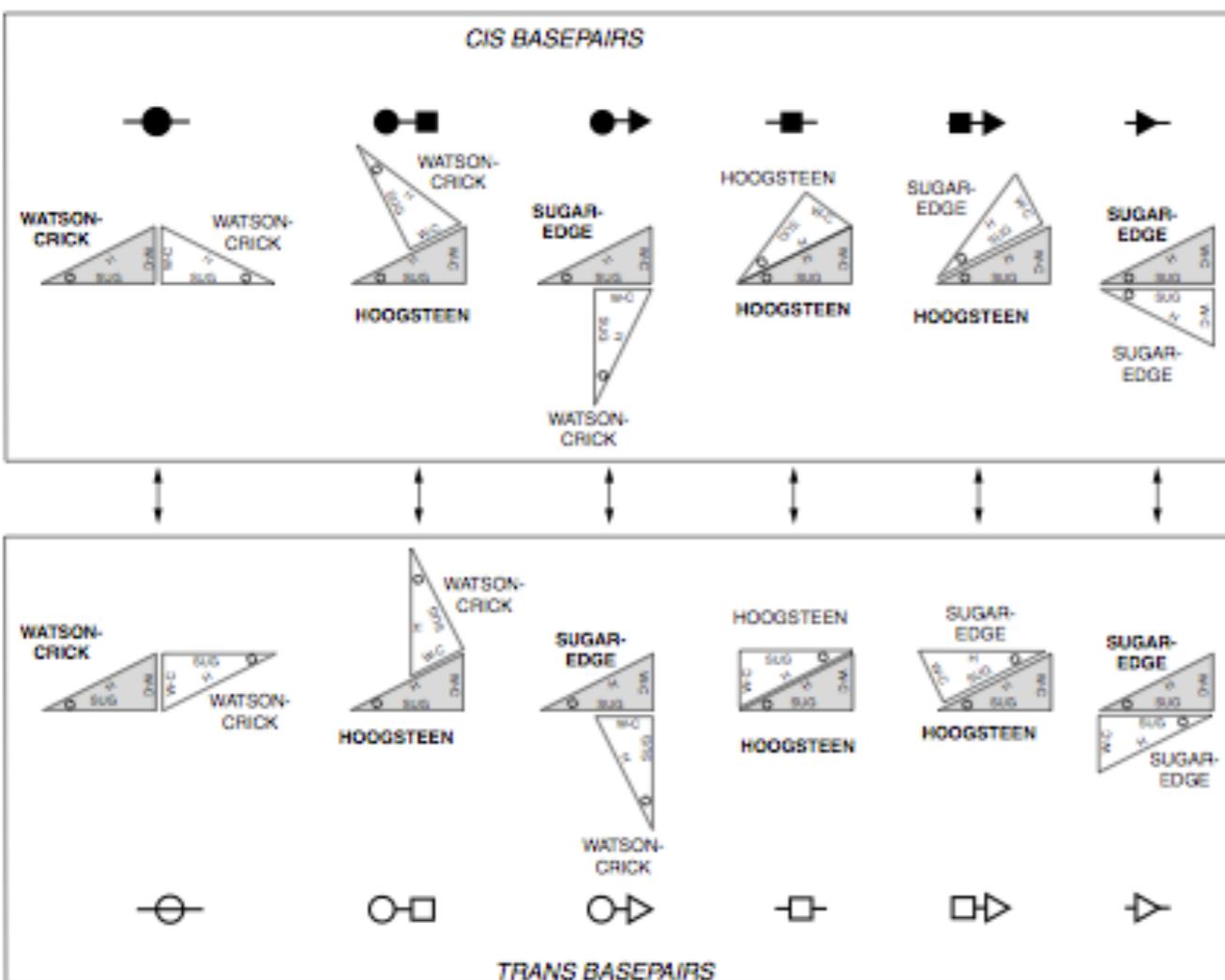
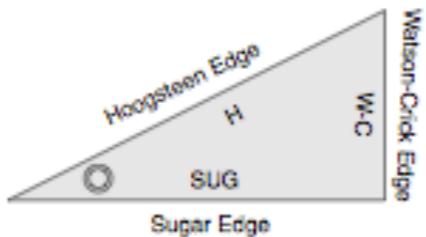
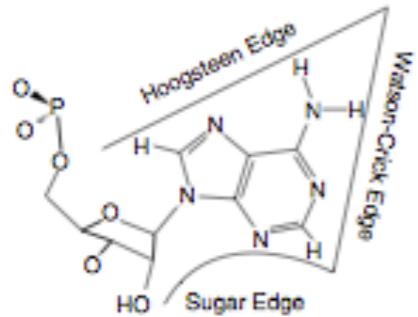
Stepping Through an RNA Structure: A Novel Approach to Conformational Analysis

Carlos M. Duarte and Anna Marie Pyle*

The motifs correspond to the helical A-RNA type conformation in the middle(b), stacked turn like GNRA's (c), w-switch(d), flip-turn (e), c-20 bend (f), stack-switching (g), base-twist (h), cross-strand stack (i).



More on the bond-based perspective, Leontis-Westhof classification of base-pairs. Others are Tinoco, Gutell, and Saenger.



Diving right in. The genesis and development of the rigid-block perspective for nucleic acids.

1953 Watson-Crick.

1980's block models are popular. Dickerson-Drew dodecamer (1981). Calladine-Drew A to B conformational change seen from the base perspective.

1988 Cambridge accord for base-pair and bp step parameters.

1999 Tsukuba / 2001 JMB Standard Reference Frame.

2003 13th Albany Conversation Standard Method. 3DNAV1

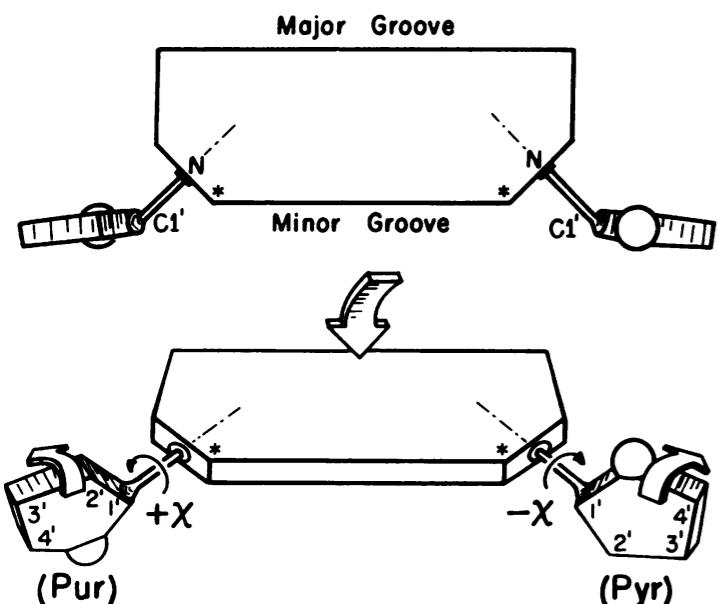
2008 3DNAV2

2009 Curves+

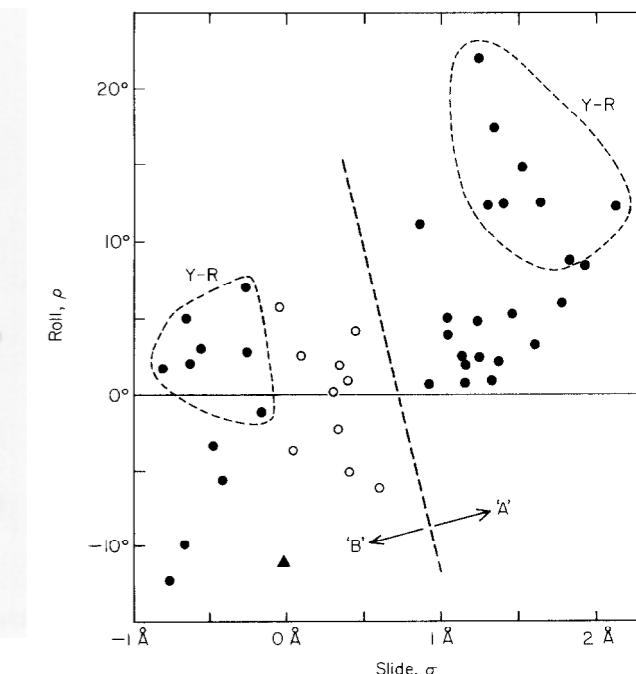
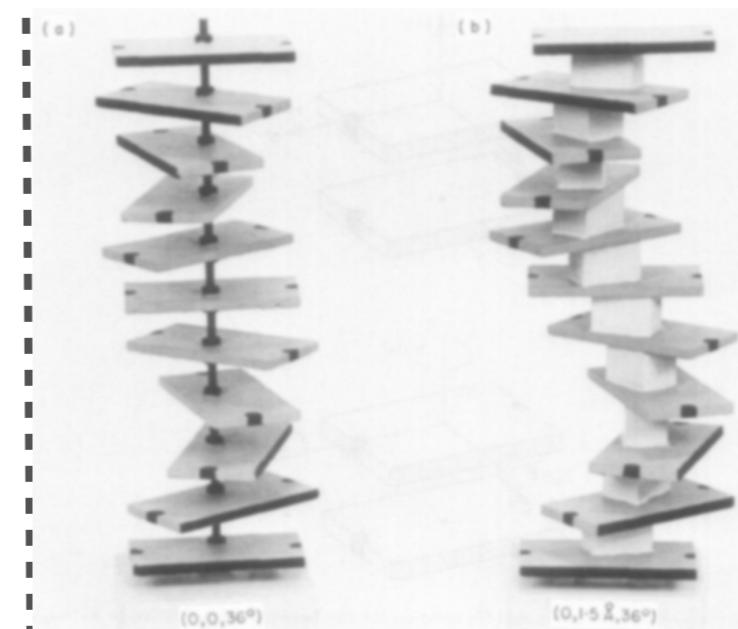
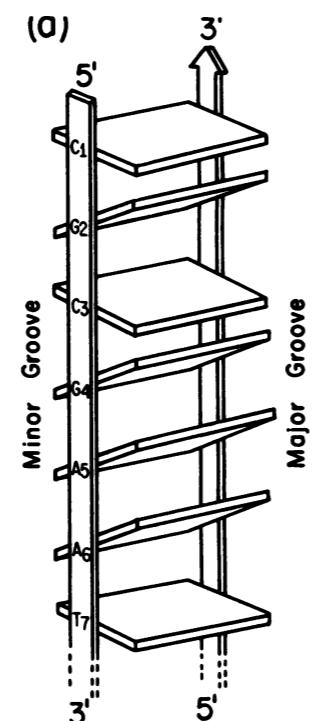


This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

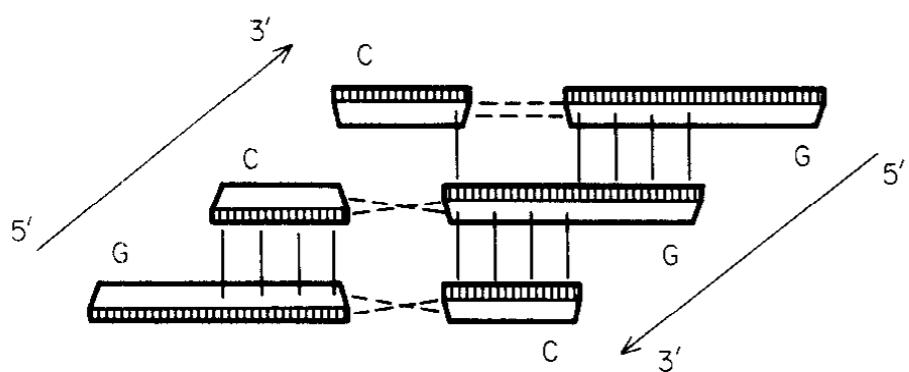
The **eighties** for nucleic acids, the mechanical (engineering) perspective develops.



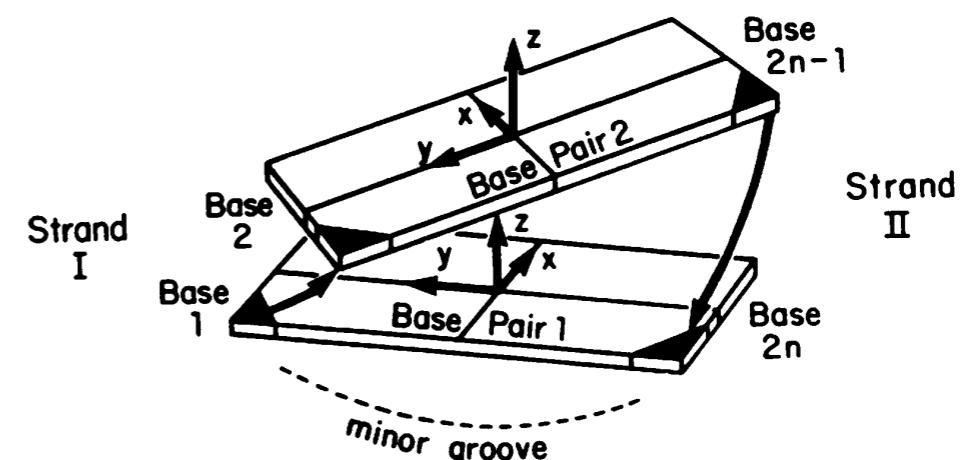
dickerson-drew 1981



calladine-drew 1984

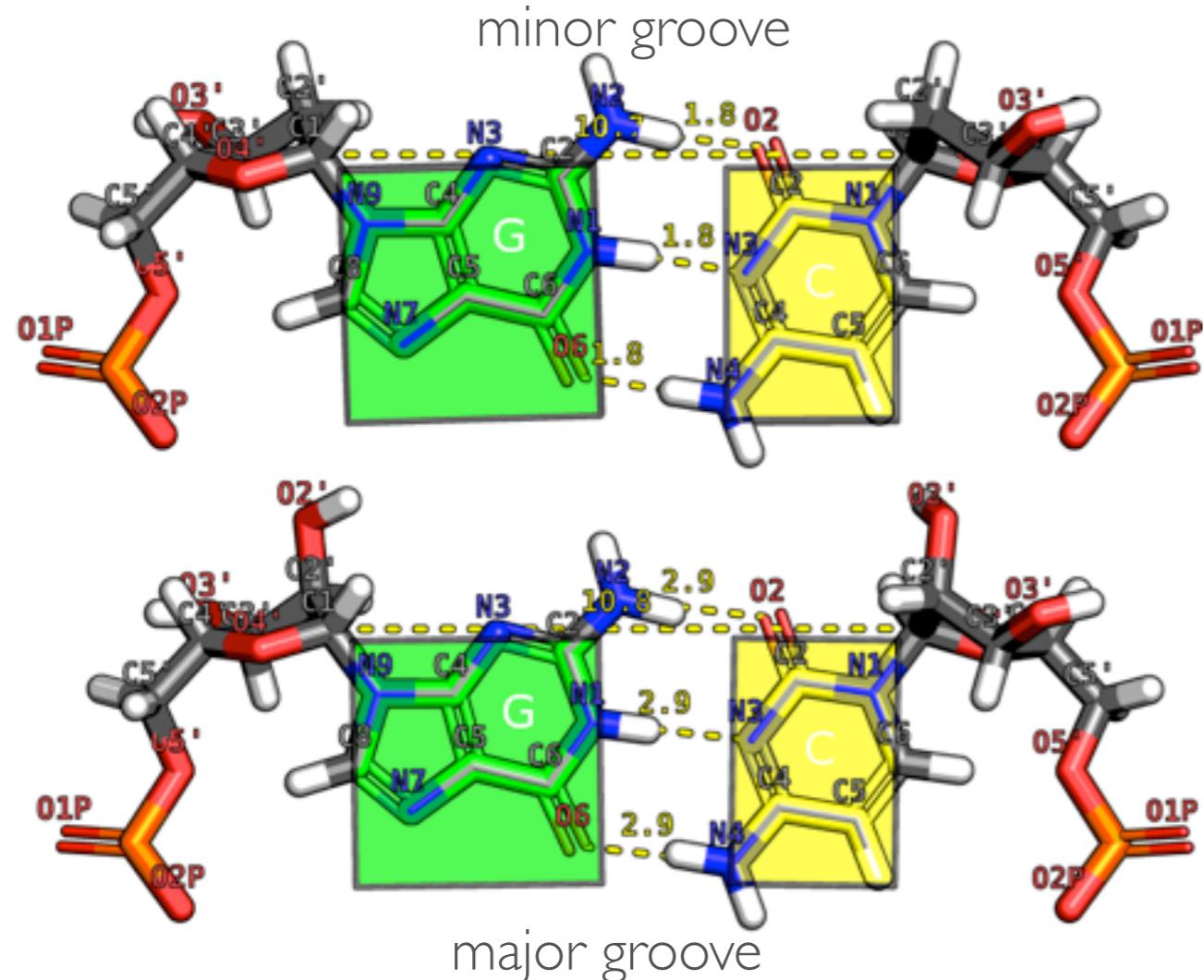


calladine-drew 1986

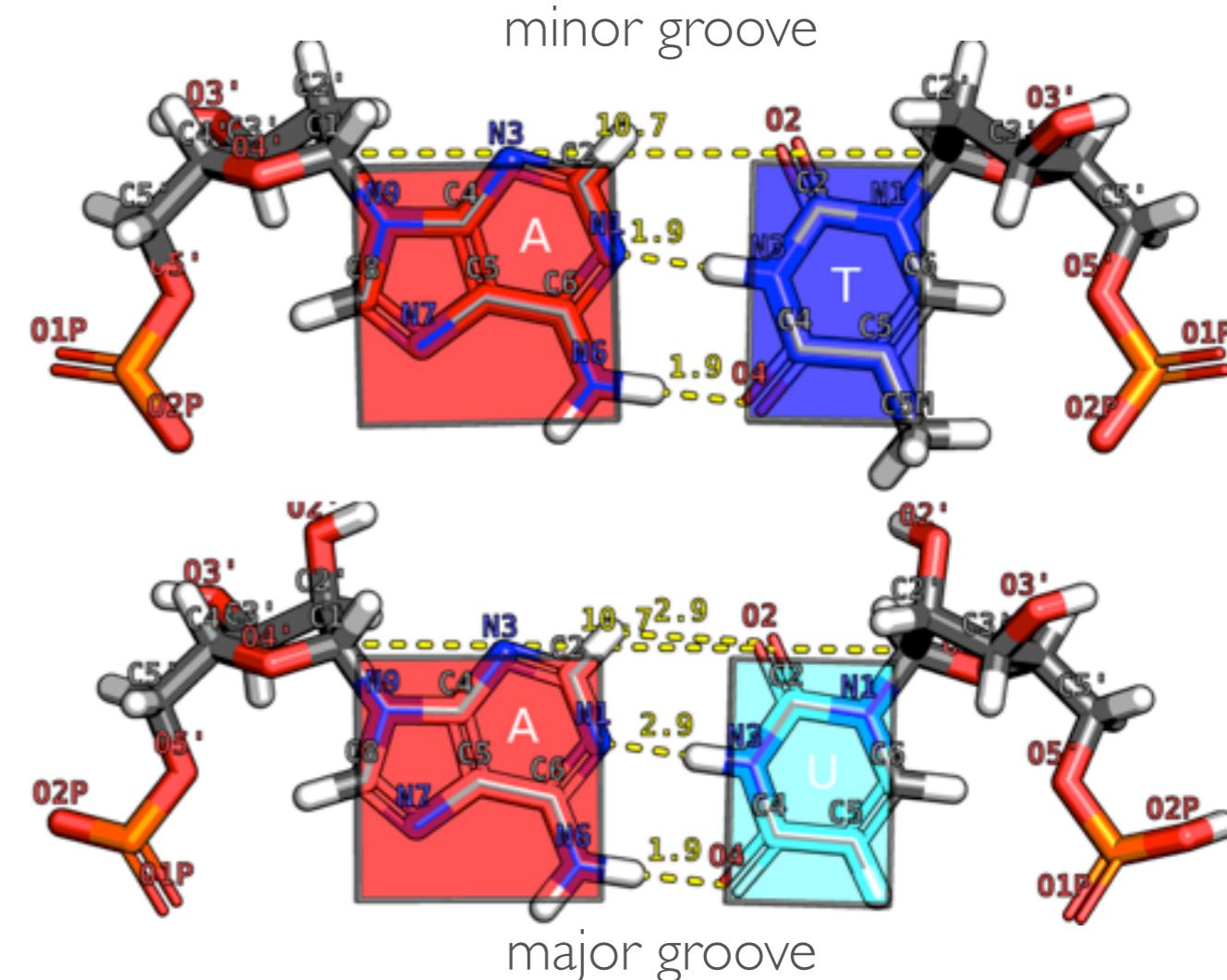


dickerson et al. 1989

Some conventions were necessary, a simple one is coloring. This one is the adopted convention at the PDB/NDB.



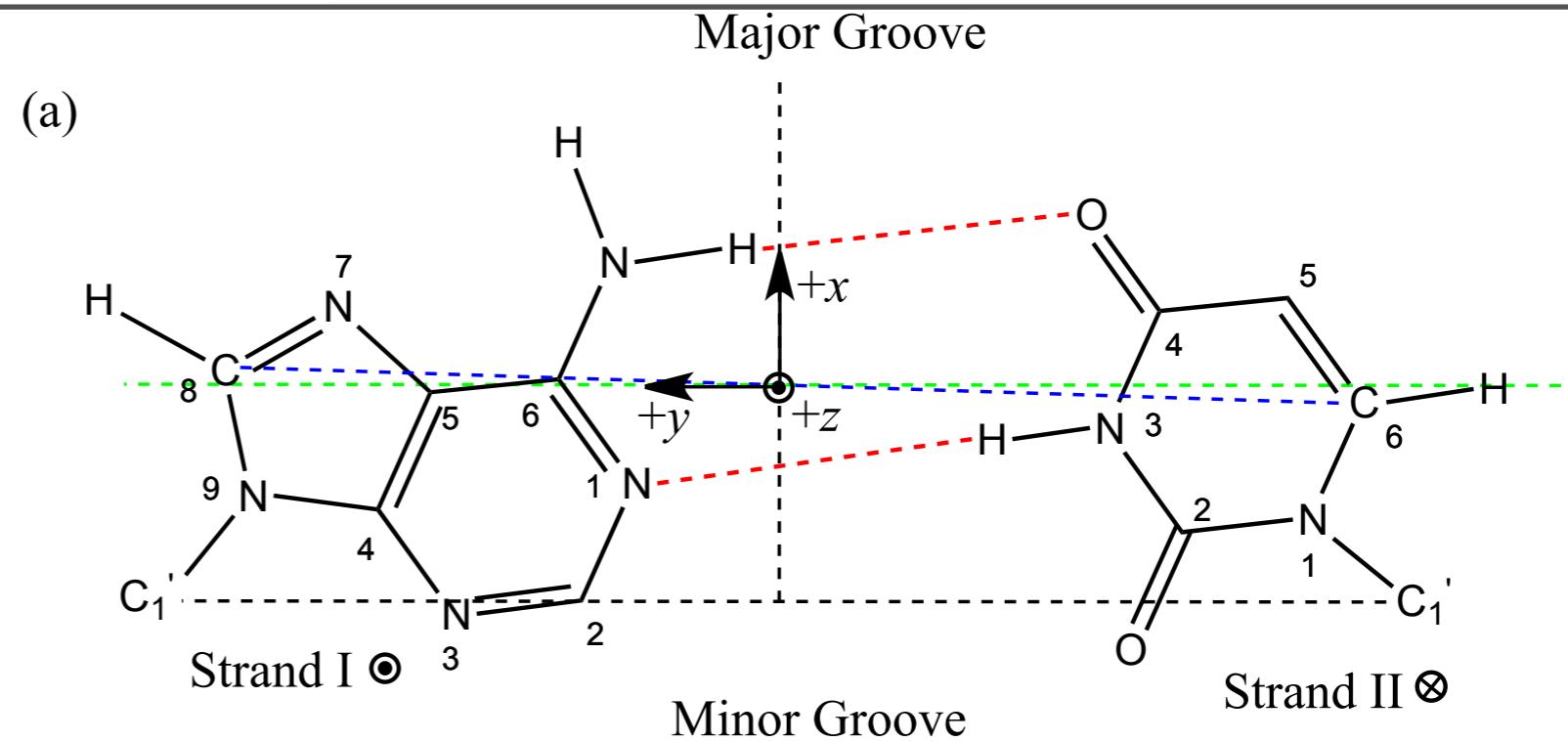
Purines aka **R** = Guanine (green) and Adenine (red)



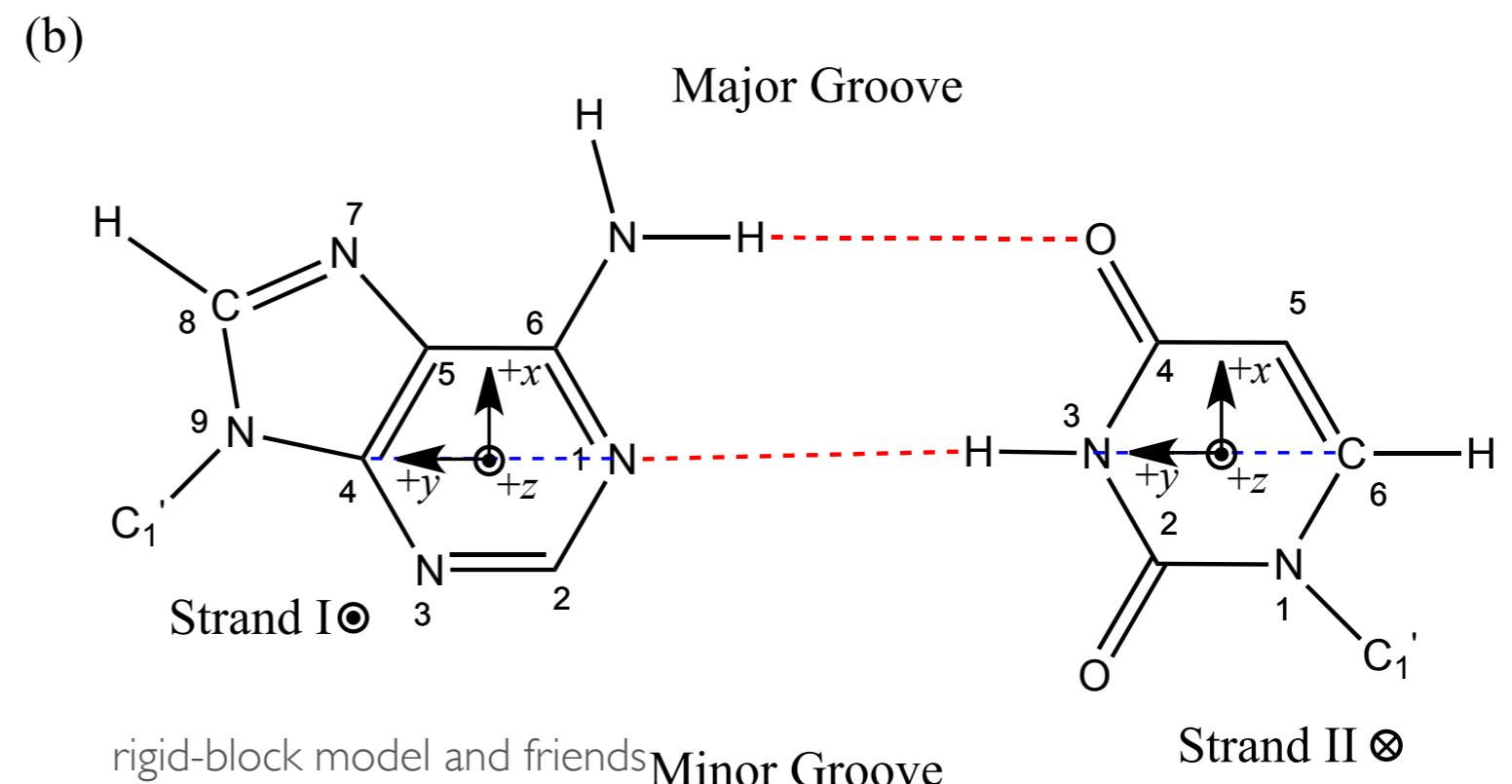
Pyrimidines aka **Y** = Cytosine (yellow), Thymine (blue) and Uracil (cyan)

The standard reference frame for nucleic acid base-pairs.

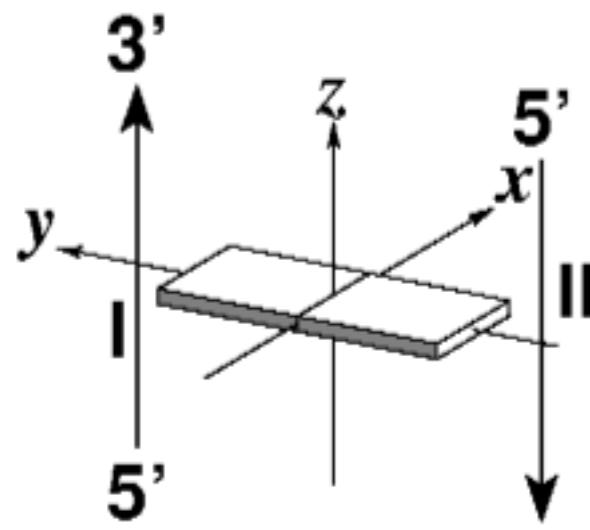
“new”
3DNA, Curves+
1999-2001



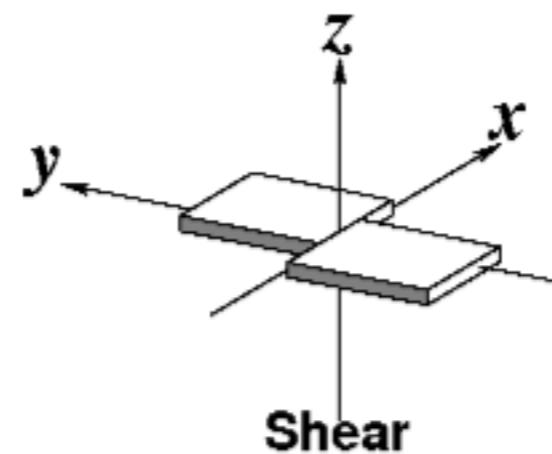
old
SCHNAaP
1997



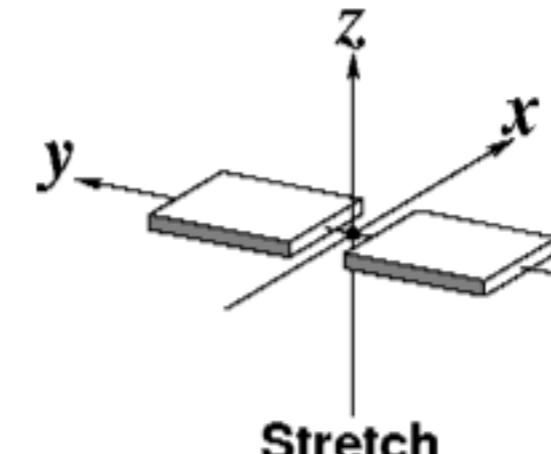
Calladine-Drew rigid-body model for describing nucleic acid base-pairs -- hydrogen bonded pairs.



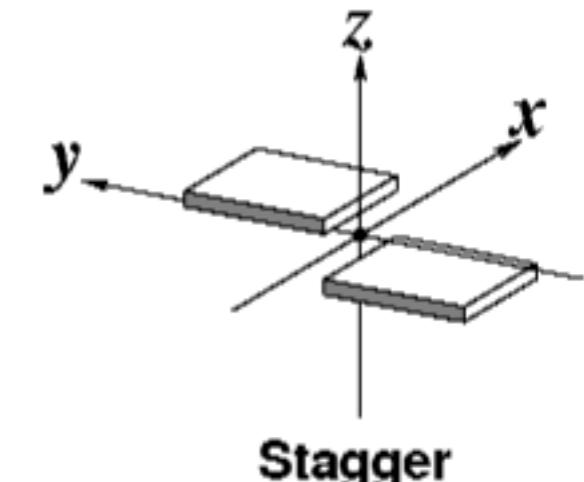
translation



Base-Pair Parameters

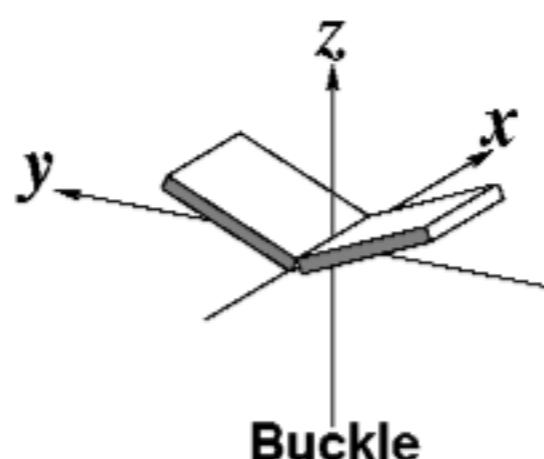


Stretch

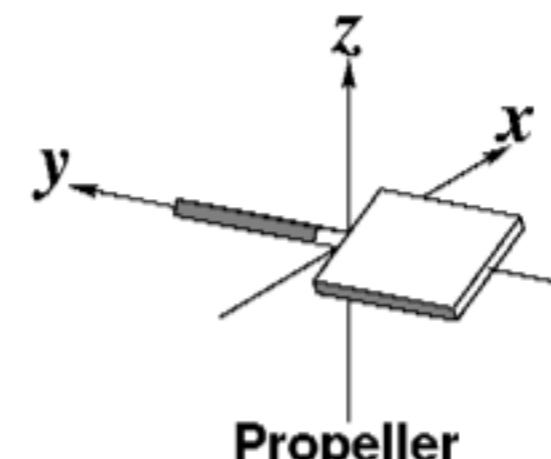


Stagger

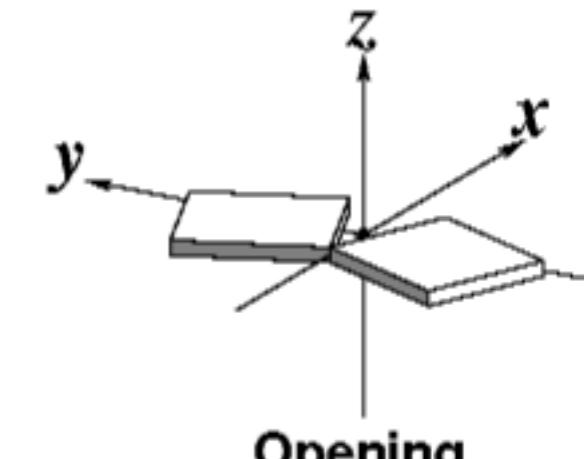
rotation



Buckle



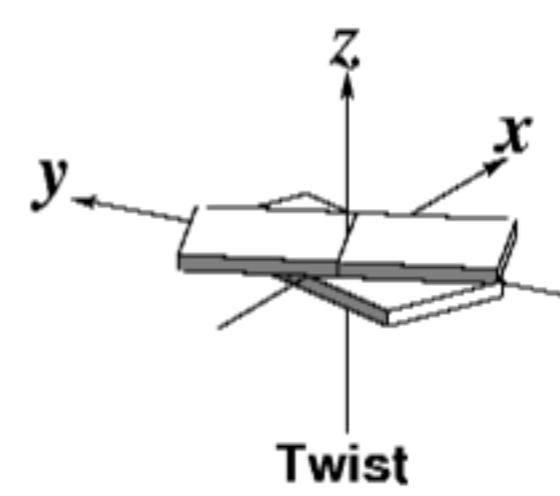
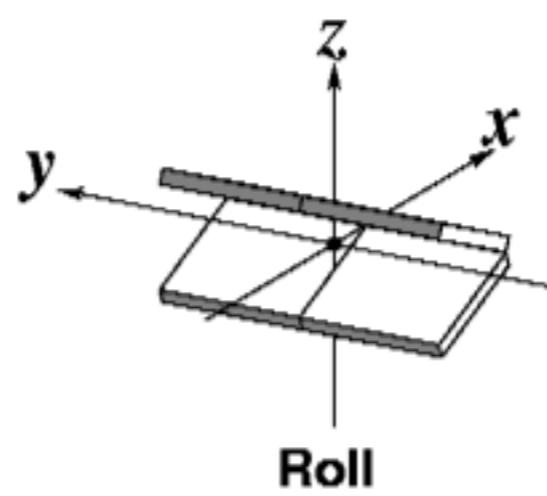
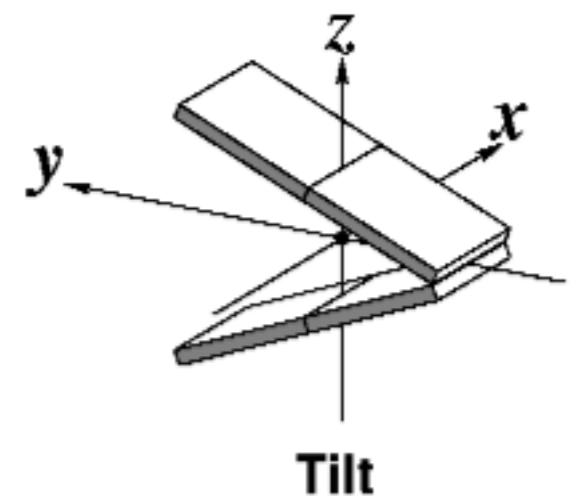
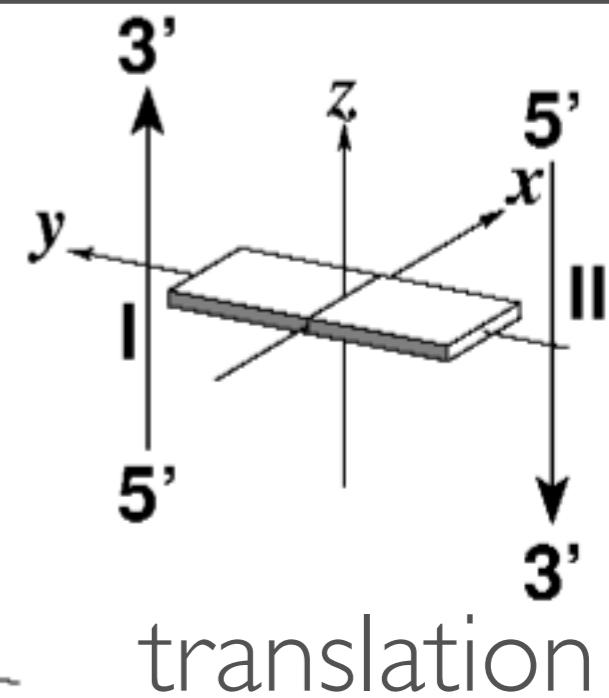
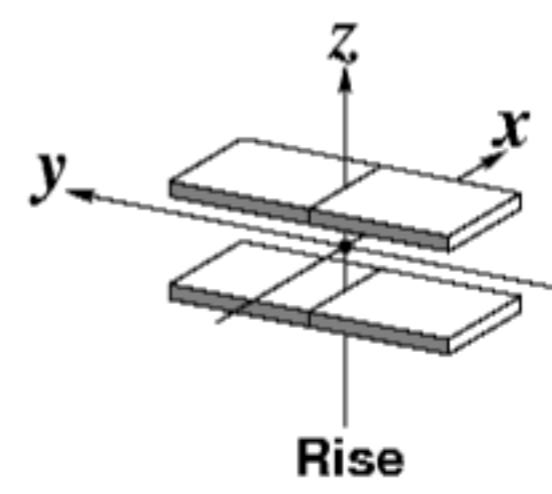
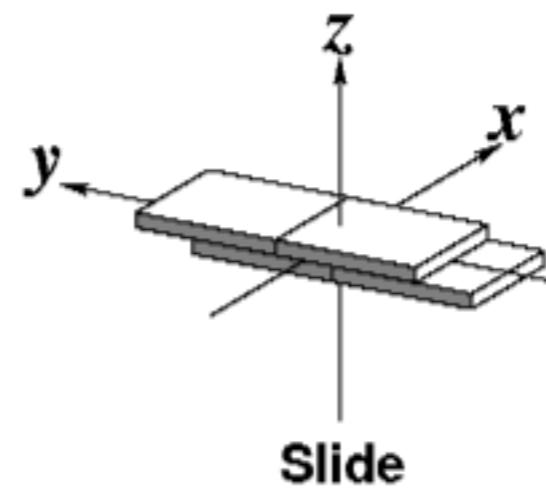
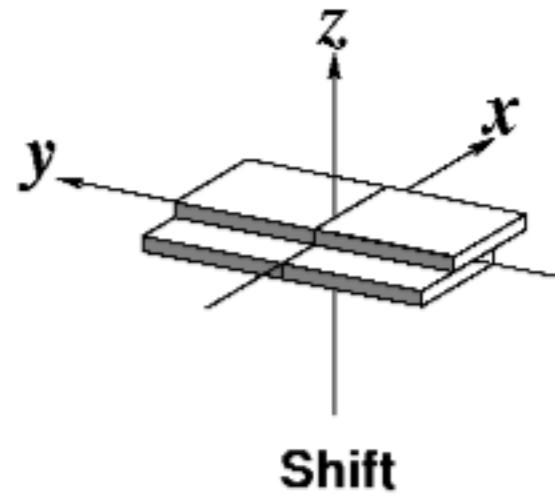
Propeller



Opening

The Calladine-Drew rigid-body model to describe base-pairs making a step among them -- **stacking**.

Base-Pair-**Step** Parameters



rotation

Know programs and servers that compute base-pair and base-pair step parameters.

NEWish

3DNA (Xiang-Jun Lu @ Columbia, Bussemaker lab.)

<http://w3dna.rutgers.edu>

<http://x3dna.org/>

<http://haddock.science.uu.nl/dna/dna.php>

3dnaV2, UNIX preferred (linux, OS-X, freebsd, cygwin “yikes!”)

curves+ (Richard Lavery @ Universite d' Lyon)

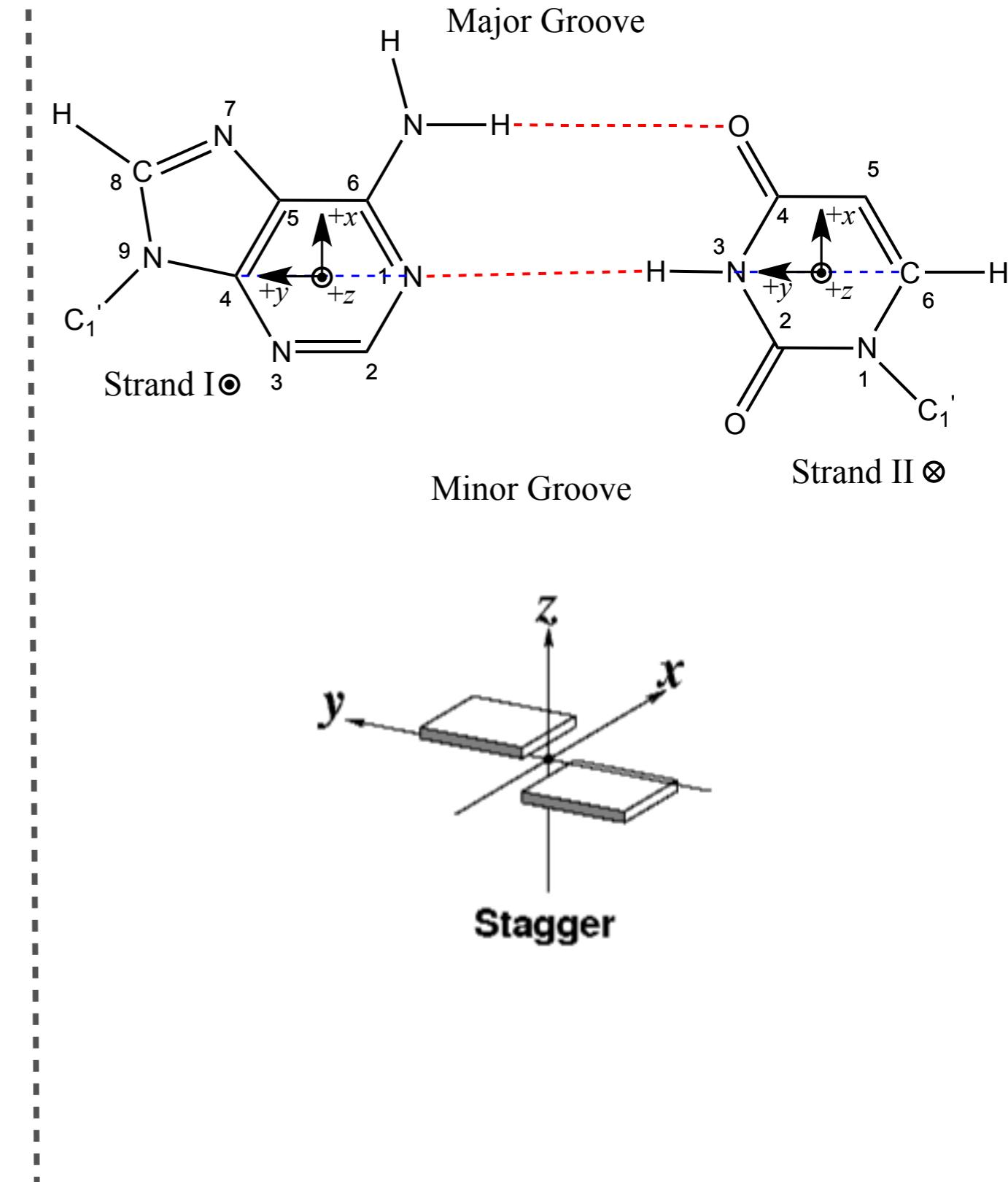
OLDish

FREEHELIX, RNA, SCHNAaP, SCHNArP, compDNA, NUPARM

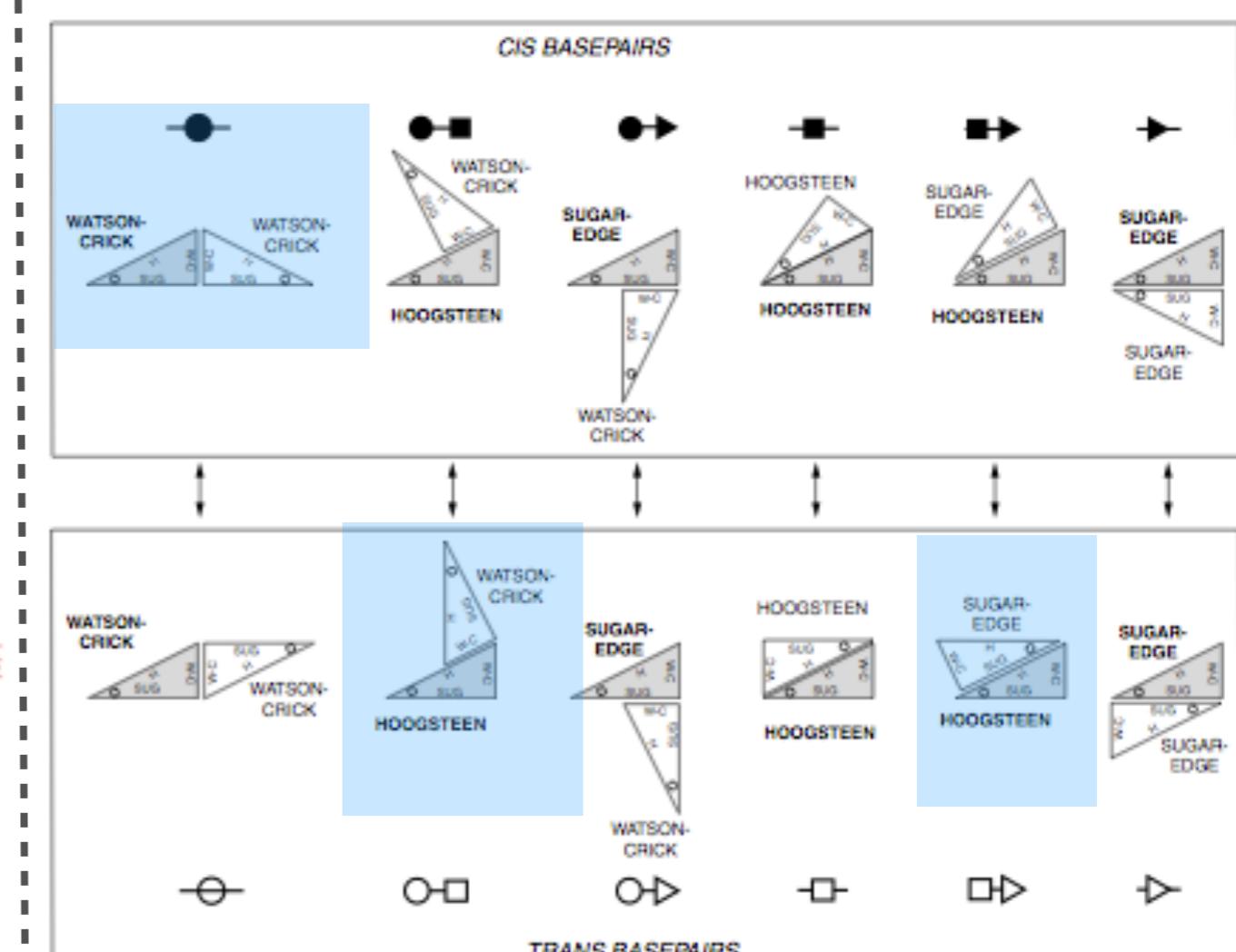
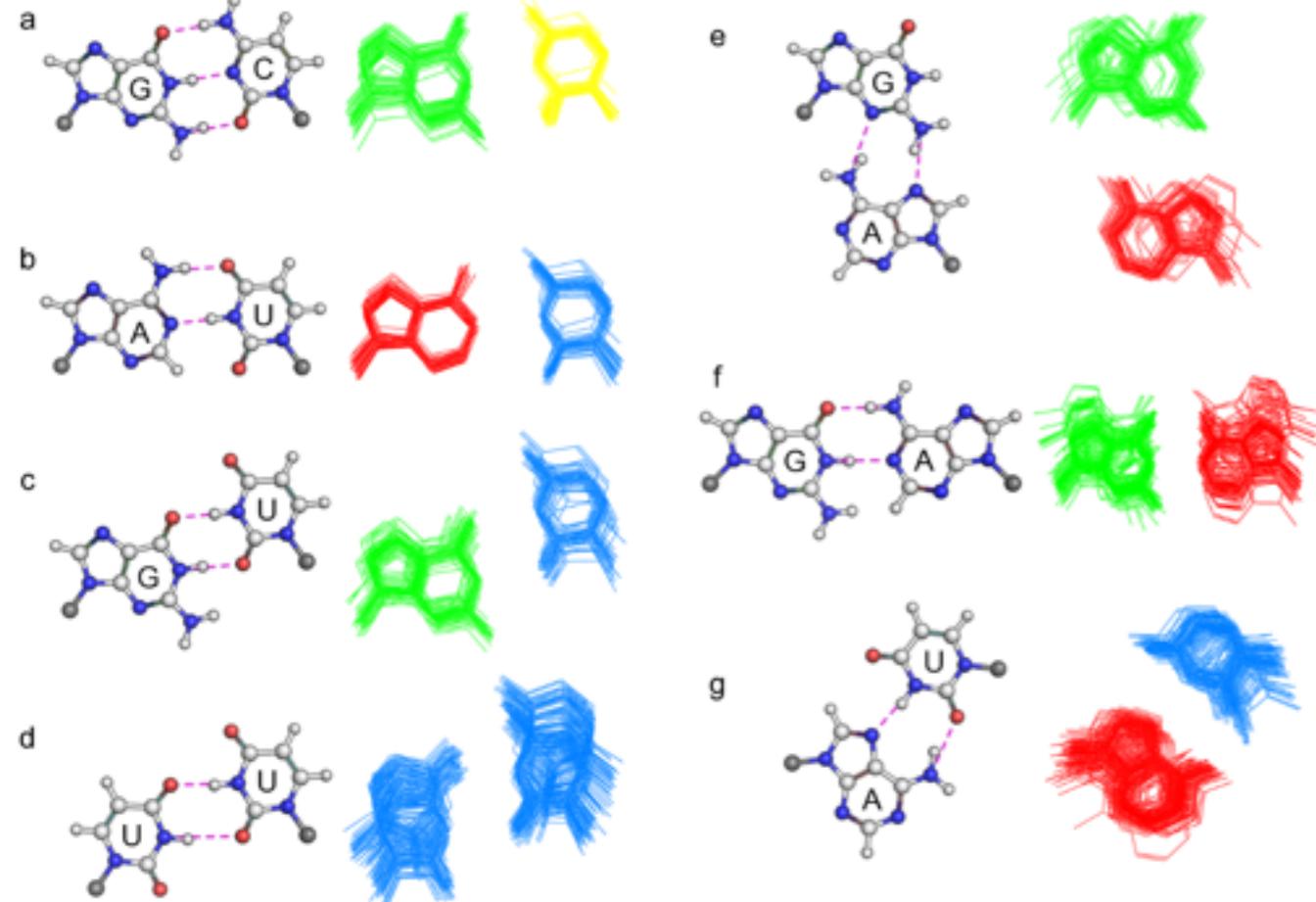
| ***** Local base-pair parameters ***** | | | | | | | |
|---|-------|---------|---------|--------|-----------|---------|--|
| bp | Shear | Stretch | Stagger | Buckle | Propeller | Opening | |
| 1 C-G | -0.42 | -0.27 | 0.06 | 2.76 | -14.20 | -3.67 | |
| 2 G-C | -0.02 | -0.27 | 0.25 | -4.46 | -10.85 | -4.02 | |
| 3 C-G | 0.00 | -0.25 | 0.21 | -6.94 | -3.93 | -2.35 | |
| 4 G-C | -0.37 | -0.44 | -0.18 | 9.31 | -10.39 | -1.30 | |
| 5 A-T | 0.27 | -0.22 | 0.03 | 5.03 | -16.36 | 1.84 | |
| 6 A-T | -0.09 | -0.04 | 0.17 | 3.54 | -18.13 | 5.56 | |
| 7 T-A | 0.32 | -0.12 | 0.13 | 0.83 | -17.70 | 7.93 | |
| 8 T-A | 0.25 | -0.21 | -0.10 | -1.33 | -17.67 | 0.83 | |
| 9 C-G | -0.02 | -0.25 | -0.06 | -10.18 | -17.25 | -0.87 | |
| 10 G-C | 0.09 | -0.28 | 0.27 | 1.67 | -5.31 | -1.13 | |
| 11 C-G | 0.07 | -0.28 | 0.59 | -3.96 | -18.05 | -5.62 | |
| 12 G-C | -0.53 | -0.11 | 0.26 | 6.60 | 1.96 | -3.86 | |
| ave. | -0.04 | -0.23 | 0.14 | 0.24 | -12.32 | -0.55 | |
| s.d. | 0.28 | 0.10 | 0.21 | 5.78 | 6.73 | 4.04 | |
| ***** Local base-pair step parameters ***** | | | | | | | |
| step | Shift | Slide | Rise | Tilt | Roll | Twist | |
| 1 CG/CG | -0.36 | 0.15 | 3.52 | -3.40 | 6.42 | 40.31 | |
| 2 GC/GC | 0.50 | 0.23 | 3.52 | 0.80 | -4.73 | 38.15 | |
| 3 CG/CG | -0.32 | 0.69 | 3.04 | 3.63 | 7.95 | 24.47 | |
| 4 GA/TC | 0.01 | 0.07 | 3.36 | -2.68 | 3.16 | 40.90 | |
| 5 AA/TT | 0.10 | -0.31 | 3.32 | -0.70 | 0.95 | 35.35 | |
| 6 AT/AT | 0.33 | -0.60 | 3.34 | 1.83 | -2.75 | 34.76 | |
| 7 TT/AA | -0.31 | -0.18 | 3.32 | 2.96 | 0.73 | 35.39 | |
| 8 TC/GA | 0.02 | -0.03 | 3.39 | 0.33 | -0.05 | 39.27 | |
| 9 CG/CG | 0.38 | 0.86 | 3.24 | -3.29 | 3.86 | 29.40 | |
| 10 GC/GC | -1.30 | 0.42 | 3.68 | -4.68 | -12.20 | 40.78 | |
| 11 CG/CG | 0.77 | 0.06 | 3.23 | 3.14 | -3.09 | 32.62 | |
| ave. | -0.02 | 0.12 | 3.36 | -0.19 | 0.02 | 35.58 | |
| s.d. | 0.56 | 0.42 | 0.17 | 2.96 | 5.65 | 5.19 | |

What is the criteria that 3DNA uses to define a base-pair, and could I change it? look at misc_3dna.par .

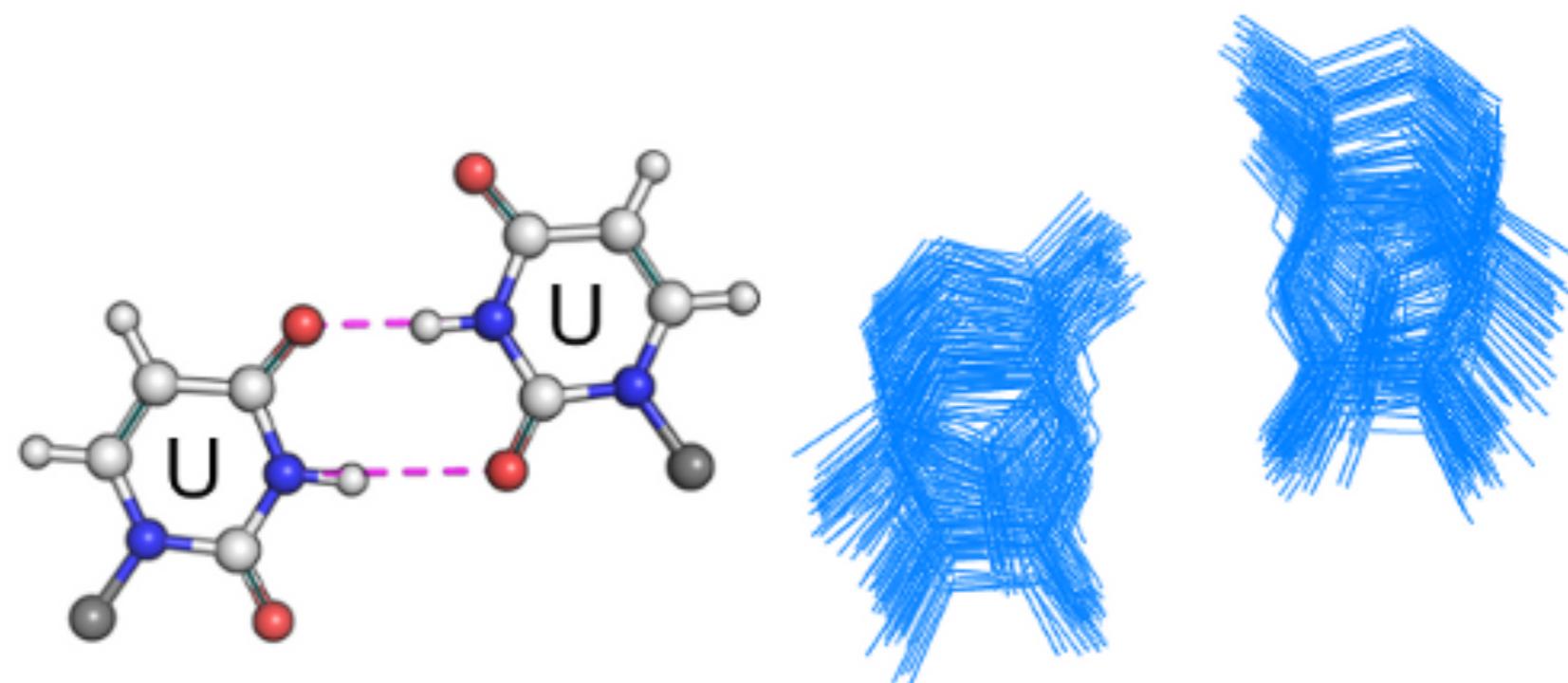
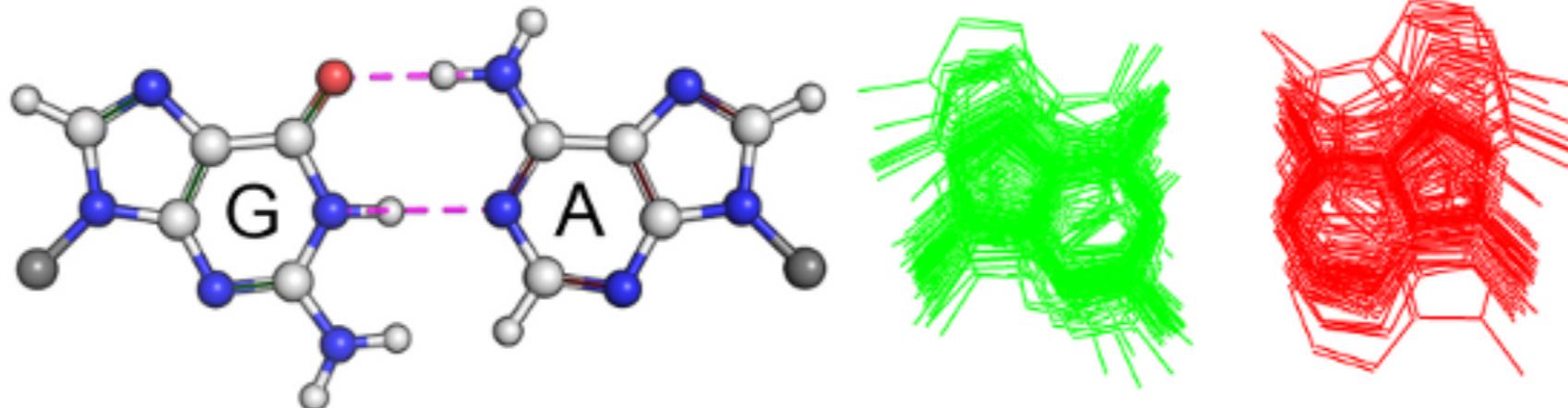
- The distance between the origins of the two bases (as defined by their standard reference frames) must be less than certain limit (15.0 \AA by default) - otherwise, they would be too far away to be called a pair.
- The vertical separation (i.e., stagger) between the two bases must be less than certain limit (2.5 \AA by default) - otherwise, they would be stacking instead of pairing.
- The angle between the two base z-axes (i.e., their normal vectors) is less than a cut-off (65.0° by default).
- There is at least one pair of nitrogen/oxygen base atoms that are within a H-bonding cut off distance (4.0 \AA by default).



Some results using 3DNA's base-pair recognition for RNA. Non-canonical base-pairs are more deformable than canonical base-pairs.



Most deformed base-pairs can be seen as oversized or undersized pieces in a puzzle.



If canonical RNA base-pairs (including G•U) are stacked in a base-pair step, then there are 21 unique possible steps.

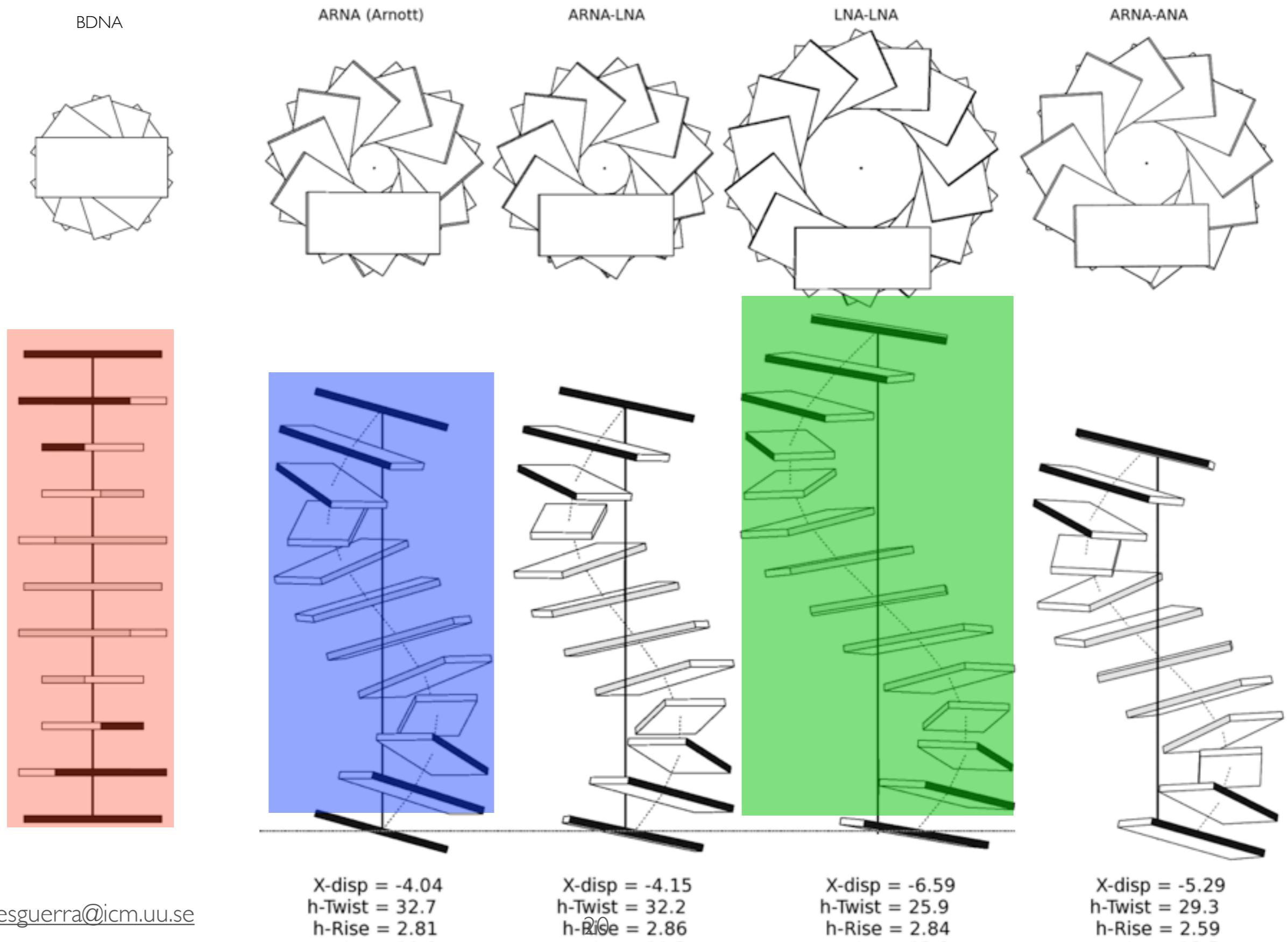
| $i \backslash i+1$ | A·U | U·A | G·C | C·G | GU | U·G |
|--------------------|-----|-----|-----|-----|----|-----|
| U·A | | | | | | |
| A·U | | | | | | |
| G·C | | | | | | |
| U·G | | | | | | |
| GU | | | | | | |

There is a database with a classification of the unique base-pair steps of RNA. Still needs more structural input.

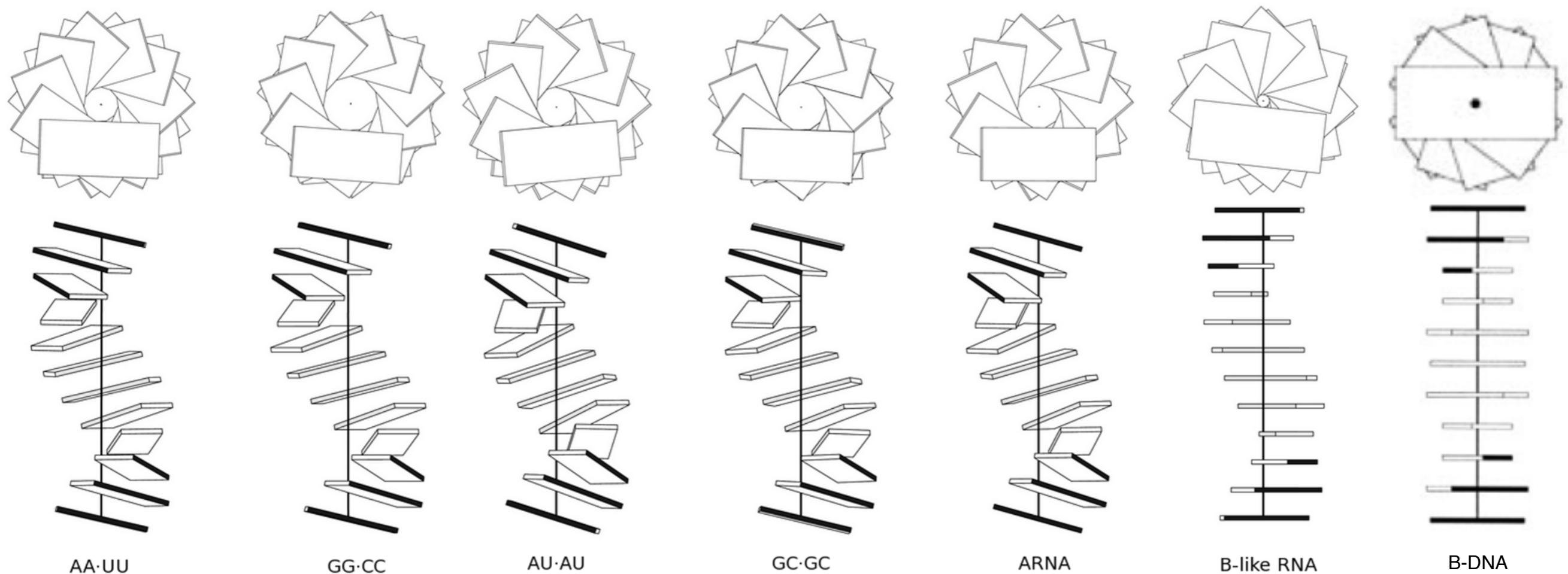
- With average base-pair-step parameters, average base-pair steps can be rebuilt.
- Go to <http://rnasteps.rutgers.edu>, download parameters and rebuild in <http://w3dna.rutgers.edu>.

| RNA BASE-PAIR STEP PARAMETERS | | | | | | | | | | |
|---|-------|-------|-------|-------|------|------|------|-------|--------|------|
| Stack Type | Step | Count | Shift | Slide | Rise | Tilt | Roll | Twist | Volume | RMSD |
| RR | GG.CC | 1274 | -0.01 | -1.85 | 3.30 | 0.0 | 7.4 | 31.1 | 2.0 | 0.35 |
| YR | UG.CA | 700 | 0.03 | -1.59 | 3.16 | 0.2 | 10.6 | 30.7 | 1.2 | 0.31 |
| RY | GC.GC | 587 | 0.02 | -1.56 | 3.20 | 0.0 | 4.2 | 33.5 | 1.3 | 0.34 |
| YR | CG.CG | 562 | 0.05 | -1.84 | 3.29 | 0.3 | 10.8 | 29.1 | 2.2 | 0.35 |
| RR | AG.CU | 547 | 0.06 | -1.66 | 3.25 | -0.1 | 8.2 | 30.1 | 0.5 | 0.35 |
| RY | AC.GU | 546 | 0.14 | -1.48 | 3.22 | 0.3 | 4.9 | 32.7 | 1.5 | 0.33 |
| RR | GA.UC | 484 | 0.02 | -1.61 | 3.20 | 0.0 | 5.9 | 32.6 | 2.2 | 0.36 |
| RR | AA.UU | 241 | -0.08 | -1.38 | 3.16 | -0.4 | 7.1 | 31.6 | 1.0 | 0.32 |
| RY | GC.GU | 237 | 0.06 | -1.25 | 3.21 | 0.0 | 4.4 | 41.4 | 0.6 | 0.28 |
| RR | GG.CU | 180 | 0.01 | -1.76 | 3.31 | -0.2 | 5.0 | 37.1 | 2.0 | 0.39 |
| « previous 1 2 3 next » | | | | | | | | | | |

With 3DNA a mechanical perspective of helicoidal polymers is easier to understand and visualize. #blocview -o file.pdb



An example of what one can do with the rigid-body model is seeing the subtle effects of RNA sequence on helical regions.



homopolymers block copolymers

slide = -1.5
rise = 3.30
tilt = 0.0
roll = 0.0
twist = 31.6

slide = 0.0
rise = 3.36
tilt = 0.0
roll = 0.0
twist = 36.0

It is of course also useful for the analysis of simulation results, even at the microsecond scale.

Published online 24 July 2008

Nucleic Acids Research, 2008, Vol. 36, No. 15 4941–4955
doi:10.1093/nar/gkn473

Molecular dynamics of a κB DNA element: base flipping via cross-strand intercalative stacking in a microsecond-scale simulation

Cameron Mura^{1,*} and J. Andrew McCammon^{1,2}

¹Department of Chemistry and Biochemistry and Center for Theoretical Biological Physics, University of California, San Diego, La Jolla, CA 92093-0365 and ²Howard Hughes Medical Institute and Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093-0636, USA

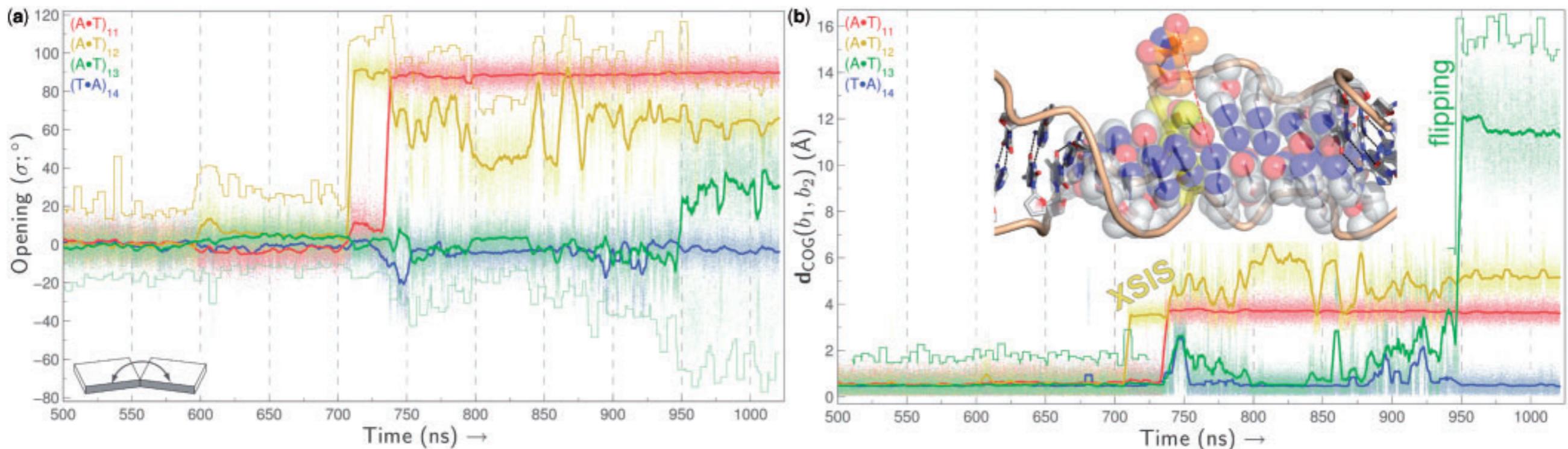
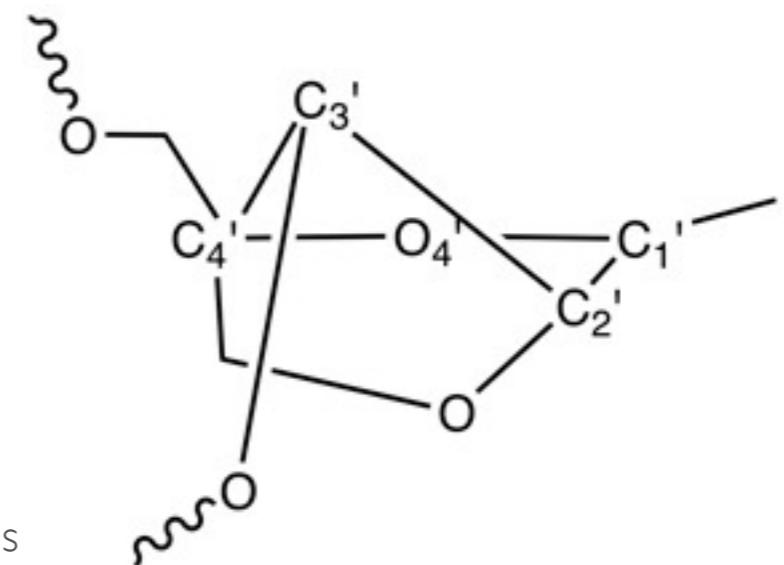


Figure 6. Correlated disruptions at the junction of κB half sites: resolution of XSIS via spontaneous base flipping. The spontaneous $(A \cdot T)_{13}$ base flipping event at roughly 950 ns (≈ 200 ns after the onset of XSIS) can be characterized by the standard base pair *Opening* (σ) parameter (a), as well as by the distance between the complementary bases in a given pair (b). The latter measure (d_{COG}) is taken as the distance between the centers of geometry of the bases, and is calculated from the *Shear*, *Stretch* and *Stagger* as $\sqrt{S_x^2 + S_y^2 + S_z^2}$. A sample structure of the flipped state is shown (b; inset), with the distance between the extrahelical base ($T_{2,8}$) and its cross-strand partner ($A_{1,13}$) indicated as a dashed magenta-colored line. Other diagrammatic and graphical conventions are as in Figures 3 and 4.

CHARMM simulations of a modified DNA triplex where normal ribose sugars are modified to so-called Locked Nucleic Acids

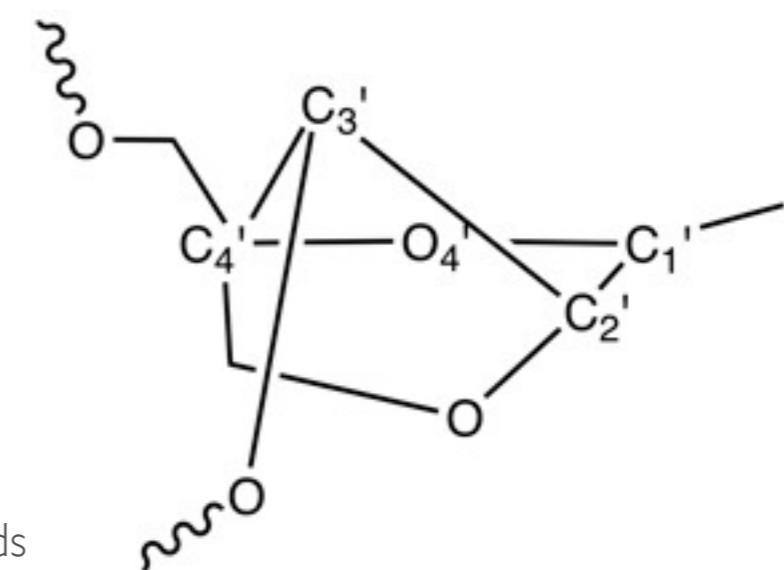
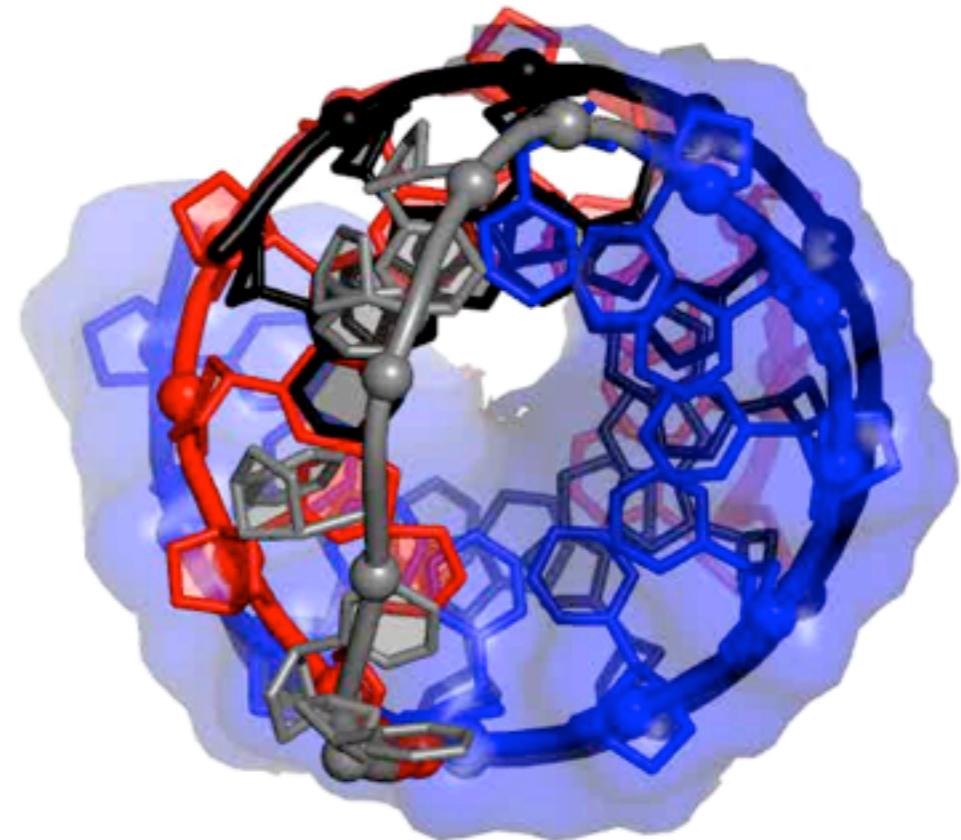
Of course shorter simulations can also be analyzed. In this case a 40ns dynamics.



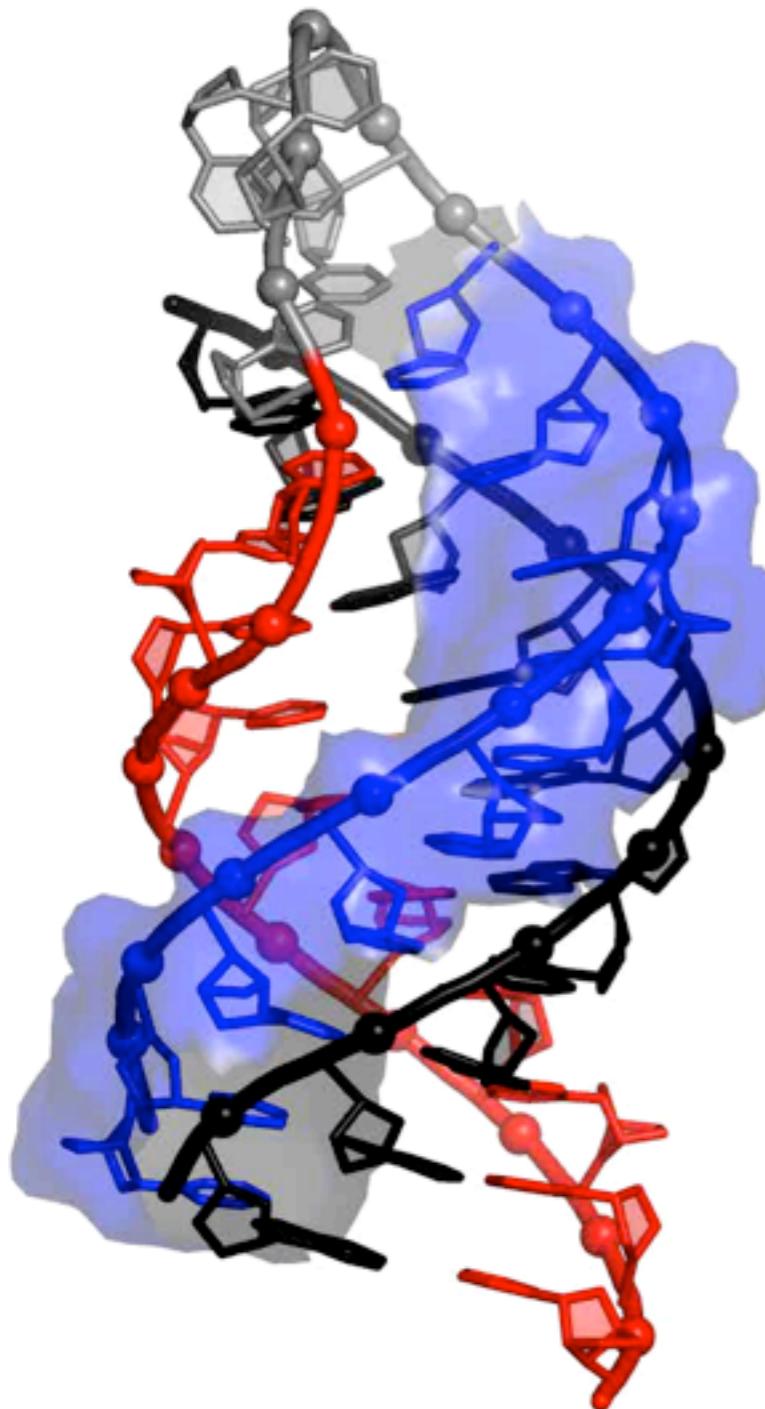
CHARMM simulations of a modified DNA triplex where normal ribose sugars are modified to so-called Locked Nucleic Acids



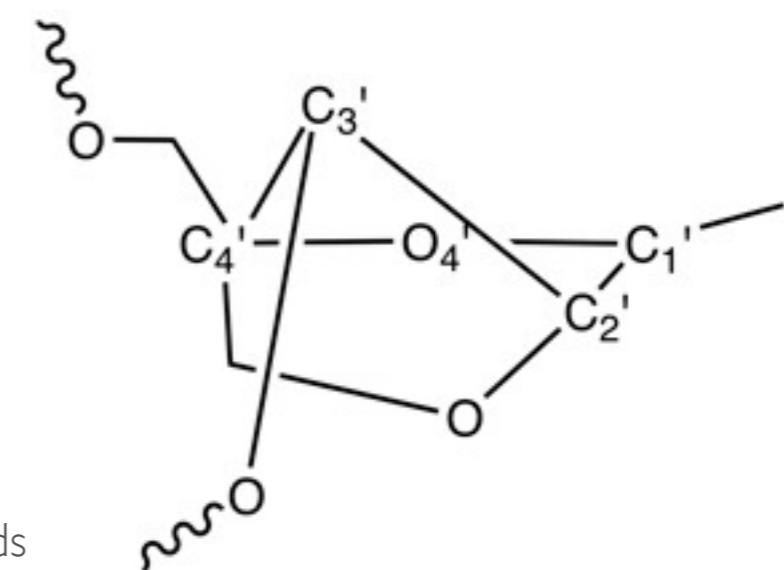
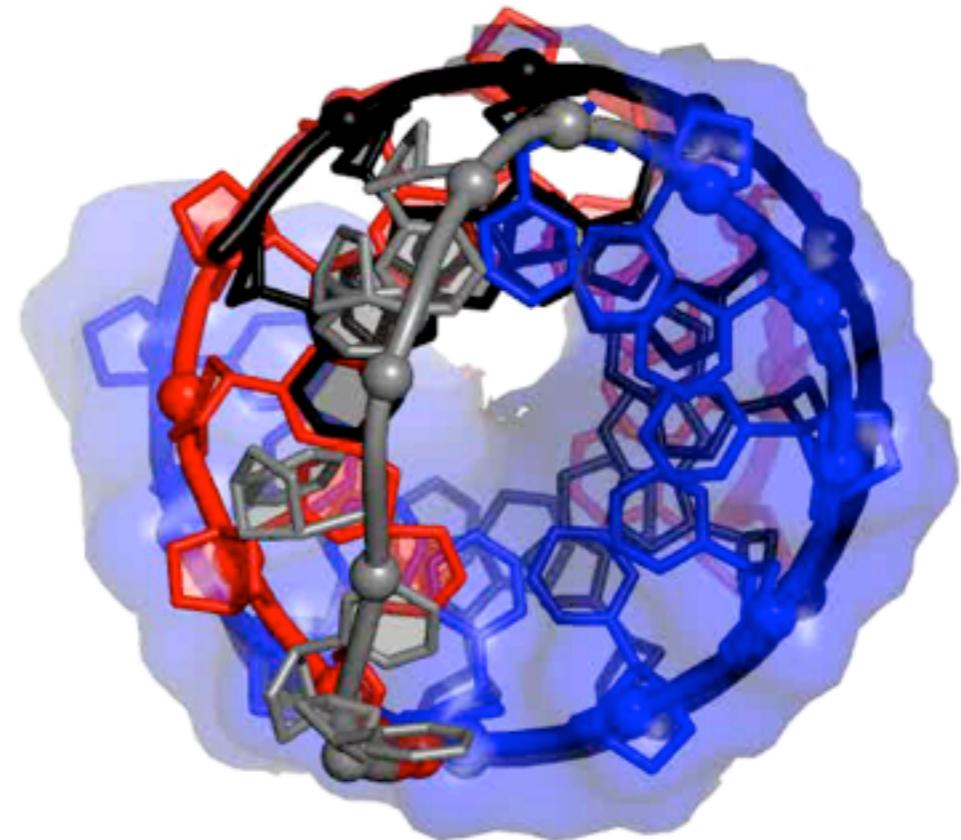
Of course shorter simulations can also be analyzed. In this case a 40ns dynamics.



CHARMM simulations of a modified DNA triplex where normal ribose sugars are modified to so-called Locked Nucleic Acids

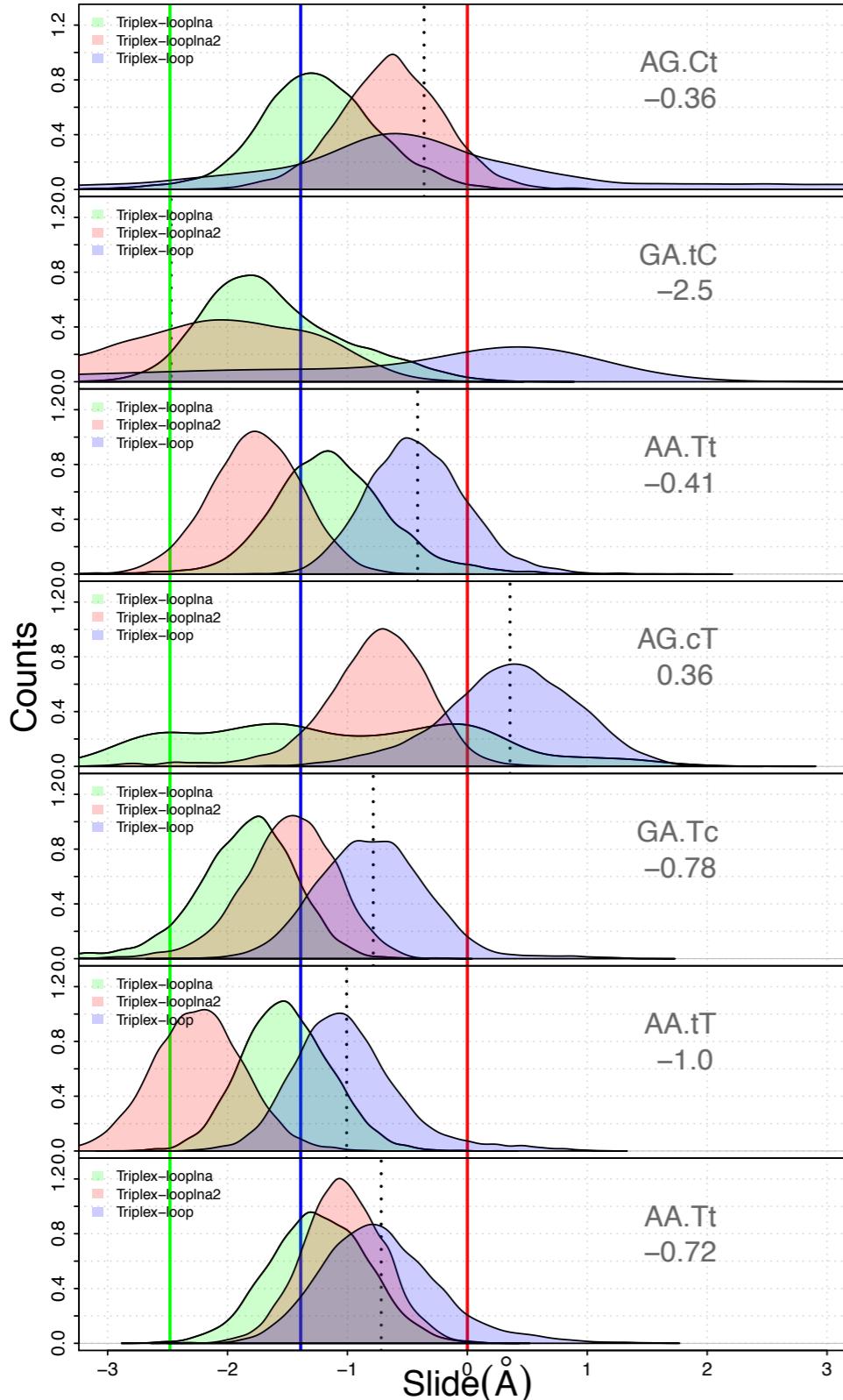


Of course shorter simulations can also be analyzed. In this case a 40ns dynamics.

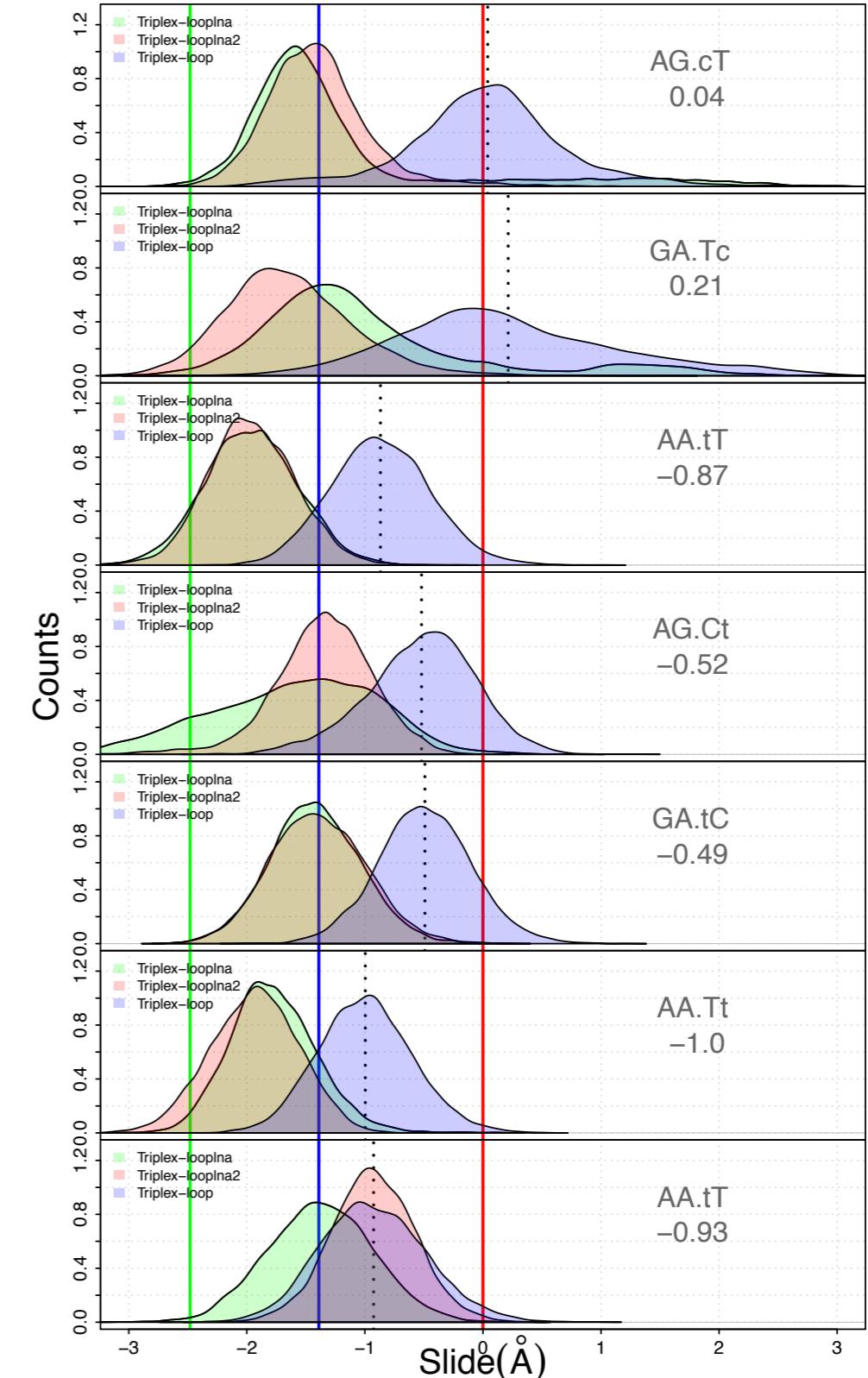


Sequence effects can be clearly seen in cases where nucleic acids are modified at the step-scale. Not only W-C, but also Hoogsteen.

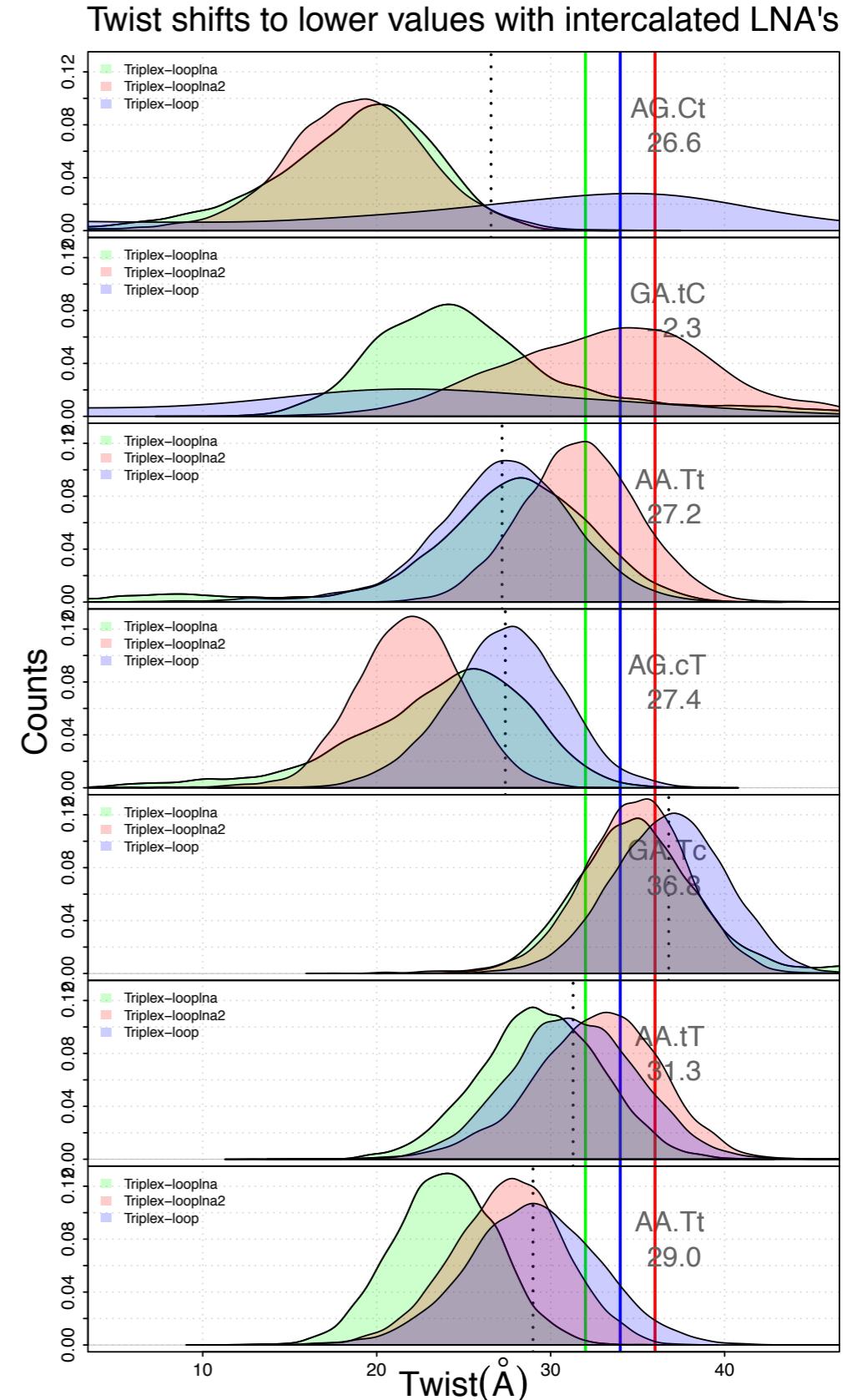
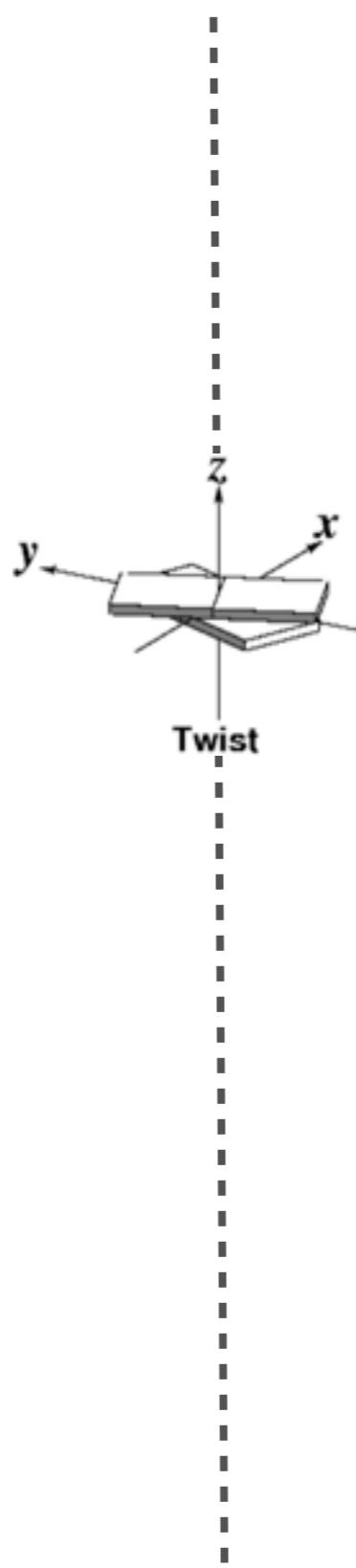
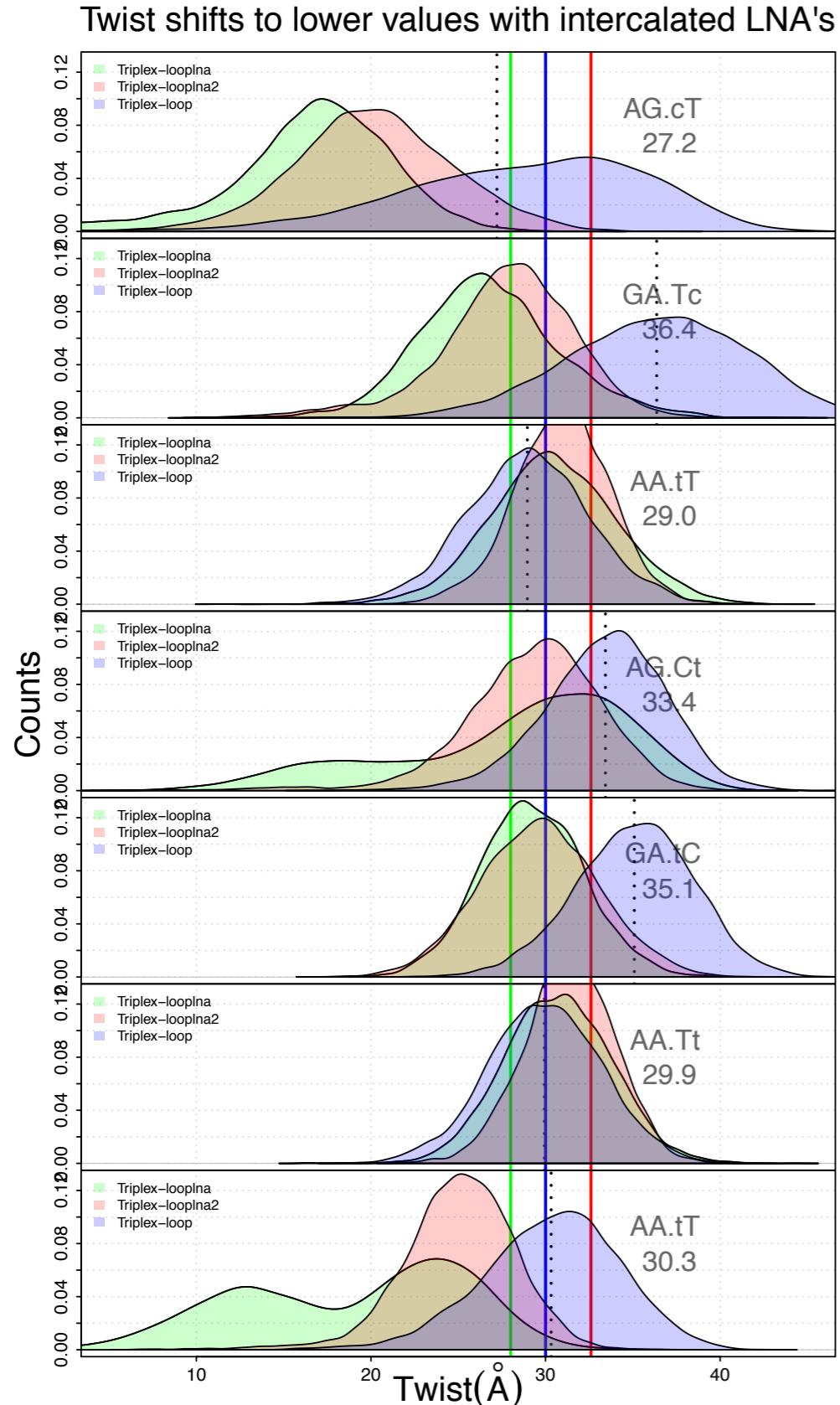
Slide shifts to lower values with intercalated LNA's



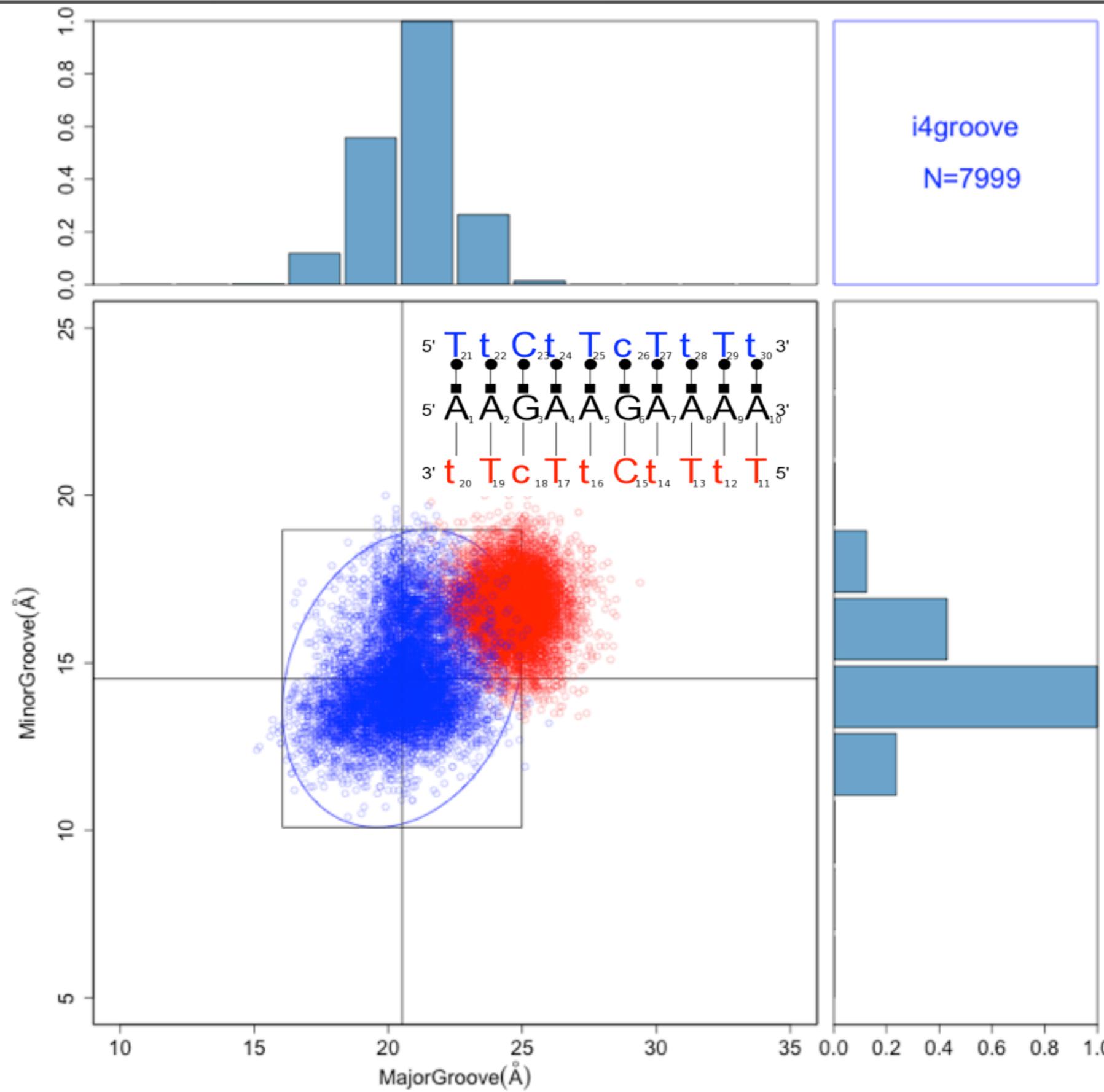
Slide shifts to lower values with intercalated LNA's



Here, on modification of the sugars within a strand in triplex DNA
we clearly see how Twist depends on sequence. W-C (left)



3DNA apart from the rigid-block model gives other geometrical properties such as the groove widths.



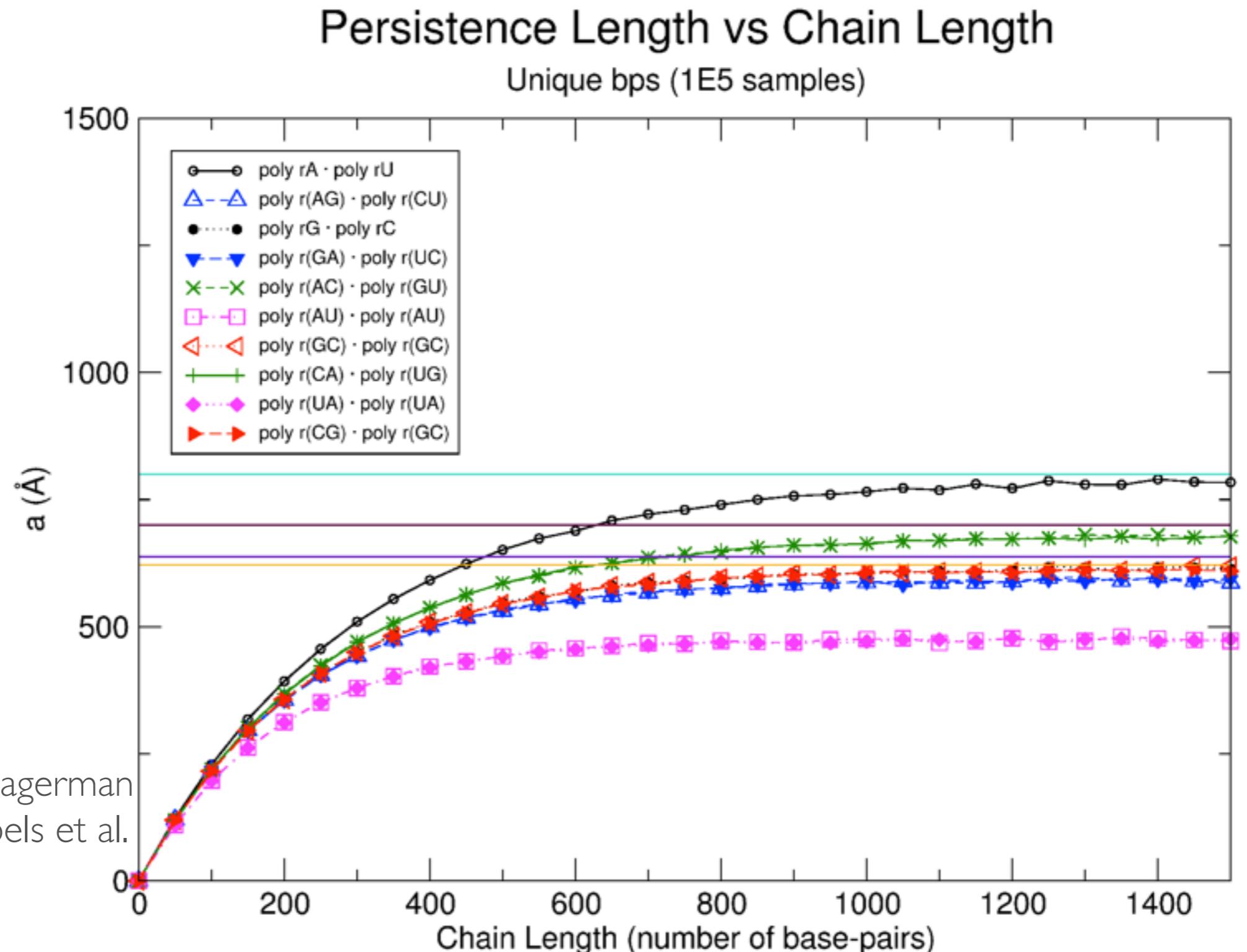
With a more polymer chemist flavor one can compute the persistence length based on the statistics of block parameters.

Porod-Kratky

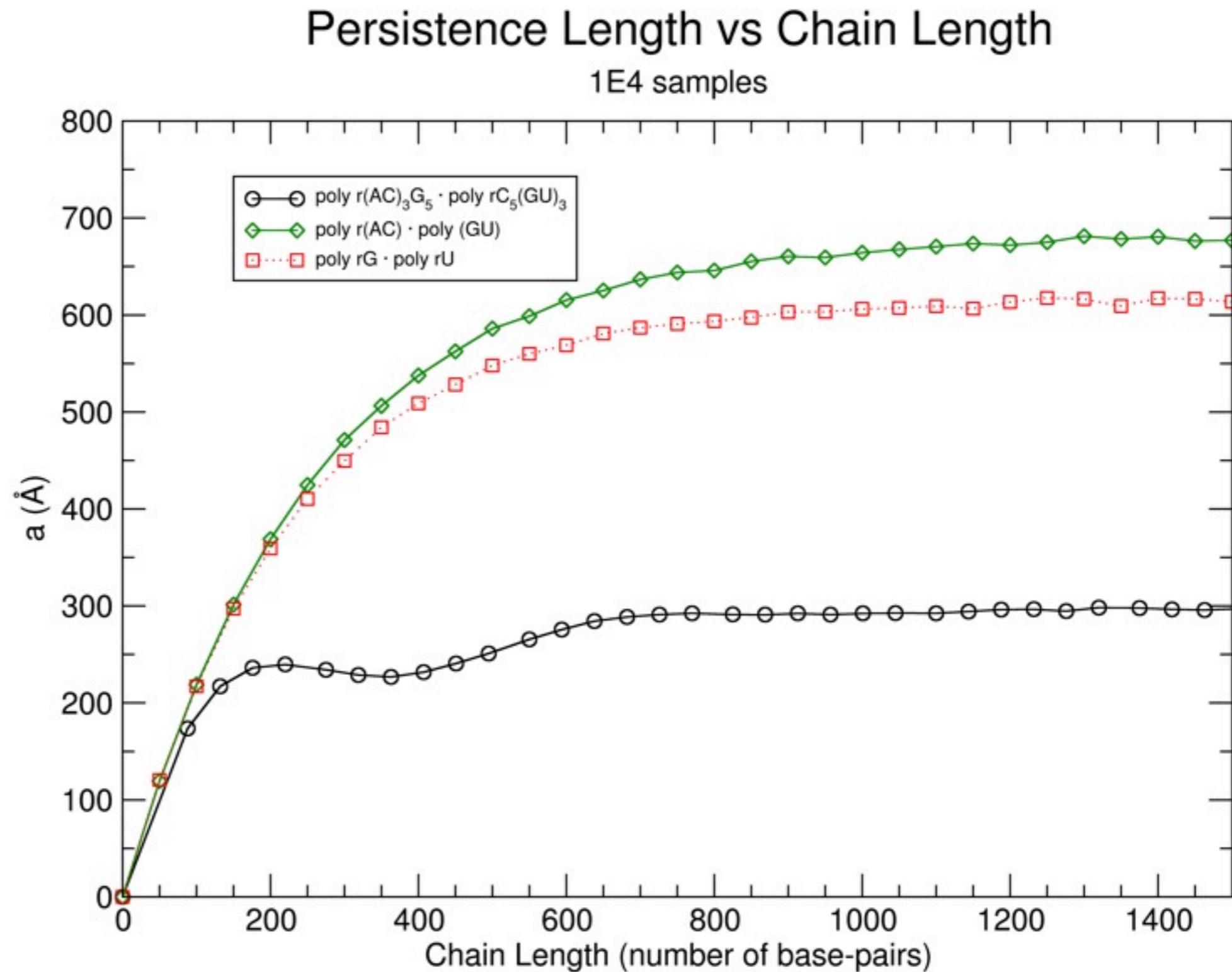
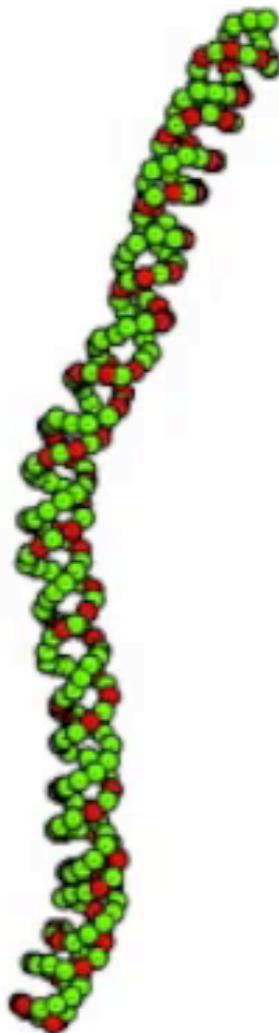
Persistence length is “the average sum of the projections of all bonds $j \geq i$ on bond i in an indefinitely long chain. The bond i is taken to be remote from either end of the chain, i.e., $1 \ll i \ll n$ ”. Paul J. Flory, Statistical Mechanics of Chain Molecules. 1969

| Polymer | a (nm) | Citation |
|-----------------------|----------|--------------------------|
| Polymethylene | 0.6 | Flory ^a |
| Polystyrene | 0.9 | Flory ^a |
| Polyglycine | 0.6 | Flory ^b |
| Poly-L-alanine | 2 | Flory ^b |
| Poly-L-proline | 22 | Cantor and Schimmel [11] |
| B-DNA | 53 | Rivetti [12] |
| A-RNA | 62-64 | Abels [13] |
| α -helix | 80-100 | Lakkaraju [14] |
| Coiled-coil | 150-300 | Lakkaraju [14] |
| Neurofilament | 500 | Nelson [2] |
| Intermediate filament | 1000 | Lakkaraju [14] |
| F-actin | 17000 | Lakkaraju [14] |
| Microtubule | 5200000 | Lakkaraju [14] |

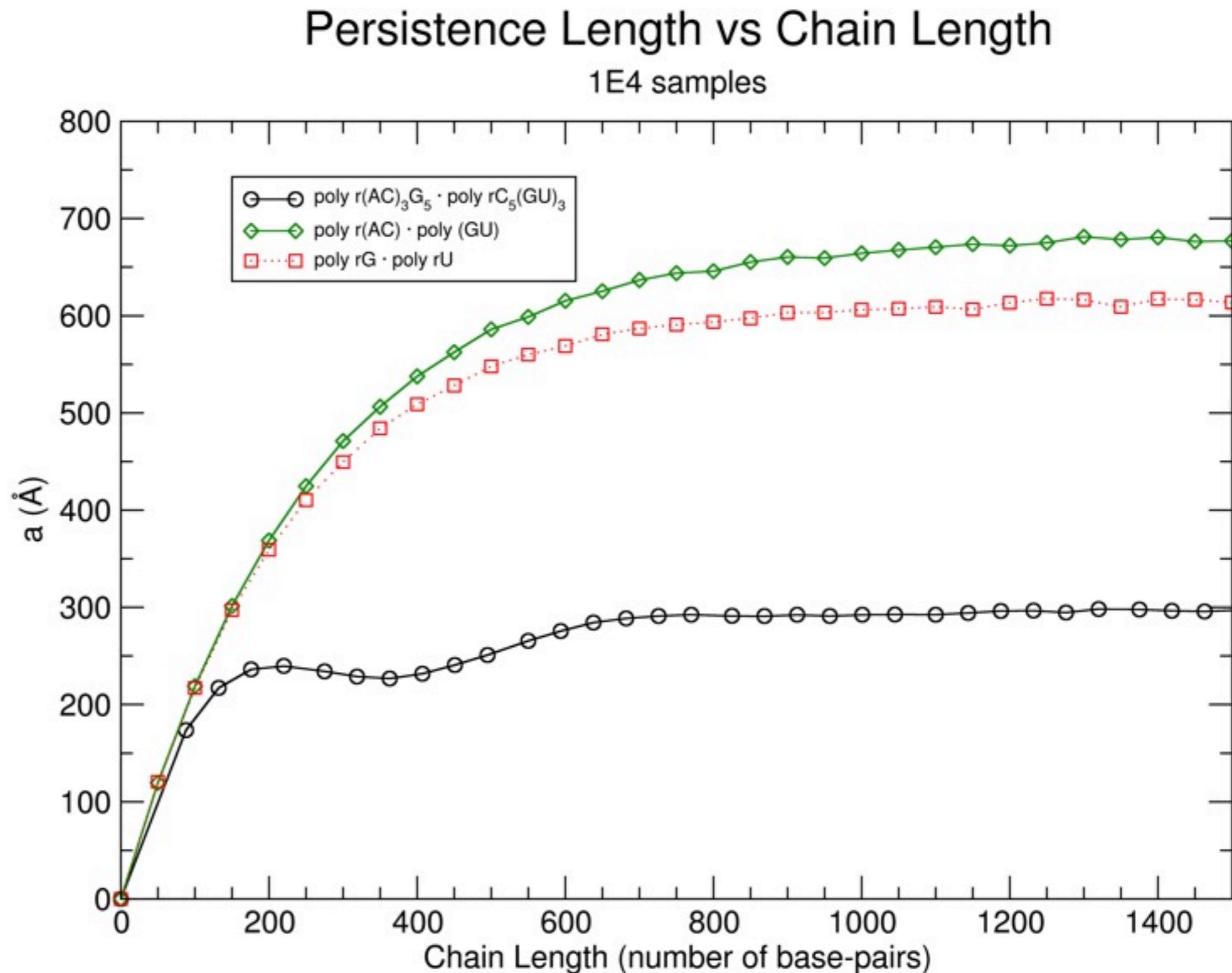
It is known that L_p is dependent on sequence for DNA. But what about for RNA? It seems also to be the case.



It even seems to be the case that some sequences of double-stranded RNA would have a preference for curving. Mini-circles?



It even seems to be the case that some sequences of double-stranded RNA would have a preference for curving. Mini-circles?



DET ÄR DAGS FÖR FIKA NU



PRACTICAL EXERCISE

Generating Nucleic Acid Structures using 3DNA and w3DNA,
interpreting rigid-body (geometric) parameters
and visualization tricks using pymol

First start with the most basic. The reconstruction of the commonly known conformations of nucleic acids.

Go to:

<http://w3dna.rutgers.edu>

Click in reconstruction.

Choose a fiber model (ADNA, BDNA, and ARNA)

Create a 24 base-pair model and download for later visualization via pymol.

Open your three pdb files in pymol.

```
set orthoscopic, 1  
bg white  
color black, elem C  
align bdna, adna  
align arna, adna  
translate [10,0,0], bdna  
translate [-10,0,0], adna
```

Put all of the previous commands into a file called conformations.pml

open with pymol:

pymol conformations.pml

Go again to w3dna.

upload one of your created conformations

get base-pair steps and download

modify

rebuild

More playing around with pairing recognition in 3DNA

Again go to w3dna, or if you can do it, do it on your machine.

Upload the Ijj2 structure of the ribosome to w3dna and analyze.

```
find_pair Ijj2.pdb stdout | analyze stdin
```

Look for the multiplets it has found.

Which is the largest multiplet?

Are there a lot of triplets?

How many?

How many quadruplets?

Take the corresponding PDB structures of the triplets and load them to pymol.

Produce a pretty plot in pymol showing the hydrogen bonding pattern of the quadruplets with the hydrogen bond distances.

Here are some suggestions for a structure or two to work with.

Pick one structure from the following options, or, if you know one related to your research use that one.

Dickerson-Drew dodecamer - classic for DNA

tRNAPhe -classic from Kim

Chromatin - classic from K. Luger

Ribosome - classic from A. Yonath

Enhanceosome -

Introns - Anna Pyle

RNAse P - Masquida Westhof

anti-NF-(kappa)B RNA aptamer

Ned Seeman nanoDNA -- 3gbi

Hammerhead Ribozyme

Locked Nucleic Acids -- 2x2q

Peptide Nucleic Acids -- 1pnn

Quadruplex -- 3ibk (Elizabeth H. Blackburn, Carol W. Greider and Jack W. Szostak)

miRNA -- 4ei I

Spliceosome -- 2kr8

Further information and selected references on nucleic acids and the rigid-body model

BOOKS

- Calladine, C. R. & Drew, H. R. & Luisi B. F. & Travers A. A
Understanding DNA: The Molecule and How It Works
Elsevier, 2004

- Saenger, W.
Principles of Nucleic Acid Structure
Springer-Verlag, 1984

- Flory, P.J.
Statistical Mechanics of Chain Molecules
Interscience Publishers, 1969

ARTICLES

- Watson, J. D. & Crick, F. H.
Molecular Structure of Nucleic Acids; A Structure for
Deoxyribose Nucleic Acid
Nature, 1953, 171, 737-738

- Calladine, C. R. & Drew, H. R.
A Base-centred Explanation of the B-to-A Transition in DNA
Journal of Molecular Biology, 1984, 178, 773-782

- Dickerson, R. E.; Bansal, M.; Calladine, C. R.; Diekman, S.;
Hunter, W. N.; Kennard, O.; von Kitzing, E.; Lavery, R.; Nelson, H.;
Olson, W. K. & Saenger, W.
Definitions and Nomenclature of Nucleic Acid Structure
Components
Nucleic Acid Research, 1989, 17, 1797-1803

- Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X.-J.; Neidle, S.; Shakkeb, Z.; Sklenar, H.; Suzuki, M.; Tung, C.-S.; Westhof, E.; Wolberger, C. & Berman, H.

A Standard Reference Frame for the Description of Nucleic Acid Base-pair Geometry
Journal of Molecular Biology, 2001, 313, 229-237

- Lu, X.-J. & Olson, W.
3DNA: A Software Package for the Analysis, Rebuilding and
Visualization of the Three-Dimensional Nucleic Acid Structures
Nucleic Acids Research, 2003, 31, 5108-5121

- Lu, X.-J. & Olson, W. K.
3DNA: A Versatile, Integrated Software System for the Analysis,
Rebuilding and Visualization of Three-Dimensional Nucleic-Acid
Structures
Nature Protocols, Department of Chemistry and Chemical
Biology and BioMaPS Institute for Quantitative Biology, Rutgers-
The State University of New Jersey, Piscataway, New Jersey
08854-8087, USA. 3dna.lu@gmail.com, 2008, 3, 1213-1227

- Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M. & Zhurkin, V. B.
DNA Sequence-Dependent Deformability Deduced from
Protein-DNA Crystal Complexes
Proceedings of the National Academy of Sciences, 1998, 95,
11163-11168

N. Seeman (2007) An overview of structural DNA
nanotechnology. Molecular Biotechnology 37, 246-257.

USE AND APPLICATION OF THE RIGID-BODY PARAMETER FORMALISM

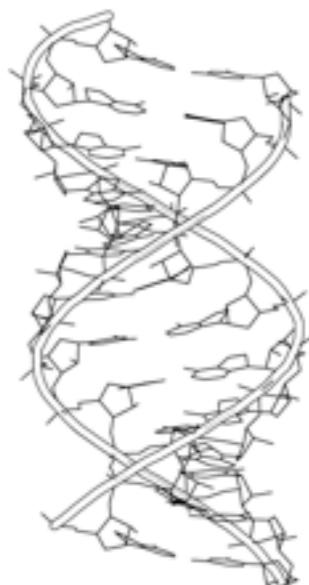
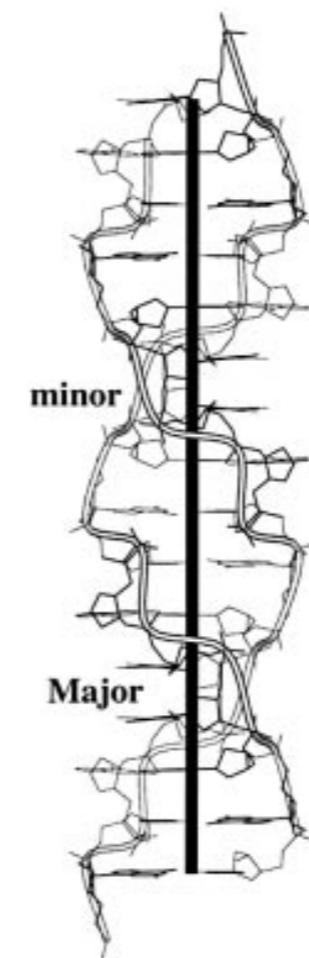
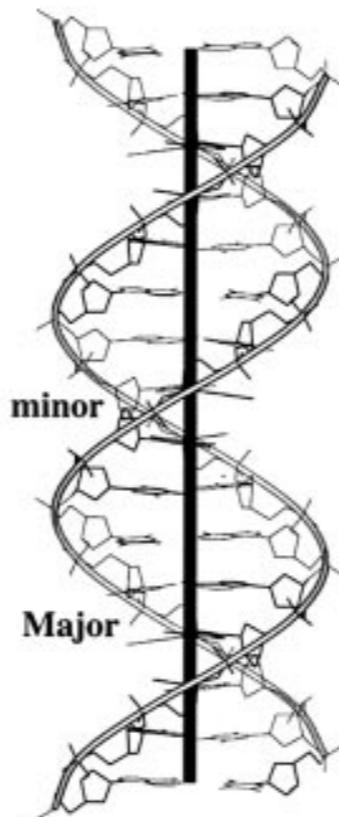
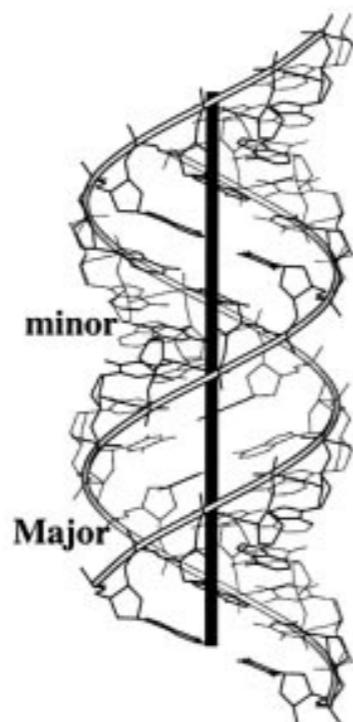
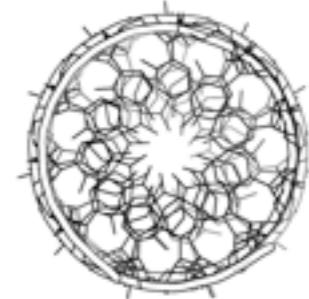
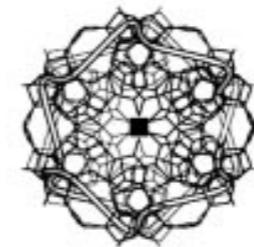
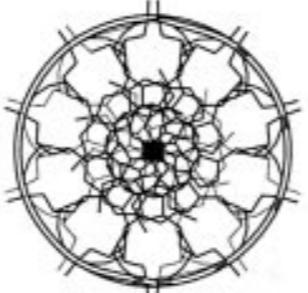
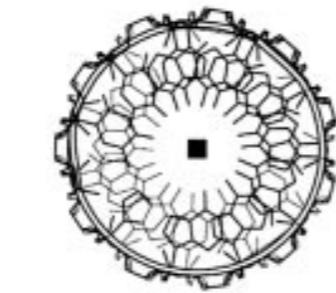
- General analysis of sequence dependent properties of DNA.
- Compute deformation scores for DNA based on X-ray data standards.
- From step-parameter information and inverse covariance analysis a link can be made to global polymer properties, e.g. persistence length, J-factors (cyclization probability)
- Compute topological properties of nucleic acids, e.g. linking number, twist, writhe.

Adding layers to DNA's simple complexity. DNA Triplex.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

ADNA BDNA ZDNA Triplex



Lu-Olson NAR, 2003, **31**, 5108-5121