**How to calculate inter-helical Euler angles using in-house script**

Huiqing Zhou, Honglue Shi, Dawn Merriman

**Materials**

All the files in **Materials** can be found at: summer:/home/hs189/ABG/Protocol/

**ABG Script**

There are two in-house scripts to calculate Euler angles: **allAtomMeas\_RNA.pl** and **allAtomMeas\_DNA.pl** for measuring ABG angles of two-way RNA and DNA motifs, respectively.

**Reference Helix**

**iAformRNA.pdb** is the reference helix used in allAtomMeas\_RNA.pl to align the lower helix of RNA (Bailor M., 2011). There is no need to change this reference frame or otherwise the script needs to be changed accordingly. **iBformDNA.pdb** is the reference helix used in allAtomMeas\_DNA.pl to align the lower helix of DNA. The reference DNA helix is built by the 3DNA program and aligned as described in the NAR paper (Zhou H., 2015), i.e. the helical axis is aligned along the z-direction and the C1'-C1' vector of the bp in the lower helix (H1) next to the junction is aligned along the Y-direction, in order to allow the output γh value describing the bending direction (towards major or minor groove). There is no need to change this reference frame or otherwise the script needs to be changed accordingly.

**Input PDB file**

The structure you want to perform ABG calculation.

**Formatting script**

Use the **convert.py** to format your input pdb file. Input pdb files has to be very well formatted or otherwise the script may give errors.

**Procedures**

To calculate the ABG angles of a two-way HJH motif, type the command below.

For RNA

|  |
| --- |
| perl allAtomMeas\_RNA.pl RNA.pdb X.4:12 X.18:23 '(((...(((' '))) )))' |

For DNA

|  |
| --- |
| perl allAtomMeas\_DNA.pl DNA.pdb X.4:12 X.18:23 '(((...(((' '))) )))' |

\* RNA.pdb and DNA.pdb are the input pdb files for which the ABG angles will be calculated;

\* X is the chain identifier of in the pdb files and should be changed according to the input pdb files;

\* 4:12 and 18:23 are residue numbers that refer to the residues that will be used in calculation. Here 3-bulge structures of RNA and DNA are used as examples;

\* The notation in quotes indicates the secondary structure of the residues that will be used in calculation. HIV-1-TAR RNA and DNA are used as examples here. “(” indicates one paired residue in 5' strand; “)” indicates one paired residue in 3' strand; “.” indicates one unpaired residue in bulge. If there is no unpaired residue, simply use a space to replace the bulge as shown above. Note that the secondary motif notation **DOES NOT** totally describe which helix in the target helix is recognized as the lower helix (H1); see FAQ.

\* For negative βh in DNA Euler angle calculation, one need to manually change (αh±180, -βh, γh±180) to make sure the βh is always positive and the γh will correspond to the projection of H2 helix onto the x-y plane and represents the bending direction of H2 relative to the aligned helix (Zhou H., 2015).

**FAQs**

1. **What is the Euler Degeneracy?**

The inter-helical Euler angles (αh, βh, γh) are two-fold degenerate when limiting αh, βh, γh to ±180. A second solution is obtained because a given set of Euler angles (αh±180, -βh, γh±180). In RNA calculation, this degeneracy is minimized by choosing the angles that minimize δ=sqrt(αh^2+βh^2+γh^2) (Bailor M., 2010). This serves to bias solutions to the pole (0,0,0), thus resulting in a compact distribution of Euler angles (Bailor M., 2011). In DNA calculation, the βh should be manually set to be 0 to 180 (also change αh and γh) based on the output value from the script (Zhou H., 2015).

1. **How does the program identify the lower helix (H1) and upper helix (H2)?**

Note that the secondary motif notation **DOES NOT** totally describe which helix in the input helix is recognized as the lower helix (H1). There are three criteria (sort by priority levels) in the perl script to determine which helix is the H1:

1. if one strand is longer than the other, H1 helix is the helix near the 5'-end of the long strand;
2. if the two strands have equal length and if residue numbers are fully numeric, the H1 helix refers to the helix containing the smallest numeric residue IDs;
3. if the two strands have equal length, identical residue numbers and different chain IDs, the H1 helix refers to the helix containing chain IDs with chain IDs earlier in the alphabetic order.

|  |
| --- |
| perl allAtomMeas\_alignedDNA.pl [file name of the input pdb] X.4:12 X.18:23 '(((...(((' '))) )))'  OR  perl allAtomMeas\_alignedDNA.pl [file name of the input pdb] X.18:23 X.4:12 '((( (((' ')))...)))'  H1 helix: X. 4:6 |
| perl allAtomMeas\_alignedDNA.pl [file name of the input pdb] X.4:10 X.18:24 '(((.(((' '))).)))'  OR  perl allAtomMeas\_alignedDNA.pl [file name of the input pdb] X.18:24 X.4:10 '(((.(((' '))).)))'  H1 helix: X. 4:6 |
| perl allAtomMeas\_alignedDNA.pl [file name of the input pdb] A.4:10 B.4:10 '(((.(((' '))).)))'  OR  perl allAtomMeas\_alignedDNA.pl [file name of the input pdb] B.4:10 A.4:10 '(((.(((' '))).)))'  H1 helix: A. 4:6 |

1. **What is defined as over- and under-twisting in the script?**

Inter-helical twist is defined as **ζh = αh + γh**.

For RNA ABG calculation using the current scripts (allAtomMeas\_RNA.pl) and the reference helix (iAformRNA.pdb), ζh < 0 and ζh > 0 corresponds to over-twisting and under-twisting, respectively (Bailor M., 2011); however, for DNA script (allAtomMeas\_DNA.pl) and reference helix (iBformDNA.pdb), due to the change in the reference helix, ζh > 0 and ζh < 0 correspond to over-twisting and under-twisting, respectively (The paper Zhou H., 2015 is incorrect about this). To use these interpretations for DNA and RNA, one needs to make sure using the exact same perl protocol and the reference helix to perform the calculation. Whenever changing the reference helix or the protocol, one needs to re-examine the over or under-twisting interpretation.

1. **How to examine the over- and under-twisting interpretation?**

For example, one can calculate ABG of any idealized B-form DNA.

|  |
| --- |
| perl allAtomMeas\_alignedDNA.pl iBformDNA.pdb A.4:10 B.4:10 '(((.(((' '))).)))' |

In this input helix, one would expect the H1 and H2 in the input helix being "over-twisting like" as they are separated by one bp: if the output αh+γh ~ +36º, positive ζh means over-twisting; if αh+γh ~ -36º, negative ζh means over-twisting.

Similar test can be performed using A-form RNA.

1. **Should we still use the ABG angles with RMSD larger than 2 Å?**

It depends. The calculated ABG angles with RMSD larger than 2 Å are not reliable and should be used with cautions. One should manually inspect the shape of the input structure with RMSD larger than 2 Å (Bailor M., 2011).

1. **What is the cavies if we have small βh?**

Theoretically, if βh = 0 or 180, then αh and γh could be arbitrary and should not be used for any conclusion. The reason is or the special case of perfectly parallel or antiparallel helices, there is a continuous coaxial degeneracy defined by (αh±D, βh=0,180, γh±D), D is an arbitrary value (Bailor M., 2011). In general, when you have smaller βh values, you have larger errors in αh and γh, and smaller errors in βh. You can still use these ABG angles with small βh, but you just have to note that αh/γh are not well defined and make sure not to over interrupt them.